

**NERMAP: APOIANDO O PROCESSO DE *ROADMAPPING* TECNOLÓGICO A
PARTIR DA TÉCNICA DE RECONHECIMENTO DE ENTIDADES NOMEADAS.**

Alan de Oliveira Lyra

Dissertação de Mestrado apresentada ao Programa
de Pós-graduação em Engenharia de Sistemas e
Computação, COPPE, da Universidade Federal do
Rio de Janeiro, como parte dos requisitos
necessários à obtenção do título de Mestre em
Engenharia de Sistemas e Computação.

Orientadores: Jano Moreira de Souza

Carlos Eduardo Barbosa

Rio de Janeiro

Junho de 2021

**NERMAP: APOIANDO O PROCESSO DE *ROADMAPPING* TECNOLÓGICO A
PARTIR DA TÉCNICA DE RECONHECIMENTO DE ENTIDADES NOMEADAS.**

Alan de Oliveira Lyra

**DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM
CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.**

Orientadores: Jano Moreira de Souza

Carlos Eduardo Barbosa

Aprovada por: Prof. Jano Moreira de Souza

Dr. Carlos Eduardo Barbosa

Prof.^a Adriana Santarosa Vivacqua

Prof.^a Adelaide Maria Souza Antunes

RIO DE JANEIRO, RJ - BRASIL

JUNHO DE 2021

Lyra, Alan de Oliveira

NERMAP: Apoiando o processo de *Roadmapping*
Tecnológico a partir da técnica de Reconhecimento de
Entidades Nomeadas. / Alan de Oliveira Lyra. – Rio de
Janeiro: UFRJ/COPPE, 2021.

XIII, 172 p.: il.; 29,7 cm.

Orientadores: Jano Moreira de Souza

Carlos Eduardo Barbosa

Dissertação (mestrado) – UFRJ / COPPE / Programa de
Engenharia de Sistemas e Computação, 2021.

Referências Bibliográficas: p. 119-126.

1. *Roadmapping* Tecnológico 2. Reconhecimento de
Entidades Nomeadas 3. Prospecção Tecnológica. I. Souza,
Jano Moreira de *et al.* II. Universidade Federal do Rio de
Janeiro, COPPE, Programa de Engenharia de Sistemas e
Computação. III. Título.

Agradecimentos

Antes de tudo, agradeço a Deus, por ter sido meu amparo em todos os momentos, por ter me dado força e perseverança, e por ter colocado em minha vida todas estas pessoas que, de diversas formas, me ajudaram a chegar até aqui.

À minha mãe e tia, que acreditaram e torceram tanto por mim, me dando todo apoio e amor incondicional.

A todos os meus familiares e amigos, pelo apoio e compreensão.

Aos meus orientadores Jano Moreira de Souza e Carlos Eduardo Barbosa, pela paciência, confiança e disponibilidade nesses anos de trabalho.

Aos professores Adriana Vivacqua e Adelaide Maria, por aceitarem fazer parte desta banca de avaliação.

A todos os colegas do PESC, por estes anos de convivência e trabalho em equipe, seja em disciplinas ou em projetos de P&D. Em especial aos colegas Herbert Salazar, Jonathan Augusto, Rebeca Motta e Renan Basílio, companheiros de projeto COPPETEC.

Também aos demais professores e funcionários do PESC, e a todos aqueles que, direta ou indiretamente, contribuíram para que este trabalho fosse realizado.

Agradeço também a CAPES pelo apoio financeiro a mim e ao meu programa, sem o qual este trabalho não poderia ter sido realizado.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

NERMAP: APOIANDO O PROCESSO DE *ROADMAPPING* TECNOLÓGICO A PARTIR DA TÉCNICA DE RECONHECIMENTO DE ENTIDADES NOMEADAS.

Alan de Oliveira Lyra

Junho/2021

Orientadores: Jano Moreira de Souza

Carlos Eduardo Barbosa

Programa: Engenharia de Sistemas e Computação

As sociedades são surpreendidas com incessantes mudanças no mundo atual de forma gradativa, acentuando não somente os níveis de incerteza, como também a complexidade dos processos envolvidos, caminhando para uma dinâmica que parece ser diferente de tudo o que se conhece e já foi experimentado atualmente. Esse futuro observado proporciona uma demanda maior por estudos referentes ao planejamento, prospecção, diagnósticos e visões de futuro, no qual Prospecção Tecnológica (*Future-oriented Technology Analysis - FTA*) tem sido considerado fundamental para promover a criação da capacidade de organizar sistemas de inovação que respondam aos interesses da sociedade. O *Roadmapping* Tecnológico configura um dos diversos métodos pertencentes a um estudo de FTA cujo produto esperado é o *roadmap*, apresentando um cronograma temporal relacionando tecnologias e produtos, de forma a apresentar as prospecções tecnológicas previstas de forma cronológica. Este trabalho propõe um *framework* para apoiar o processo de FTA capaz de semiautomatizar o processo de *Roadmapping* Tecnológico fazendo uso de Reconhecimento de Entidades Nomeadas, denominado NERMAP. O sistema NERMAP foi implementado e avaliado, sob à luz do *framework Design Science* proporcionando o apoio ao FTA.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

**NERMAP: SUPPORTING THE TECHNOLOGY ROADMAPPING PROCESS
USING THE NAMED ENTITY RECOGNITION TECHNIQUE.**

Alan de Oliveira Lyra

June/2021

Advisors: Jano Moreira de Souza

Carlos Eduardo Barbosa

Department: Computer Science and Systems Engineering

Societies are gradually surprised by unceasing changes in the current world, accentuating not only the levels of uncertainty, but also the complexity of the involved processes, moving towards a dynamic that seems to be different from all that is known and has already been experienced lately. This observed future provides a greater demand for studies related to planning, prospecting, diagnostics and future visions, in which Future-oriented Technology Analysis (FTA) has been considered fundamental to promote the creation of the capacity to organize innovation systems that respond to society interests. Technology Roadmapping configures one of the several methods belonging to an FTA study which expected final product is the roadmap, presenting a timeline relating technologies and products, in order to present the technological prospects chronologically organized. This study proposes a framework to support the FTA process capable of semiautomating the Technology Roadmapping process using Named Entity Recognition, called NERMAP. The NERMAP system was implemented and evaluated, in the light of the Design Science framework, providing support to the FTA.

Sumário

Capítulo 1 – Introdução	1
1.1. Contexto e Motivação.....	2
1.2. Objetivo e Metodologia	4
1.3. Organização do Trabalho.....	5
Capítulo 2 – Future-oriented Technology Analysis	6
2.1. Tipos de FTA	6
2.1.1. <i>Technology Forecasting</i>	7
2.1.2. <i>Foresight</i>	7
2.1.3. <i>Technology Assessment</i>	8
2.2. Classificação dos métodos de FTA.....	8
2.2.1. Pela sua Natureza	8
2.2.2. Por Abordagem.....	9
2.2.3. Pela sua Capacidade	9
2.2.4. Por Famílias de Métodos	10
2.3. Métodos de FTA	13
Capítulo 3 – Roadmapping Tecnológico	15
3.1. Por que Elaborar <i>Roadmaps</i> ?.....	16
3.2. Metodologias de <i>Roadmaps</i>	18
3.3. Tipologia de <i>Roadmaps</i>	22
3.3.1. Propósitos	22
3.3.2. Formatos.....	27
3.3.3. Aplicações, Ênfase e Propósito.....	32
3.4. Abordagens para Implantações de TRM.....	36
3.5. Benefícios de Realizar TRM.....	42
Capítulo 4 – Identificando Prospecções em Textos	44
4.1. Conceitos Fundamentais	44
4.2. Uma Visão Geral de Mineração de Textos	46
4.2.1. Processamento de Linguagem Natural.....	49
4.2.2. Corpus	53
4.2.3. Texto Marcado.....	55
4.2.4. Extração da Informação	56
4.3. Reconhecimento de Entidades Nomeadas	59
4.3.1. Conferências de REN	60
4.3.2. Resolução e Anotação de Homônimos	65
4.3.3. Métodos Probabilísticos.....	66

4.3.4.	Medidas para Avaliação e Desempenho	69
4.3.5.	Aprendizado de Máquina.....	72
Capítulo 5 – O framework NERMAP	76
5.1.	Conceitos e Modelo	76
5.1.1.	Processamento de Documentos.....	78
5.2.	Método: Processo de <i>Roadmapping</i> Tecnológico.....	82
5.3.	Instanciação: O sistema NERMAP	84
5.3.1.	Arquitetura do NERMAP	84
5.3.2.	Modelagem do NERMAP.....	93
Capítulo 6 – Avaliação do NERMAP	99
6.1.	Metodologia	99
6.2.	Avaliação dos Conceitos.....	101
6.3.	Avaliação do Modelo.....	102
6.3.1.	Métricas do Tagueamento.....	104
6.4.	Avaliação do Método.....	105
6.5.	Avaliação da Instanciação.....	106
6.5.1.	<i>Journal Futures</i> (Estudo de Caso)	107
6.5.2.	Discussão.....	114
Capítulo 7 – Considerações Finais	116
Referências Bibliográficas	119
Anexos	127
1.	Experimento: Prospecções do <i>Journal Futures</i> (2020)	127
2.	Ciclo 1: Prospecções obtidas pelo NERMAP (<i>Journal Futures</i>)	141
3.	Ciclo 2: Prospecções obtidas pelo NERMAP (<i>Journal Futures</i>)	147
4.	Ciclo 3: Prospecções obtidas pelo NERMAP (<i>Journal Futures</i>)	160

Lista de Figuras

Figura 1. Exemplo de representação de um método de FTA por sua capacidade. Adaptado de BARBOSA (2018).	10
Figura 2. Modelo de um <i>roadmap</i> , adaptado de PHAAL et al. (2004).	19
Figura 3. Formato característico de um <i>roadmap</i> . Adaptado de (ALBRIGHT, 2005).	20
Figura 4. Caracterização de um TRM genérico – propósitos e formatos. Adaptado de PHAAL et al. (2004).	22
Figura 5. <i>Roadmap</i> de planejamento de produto – realizado pela <i>Philips</i> . Adaptado de PHAAL et al. (2004).	23
Figura 6. <i>Roadmap</i> de planejamento de serviços/capacidade – <i>Royal Mail</i> . Adaptado de PHAAL et al. (2004).	23
Figura 7. <i>Roadmap</i> de planejamento estratégico. Adaptado de PHAAL et al. (2004).	24
Figura 8. <i>Roadmap</i> de planejamento de longo prazo – realizado pela <i>U.S. Integrated Manufacturing Technology Roadmapping Initiative</i> . Adaptado de PHAAL et al. (2004).	25
Figura 9. <i>Roadmap</i> de planejamento do capital intelectual – realizado pela <i>Artificial Intelligence Applications Unit - University of Edinburgh</i> . Adaptado de PHAAL et al. (2004).	25
Figura 10. <i>Roadmap</i> de planejamento de programas – realizado pela NASA para o projeto <i>Origins</i> . Adaptado de PHAAL et al. (2004).	26
Figura 11. <i>Roadmap</i> de planejamento de processos. Adaptado de PHAAL et al. (2004).	26
Figura 12. <i>Roadmap</i> de planejamento de integração – realizado pela NASA. Adaptado de PHAAL et al. (2004).	27
Figura 13. <i>Roadmap</i> de múltiplos níveis – realizado pela <i>Philips</i> . Adaptado de PHAAL et al. (2004).	28
Figura 14. <i>Roadmap</i> em forma de barra – realizado pela <i>Motorola</i> . Adaptado de PHAAL et al. (2004).	28
Figura 15. <i>Roadmap</i> em forma de tabela. Adaptado de PHAAL et al. (2004).	29
Figura 16. <i>Roadmap</i> em forma de gráfico. Adaptado de PHAAL et al. (2004).	30
Figura 17. <i>Roadmap</i> em forma imagens – realizado pela <i>Sharp</i> . Adaptado de PHAAL et al. (2004).	30
Figura 18. <i>Roadmap</i> em forma de fluxograma – realizado pela NASA. Adaptado de PHAAL et al. (2004).	31
Figura 19. Taxonomias de <i>roadmaps</i> . Adaptado de KAPPEL (2001).	32
Figura 20. Arquitetura do <i>roadmap</i> para tecnologias convergentes. Adaptado de ALBRIGHT (2005).	35
Figura 21. Passos da abordagem <i>T-Plan</i> padrão. Adaptado de PHAAL et al. (2004).	37
Figura 22. Arquitetura de <i>roadmap</i> generalizada. Adaptado de PHAAL et al. (2004).	39
Figura 23. <i>Roadmaps</i> integram conhecimento comercial e tecnológico. Adaptado de PHAAL et al. (2004).	40
Figura 24. Estágios de implantação de um TRM. Adaptado de VATANANAN e DANSAMASADIT (2009).	41
Figura 25. Diagrama da metodologia de Mineração de Textos. Adaptado de ARANHA (2007).	46
Figura 26. Tipos de Descoberta de Conhecimento. Adaptado de MORAIS e AMBRÓSIO (2007).	47
Figura 27. Etapas de Mineração de Textos (KDD). Adaptado de FAYYAD et al. (1996).	47

Figura 28. Intersecção das áreas de atuação e campos de trabalho de Mineração de Textos. Adaptado de MINER et al. (2012) e CARVALHO (2017).....	49
Figura 29. <i>Template</i> de um sistema de extração de informações. Adaptado de ZAMBENEDETTI (2002).....	58
Figura 30. Árvore de categorias do HAREM. Adaptado de (MOTA; SANTOS, 2008).....	64
Figura 31. Relevância dos termos. Adaptado de (MAGALHÃES, 2016).....	70
Figura 32. Fronteiras de Classificação dos Termos - Baseado de <i>Precision</i> e <i>Recall</i> . Adaptado de (KONKOL, 2012).....	71
Figura 33. Arquitetura simplificada do Aprendizado Supervisionado. Adaptado de (SILVA, 2012).....	73
Figura 34. Arquitetura simplificada do Aprendizado Semissupervisionado. Adaptado de (SILVA, 2012).....	74
Figura 35. Modelo de geração do <i>roadmap</i>	77
Figura 36. Composição de TRMs.....	78
Figura 37. Melhor resultado de parâmetros.....	81
Figura 38. Processo de Testes.....	82
Figura 39. Processo NERMAP.....	83
Figura 40. Painel de Início do NERMAP.....	84
Figura 41. Camadas da arquitetura do NERMAP.....	85
Figura 42. A Arquitetura do NERMAP.....	86
Figura 43. Cadastro de usuário do NERMAP.....	87
Figura 44. Edição de perfil do usuário do NERMAP.....	87
Figura 45. Gerenciamento de TRMs do NERMAP.....	88
Figura 46. Compartilhamento de TRMs do NERMAP.....	88
Figura 47. Gerenciamento de Documentos do NERMAP.....	89
Figura 48. Arquivos do TRM.....	89
Figura 49. <i>Roadmap</i> sendo exibido no NERMAP.....	90
Figura 50. Filtro sendo aplicado ao <i>roadmap</i>	90
Figura 51. Edição da prospecção.....	91
Figura 52. Visualização das prospecções nos documentos no NERMAP.....	92
Figura 53. Adicionando prospecção manualmente ao <i>roadmap</i>	92
Figura 54. Modos de exportação do <i>roadmap</i>	93
Figura 55. Diagrama de Casos de Uso do NERMAP.....	94
Figura 56. Diagrama de Classes do NERMAP.....	95
Figura 57. Diagrama de Atividades do NERMAP.....	97
Figura 58. Diagrama Entidade Relacionamento do Banco de Dados do NERMAP.....	98
Figura 59. Aplicação do Modelo de geração do <i>roadmap</i> para o cenário proposto.....	102
Figura 60. Composto de <i>roadmaps</i> para o cenário proposto.....	103
Figura 61. Processo NERMAP aplicado ao cenário proposto.....	105
Figura 62. Aplicação do Modelo de geração do <i>roadmap</i> para o Estudo de Caso.....	107
Figura 63. Ciclo 1: Gráfico das prospecções das publicações da <i>Futures</i> (2020) analisadas manualmente.....	109
Figura 64. Ciclo 1: Gráfico das prospecções obtidas no NERMAP.....	110
Figura 65. Ciclo 2: Gráfico das prospecções das publicações da <i>Futures</i> (2020) analisadas manualmente.....	111
Figura 66. Ciclo 2: Gráfico das prospecções obtidas no NERMAP.....	112
Figura 67. Ciclo 3: Gráfico das prospecções das publicações da <i>Futures</i> (2020) analisadas manualmente.....	113
Figura 68. Ciclo 3: Gráfico das prospecções obtidas no NERMAP.....	114

Lista de Tabelas

Tabela 1. Métodos de FTA agrupados pela classificação de famílias. Adaptado de PORTER et al. (2004).....	14
Tabela 2. Segmentos dos <i>roadmaps</i> sob a perspectiva “ <i>why-what-how-when</i> ”, listando seus objetivos. Adaptado de ALBRIGHT (2005).....	21
Tabela 3. Componentes de um <i>roadmap</i> em CT&I. Adaptado de ALBRIGHT (2005).....	33
Tabela 4. Exemplo de identificação de prospecções em textos.....	45
Tabela 5. Exemplo de <i>roadmap</i> a ser formado.....	45
Tabela 6. Excerto de texto do corpus <i>Tycho Brahe</i> (GALVES; BRITTO, 2002).....	56
Tabela 7. Excerto de texto do corpora <i>Penn Parsed</i> (KROCH; TAYLOR, 2000).....	56
Tabela 8. Categorias do segundo HAREM. Adaptado de (OLIVEIRA; MOTA; FREITAS, 2008).....	64
Tabela 9. Exemplo de Categorização Ambígua de Entidades.....	65
Tabela 10. Exemplo tagueamento usando a notação BILOU.....	80
Tabela 11. Parâmetros Utilizados.....	81
Tabela 12. Descrição dos Casos de Uso do NERMAP.....	94
Tabela 13. Descrição das Classes do NERMAP.....	96
Tabela 14. Descrição das Atividades do NERMAP.....	97
Tabela 15. Metodologia de Avaliação. Adaptado de MARCH e SMITH (1995).....	100
Tabela 16. Resultado do experimento para as EN.....	104
Tabela 17. Resultado do experimento para o Modelo NERMAP.....	104
Tabela 18. Resultado do experimento para o novo Modelo NERMAP.....	113

Lista de Acrônimos e Siglas

- ACE - *Automatic Content Extraction*
AHP - *Analytical Hierarchy Process*
AI - *Artificial Intelligence*
API - *Application Programming Interface*
BPMN - *Business Process Model and Notation*
CAS - *Complex Adaptive System*
CD - Coleção Dourada
CoNLL - *Conference on Computational Natural Language Learning*
CRF - *Conditional Random Field*
CT&I - Ciência, Tecnologia e Inovação
DC - Descoberta de Conhecimento
DEA - *Data Envelopment Analysis*
DER - Diagrama Entidade Relacionamento
EM - Entidade Mencionada
EN - Entidades Nomeadas
EI - Extração de Informação
EIRMA - *European Industrial Research Management Association*
ER - Extração de Relações
EUA - Estados Unidos da América
FAR - *Field Anomaly Relaxation*
FN - Falso Negativo
FP - Falso Positivo
FTA - *Future-oriented Technology Analysis*
GBN - *Global Business Network*
GC - Gestão do Conhecimento
GDP - *Gross Domestic Product*
HAREM - Avaliação de Sistemas de Reconhecimento de Entidades Mencionadas
HMM - *Hidden Markov Models*
IMTR - *Integrated Manufacturing Technology Roadmapping*
IPTS - *Institute for Prospective Technological Studies*
ITRS - *International Technology Roadmap for Semiconductors*
KD - *Knowledge Discovery*
KDD - *Knowledge Discovery in Databases*
KDT - *Knowledge Discovery from Text*
MD - Mineração de Dados
ME - *Maximum Entropy*
MEMM - *Maximum-Entropy Markov Model*
MT - Mineração de Textos
MTC - *Mail Technology Conference*
MUC - *Message Understanding Conference*
NASA - *National Aeronautics and Space Administration*
NER - *Named Entity Recognition*
NSF - *National Science Foundation*
NGP - *Nominal Group Process*
OTA - *Office of Technology Assessment*
P&D - Pesquisa e Desenvolvimento
PLN - Processamento de Linguagem Natural
POS - *Part-of-Speech*

REM - Reconhecimento de Entidades Mencionadas
REN - Reconhecimento de Entidades Nomeadas
REX - *Rosette Entity Extractor*
RI - Recuperação de Informação
SIA - *Semicondutor Industry Association*
SOFI - *State of the Future Index*
TIA - *Trend Impact Analysis*
TFA - *Technology Futures Analysis*
TRM - *Technology Roadmapping*
TRIZ - *Teoriya Resheniya Izobretatelskikh Zadatch*
UML - *Unified Modeling Language*
VN - Verdadeiro Negativo
VP - Verdadeiro Positivo

Capítulo 1 – Introdução

As constantes mudanças com a quais as sociedades se deparam crescem de forma rápida, intensificando não somente os níveis de incerteza, como também a complexidade dos processos, apontando para uma dinâmica que parece ser diferente de tudo o que se conhece e já foi experimentado atualmente.

Essa visão de futuro proporciona uma demanda maior por estudos referentes ao planejamento, prospecção, diagnósticos e visões de futuro por parte de governos, instituições voltadas para P&D e corporações.

O esforço dos pesquisadores para encontrar procedimentos para estudos sistemáticos das tendências e fatos futuros gerou grande variedade de métodos e técnicas de prospecção. A Prospecção Tecnológica ou *Future-oriented Technology Analysis* (FTA), de acordo com PORTER et al. (2004), é a coleção de diversos métodos e práticas sistemáticas para prever e apoiar as decisões sobre tecnologias emergentes, incluindo o seu desenvolvimento e impactos futuros.

Dentre os vastos métodos que abrangem o FTA, este trabalho destaca o *Roadmapping* Tecnológico – do inglês, *Technology Roadmapping* (TRM), que é uma técnica flexível e amplamente utilizada para apoiar a gestão e o planejamento tecnológico de uma organização. O seu uso mais comum é para apoiar o planejamento de produtos, onde é utilizado para relacionar as novas tecnologias com os produtos reais (PHAAL; FARRUKH; PROBERT, 2004). O produto esperado de um estudo de TRM, o documento resultante do processo, é o *roadmap* tecnológico, considerado como um dos primeiros passos para a inovação tecnológica e que, tipicamente, apresenta um cronograma temporal relacionando tecnologias e produtos (COELHO et al., 2005). *Roadmaps* foram popularizados na década de 80, a partir da abordagem utilizada pela *Motorola*, podendo ser definidos como uma visão de futuro, geralmente coletiva, que considera as tendências atuais de um campo em um formato de cronograma temporal (GALVIN, 2004).

O processo de planejamento do TRM é impulsionado pela necessidade de tecnologias, ajudando a identificar, selecionar e desenvolver tecnologias alternativas para satisfazer um determinado conjunto de necessidades ou de produtos já definidos. Esse processo para gerar *roadmaps* é realizado por grupos de especialistas que coletam informações de documentos – normalmente publicações que prospectam o futuro de

tecnologias – para serem usadas como insumo. Se trata de um processo *time-consuming*¹, cansativo e que exige uma densa e boa equipe, o que gera custos e tempo para ser executado. Esse fato demonstra um caso importante para contar com um processo de automação, de modo a reduzir esses custos e tornar todo o processo mais rápido e eficiente. Desse modo, aumentar a capacidade da organização de se preparar para o futuro, aumentando a sua competitividade.

Neste trabalho, propomos um sistema capaz de semiautomatizar o processo de *Roadmapping* Tecnológico usando a técnica de Reconhecimento de Entidades Nomeadas (REN), contribuindo para à área de FTA de modo a ser possível aos pesquisadores aplicarem o processo de TRM em uma grande massa de documentos de prospecção gastando muito menos tempo e esforço. O sistema proposto obteve 83% de eficácia em prospecções retornadas para composição do *roadmap* e se mostrou uma ferramenta bastante útil encontrando prospecções em grande volume de textos que, ao serem analisados manualmente, dificilmente seriam encontradas.

1.1. Contexto e Motivação

No âmbito de sistemas de Ciência, Tecnologia e Inovação (CT&I), os exercícios prospectivos ou de prospecção tecnológica têm sido considerados fundamentais para promover a criação da capacidade de organizar sistemas de inovação que respondam aos interesses da sociedade. A partir de intervenções planejadas em sistemas de inovação, fazer prospecção significa identificar quais são as oportunidades e necessidades mais importantes para a pesquisa e desenvolvimento (P&D) no futuro (COELHO et al., 2004).

Roadmaps são populares tanto na indústria quanto em instituições governamentais, pois permitem uma visão ampliada do futuro de um determinado campo de estudo, desenhado a partir do conhecimento coletivo e da imaginação de grupos e indivíduos sobre as forças motrizes que atuam no campo sob estudo. Incluem a identificação de teorias e tendências, formulação de modelos, identificação de elos da ciência, descontinuidades e *gaps*² do conhecimento, interpretação de pesquisas e experimentos (COELHO et al., 2005).

¹ *Time-consuming* é um termo em inglês usado para algo que consome/gasta muito tempo.

² *Gaps* é um termo em inglês que significa um distanciamento; afastamento, separação, uma lacuna ou um vácuo.

Elaborar um *roadmap* tecnológico é considerado crítico quando as decisões sobre as medidas a serem tomadas são complexas. O principal benefício do *roadmap* é que este provê informações que permitem tomar melhores decisões sobre investimentos em P&D, pela identificação das tecnologias críticas e dos *gaps* existentes e identificar formas de alavancar investimentos (COELHO et al., 2005).

Em contrapartida, todo o processo de TRM realizado por grupos de especialistas que tem por resultado um documento de *roadmap* tecnológico, embora fácil, se configura um processo complexo e que demanda bastante tempo conforme o número de eventos aumenta, em especial quando são usadas diversas fontes de dados para coletar esses eventos. Esse processo demanda, além de tempo, uma equipe composta de especialistas que geram custos para as organizações.

Um sistema capaz de semiautomatizar esse método se configura em um amplo salto para a área de FTA, apoiando os pesquisadores durante o processo de TRM e permitindo uma análise de uma grande massa de documentos de diferentes fontes de dados. Dessa forma, entregando melhores resultados, reduzindo custos e otimizando o tempo durante um estudo de FTA.

Como exemplo desse processo, o pesquisador de FTA busca em um estudo de TRM extrair frases com perspectiva temporal em textos, como exemplo: “*In 2030, China will lead the world in Artificial Intelligence.*” e as organizar em um formato de linha do tempo – o *roadmap*. Dessa forma, inserindo documentos manualmente para uma determinada área de estudo definida, o sistema proposto auxilia o pesquisador a realizar o processo de TRM – em especial na extração e ordenamento das prospecções – analisando essa gama de documentos e identificando trechos nos textos que indicam um acontecimento futuro, podendo ou não serem classificados como prospecções pelo sistema para serem, posteriormente, refinadas pelo pesquisador.

A análise e a geração do *roadmap* é feita, de maneira semiautomatizada e de forma “macro” para a área escolhida, de modo que o usuário pode, dentro da ferramenta, direcionar o resultado para subáreas específicas através de filtros de palavras/termos chaves.

Embora o problema deste trabalho esteja organizado para descobrir frases com perspectiva temporal futura, o sistema proposto também é capaz de fazer análises do passado. De modo a identificar, em documentos antigos, previsões de acontecimentos futuros que já se tornaram passados, para assim, poder validar tais prospecções e obter *insights* para onde uma tendência do passado caminhou.

Diversas técnicas inseridas no contexto de Mineração de Textos e Inteligência Artificial podem ser aplicadas para a realização deste trabalho e o Reconhecimento de Entidades Nomeadas foi escolhida dentre essas técnicas, por ser consolidada como uma tarefa primordial em Extração de Informação (JIANG, 2012) e utiliza modelos de Inteligência Artificial para detectar e classificar entidades como pessoa, lugar e organização em textos, sendo amplamente utilizado associado à métodos de Aprendizado Supervisionado (CHINCHOR; HIRSCHMAN; LEWIS, 1993).

1.2. Objetivo e Metodologia

O objetivo deste trabalho é apoiar o pesquisador durante uma prospecção tecnológica a semiautomatizar o processo de geração de *roadmaps* a partir de um conjunto de textos. De maneira a entregar resultados mais eficientes durante um processo de TRM e reduzir tanto os custos quanto o tempo gastos pelas organizações durante um estudo de FTA, aumentando a sua competitividade.

Para atingir esse objetivo, serão apresentados um modelo e um processo. Esses artefatos são então incorporados (ou implementados) para produzir um sistema de Tecnologia da Informação (TI), denominado **NERMAP**, capaz de realizar um trabalho semiautomatizado para a geração de *roadmaps* no contexto de prospecção tecnológica.

O *framework* proposto foi avaliado através do paradigma de pesquisa *Design Science*. Dentro deste paradigma existem diferentes *frameworks* de trabalho. Nesta pesquisa utilizamos o *framework* de MARCH e SMITH (1995) devido à sua compatibilidade com os artefatos propostos deste trabalho, os avaliando em relação aos conceitos, modelo, método e instanciação.

De acordo com MARCH e SMITH (1995), o *Design Science* representa melhor o conceito de “computação aplicada”, enquanto as ciências naturais representam melhor os fundamentos da computação. O *Design Science* é melhor aplicável pelo fato da computação aplicada estudar fenômenos artificiais, que somente possuem sentido por causa de criações humanas como, por exemplo, as organizações. As ciências naturais se baseiam em explicar como as coisas são e os seus porquês; o *Design Science* está focado em construir artefatos para atingir objetivos (MARCH; SMITH, 1995).

1.3. Organização do Trabalho

Este trabalho está estruturado em 6 partes, organizadas da seguinte forma:

O Capítulo 1 apresenta o contexto, as motivações desta pesquisa, a metodologia adotada e os objetivos a serem alcançados.

O Capítulo 2 apresenta a revisão literária expondo as origens e os conceitos do *Future-oriented Technology Analysis*, apresentando suas classificações e tipologia de métodos.

O Capítulo 3 apresenta uma revisão literária do método de FTA explorado neste trabalho, o *Roadmapping Tecnológico*. Essa revisão fornece a base necessária para que o leitor tenha uma visão geral sobre as técnicas utilizadas e o contexto dos problemas propostos.

No Capítulo 4 é apresentado os elementos relacionados ao Reconhecimento de Entidades Nomeadas, parte do processo de Mineração de Textos, discutindo também sobre métricas de avaliação e desempenho e aprendizado de máquina, conceitos fundamentais para o desenvolvimento deste trabalho.

O Capítulo 5 apresenta o *framework NERMAP*, desde a caracterização dos seus **conceitos** até a proposta de um **modelo**, que descreve o relacionamento desses conceitos. Todo o processo também é apresentado, cobrindo a definição de **método** em termos de *Design Science*, seguido da apresentação da **implementação** do *framework NERMAP*, com o objetivo de apoiar o processo de *Roadmapping Tecnológico*.

O Capítulo 6 apresenta a avaliação dos artefatos desenvolvidos durante este trabalho – conceitos, modelo, método e implementação.

O Capítulo 7 apresenta as considerações finais sobre esta pesquisa, incluindo as limitações, as contribuições e os trabalhos futuros.

Capítulo 2 – *Future-oriented Technology Analysis*

O termo *Future-oriented Technology Analysis* foi usado pela primeira vez em 2004 no título do seminário “*New Horizons and Challenges for Future-oriented Technology Analysis: New Technology Foresight, Forecasting and Assessment*” (SCAPOLO, 2005), organizado pelo Instituto de Estudos Tecnológicos Prospectivos (*Institute for Prospective Technological Studies – IPTS*). É uma evolução do termo *Technology Futures Analysis* (TFA), que foi abandonado porque não passava a ideia de que era orientado ao futuro, mas orientado a tecnologia (JOHNSTON, 2008).

FTA é um conceito amplo, que inclui qualquer metodologia sistemática para prever e apoiar as decisões sobre tecnologias emergentes, incluindo o seu desenvolvimento e impactos futuros (PORTER et al., 2004). Seu conceito foi definido como o chamado “*umbrella*” – em português, “guarda-chuva” – que cobriu vários métodos diferentes de Prospecção Tecnológica no campo do *Technology Foresight*, *Technology Forecasting* e *Technology Assessment* (CAGNIN et al., 2008).

Segundo MARINELLI et al. (2014), o FTA é particularmente útil quando as tecnologias são caras porém essenciais para o desenvolvimento de um país, região ou empresa.

Neste trabalho, por questões de simplicidade, consideraremos o termo **Prospecção Tecnológica** como o termo que engloba todos esses campos do FTA.

2.1. Tipos de FTA

Atualmente coexistem muitas formas de analisar o futuro e suas consequências – dos quais *Technology Forecasting*, *Foresight*, *Futuribles*, *La Prospective*, *Scenarios*, *Technology Assessment*, *Technological Watch*, *Veille Technologique*, *Environmental Scanning* e *Vigilancia Tecnológica* constituem-se em alguns exemplos.

O FTA é um processo que aborda essas técnicas e uma variedade de métodos, permitindo uma caracterização detalhada e uma análise sistemática da tecnologia, bem como a identificação e apresentação de seus caminhos de desenvolvimento. Esse processo compõe uma ampla gama de atividades voltadas para o futuro que podem ser divididas em três principais campos do FTA, que são: o *Technology Forecasting*, o *Foresight* e o *Technology Assessment*.

2.1.1. *Technology Forecasting*

Technology Forecasting trabalha informações de evolução histórica, modelos matemáticos e projeção de situações futuras. Segundo PORTER et al. (2004) é o processo de descrever a emergência, desempenho, características ou os impactos de uma tecnologia em algum momento no futuro.

Designa as atividades de prospecção que têm foco nas mudanças tecnológicas, normalmente centradas nas mudanças na capacidade funcional, no tempo e no significado de uma inovação (COELHO et al., 2004).

As organizações sistematizam seu *Technology Forecasting* para prever eventos incertos. Para isso, é necessário identificar os problemas ou desafios que podem utilizar o *Technology Forecasting*, selecionar e aplicar os métodos apropriados para cada problema, e só assim, avaliar e refinar o processo com o passar do tempo (HYNDMAN; ATHANASOPOULOS, 2014).

2.1.2. *Foresight*

Foresight é definido como um processo que se ocupa em, sistematicamente, examinar o futuro de longo prazo da ciência, da tecnologia, da economia e da sociedade, com o objetivo de identificar as áreas de pesquisas estratégicas e as tecnologias emergentes que tenham a propensão de gerar os maiores benefícios econômicos e sociais (MACLEAN; ANDERSON; MARTIN, 1998).

Segundo AMANATIDOU (2014), *Foresight* é um instrumento que estrutura e facilita a definição de estratégias através da antecipação dos problemas, considerando as alternativas futuras.

Já a abordagem de HORTON (1999) defende *Foresight* como um “processo de desenvolvimento de visões de possíveis caminhos nos quais o futuro pode ser construído, compreendendo que as ações do presente contribuirão com a construção da melhor possibilidade do amanhã”.

O *Foresight* requer dos seus praticantes um pensamento criativo e perspectivas multidisciplinares. Ao contrário do *Technology Forecast*, o *Foresight* dá ênfase ao processo, e não ao conjunto de métodos; seu objetivo é construir sistematicamente os caminhos de desenvolvimento tecnológico do ponto presente ao ponto futuro baseado nas ações passíveis de serem tomadas, e não analisar tendências de futuro para apontar o mais provável (MARTIN, 2010).

2.1.3. *Technology Assessment*

Technology Assessment – ou Avaliação de Tecnologia – é um conceito que começou a ser aplicado pelo *Office of Technology Assessment* (OTA), nos Estados Unidos, em 1972, a partir da constatação de que a tecnologia muda e se expande rápida e continuamente, tendo como propósito fornecer indicações antecipadas dos benefícios prováveis ou impactos adversos das aplicações de uma tecnologia emergente (BLAIR, 1994).

A *National Science Foundation*³ define *Technology Assessment* como um estudo de políticas destinado a melhor entender as consequências para a sociedade, da expansão de tecnologias existentes ou da introdução de novas tecnologias cujos efeitos normalmente não seriam planejados ou antecipados (PORTER et al., 2004).

Atualmente, apenas poucos países europeus realizam *Technology Assessment*, sendo o termo muitas vezes associado com atividades parlamentares (JOHNSTON, 2008).

2.2. Classificação dos métodos de FTA

A lista de campos de estudo e métodos relacionados com a temática de explorar o futuro é grande e tende a crescer ainda mais e, devido a isso, é possível realizarmos diversos tipos de agrupamento dos mesmos. Algumas das classificações de métodos de FTA comuns são por: Natureza, Abordagem, Capacidade, e Famílias de Métodos. Entretanto, outras classificações podem ser realizadas, dependendo do motivo da análise dos métodos.

2.2.1. Pela sua Natureza

Os métodos de FTA tendem a diferir em abordagens e em habilidades requeridas, podendo ser classificados como: “hard” (quantitativos, empíricos, numéricos), “soft” (qualitativos baseados em julgamentos ou refletindo conhecimentos tácitos) ou semiquantitativos (POPPER, 2008). Os métodos quantitativos aplicam análises

³ A *National Science Foundation* (NSF) é uma agência federal americana independente criada em 1950 para promover o progresso da ciência, promover a saúde, prosperidade e bem-estar nacional e garantir a defesa nacional. A NSF é essencial, pois apoia pesquisas básicas e pessoas para criar conhecimento que transforma o futuro.

estatísticas em variáveis provenientes de um conjunto de dados, em sua maioria, de fontes diversas, como indicadores econômicos. Os métodos qualitativos se baseiam na percepção subjetiva e criativa de eventos e seus significados. Os métodos semiquantitativos aplicam modelos matemáticos para quantificar a subjetividade, pesando as opiniões de pessoas leigas ou especialistas.

2.2.2. Por Abordagem

Os métodos de FTA também podem ser classificados dependendo da sua abordagem com relação ao futuro como sendo “*normativos*” ou “*exploratórios*” (PORTER et al., 2004). Nos métodos normativos de FTA, o processo é iniciado com uma nítida percepção de necessidade futura. Nos métodos exploratórios de FTA, o processo é iniciado a partir extração das tecnologias atuais.

2.2.3. Pela sua Capacidade

De acordo com POPPER (2008), outra forma de classificação dos métodos de FTA é pela habilidade dos métodos em coletar ou processar informação baseada em evidência, experiência, interação ou criatividade. Esses atributos não são mutualmente exclusivos, devendo ser vistos como dimensões nos quais os métodos se apoiam. Um método, por exemplo, pode conter 25% criatividade, 25% experiência, 15% evidência, 40% interação – conforme apresentado na Figura 1. Essa classificação forma a base do *Foresight Diamond*⁴, um critério de seleção de métodos de FTA.

Os métodos baseados em evidências são aqueles que elaboram visões e definem prioridades com base em conhecimentos que normalmente já são públicos, como estatísticas ou análises das inovações. Os métodos baseados em criatividade são fortemente influenciados pela imaginação, tais como Análise de Ficção Científica e Jogos de Simulação. Os métodos baseados em experiência procuram obter pareceres informados e os dados subjacentes a esses pareceres. Por fim, as técnicas baseadas em interação envolvem a troca de conhecimento e discussões (BARBOSA, 2018).

⁴ *Foresight Diamond* é uma abordagem para escolha de métodos de FTA onde vários métodos são espalhados em forma de diamante. Onde cada canto representa uma propriedade inerente ao método de FTA – criatividade, experiência, evidência e interação (POPPER, 2008).

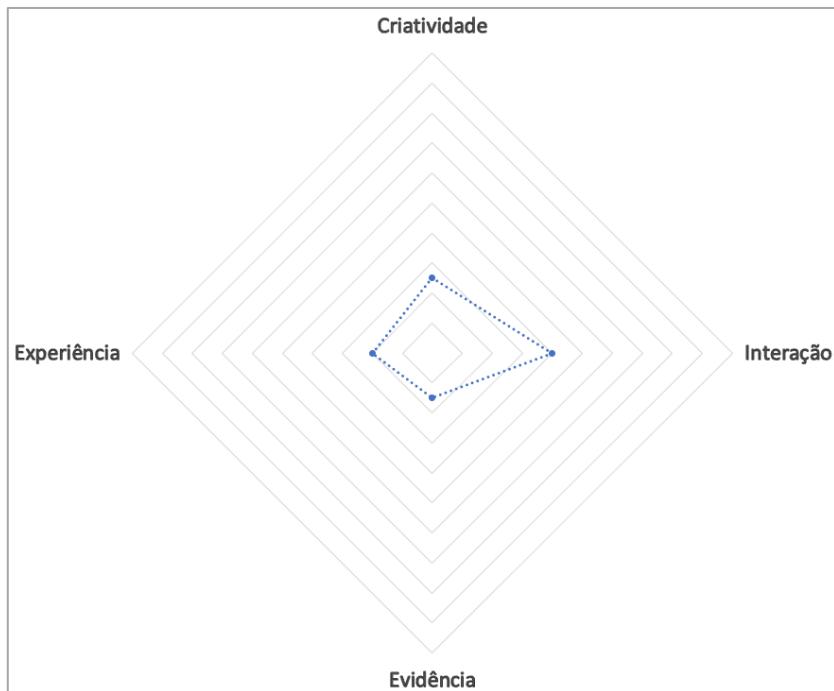


Figura 1. Exemplo de representação de um método de FTA por sua capacidade.
Adaptado de BARBOSA (2018).

2.2.4. Por Famílias de Métodos

Algumas propostas de classificação dos métodos e técnicas existentes e em uso nas atividades de prospecção tecnológica foram sugeridas por (PORTER et al., 1991, 2004), (SKUMANICH; SILBERNAGEL, 1997) e (COELHO, 2003) dividindo os métodos de prospecção em famílias.

A classificação mais recente, proposta pelo *Technology Futures Analysis Methods Working Group* (PORTER et al., 2004), identifica as seguintes (nove) famílias: Criatividade, Métodos Descritivos e Matrizes, Métodos Estatísticos, Opinião de Especialistas, Monitoramento e Sistemas de Inteligência, Modelagem e Simulação, Cenários, Análises de Tendências, e Sistemas de Avaliação e Decisão.

Um fato a destacar é que um método pode pertencer a mais de uma família. Esta seção descreve em detalhe cada uma dessas famílias de métodos.

2.2.4.1. Métodos de Criatividade

A criatividade é uma característica que deve estar presente em todos os estudos prospectivos, pois há a necessidade de se evitar visões pré-concebidas de problemas e situações e encorajar um novo padrão de percepção. É um meio de ampliar a habilidade de visualizar futuros alternativos (COELHO et al., 2004). Os métodos de Criatividade evitam visões pré-concebidas de problemas e situações e encorajam um padrão novo de

percepção, aumentando a habilidade dos pesquisadores de visualizar futuros alternativos (BARBOSA, 2018).

2.2.4.2. Métodos Descritivos e Matrizes

Métodos Descritivos e Matrizes podem ser usados para ampliar a criatividade e para possibilitar a identificação de futuros alternativos. Além disso, assim como outros métodos e técnicas, dependem da existência de especialistas, de boas séries de dados, de boas estruturas e da compreensão da modelagem e das tecnologias da informação e da comunicação (COELHO et al., 2004).

2.2.4.3. Métodos Estatísticos

Métodos Estatísticos procuram medir o efeito de uma ou mais variáveis independentes sobre o comportamento futuro de uma variável dependente. O procedimento padrão é testar modelos de ajuste (linear, exponencial, quadrado ou cúbico) para a variável dependente, procurando definir os parâmetros do modelo para minimizar o erro (COELHO et al., 2004).

2.2.4.4. Métodos de Opinião de Especialistas

Os métodos de Opinião de Especialistas têm seus limites estabelecidos naquilo que as pessoas percebem como factível, de acordo com sua imaginação e crenças, e deve ser usada sempre que a informação não puder ser quantificada ou quando os dados históricos não estão disponíveis ou não são aplicáveis. Mesmo quando há dados históricos, a opinião de especialistas pode e deve ser usada como uma forma de complementar as informações obtidas e de captação de conhecimentos tácitos, sinais fracos e *insights*. Por isso, tais métodos são considerados qualitativos (COELHO et al., 2004).

2.2.4.5. Métodos de Monitoramento e Inteligência

Os métodos de Monitoramento fornecem fontes básicas de informação relevante e por isso são comumente utilizados em estudos prospectivos. Monitorar significa observar, checar e atualizar-se em relação aos desenvolvimentos numa área de interesse, definida para uma finalidade bem específica (PORTER et al., 1991).

Alguns objetivos possíveis do monitoramento: identificar eventos científicos, técnicos ou socioeconômicos importantes; definir ameaças potenciais que são implícitas nesses eventos; identificar oportunidades envolvidas nas mudanças no ambiente; e alertar tomadores de decisão sobre tendências que estão convergindo, divergindo, ampliando, diminuindo ou interagindo (COELHO et al., 2004).

Segundo PORTER et al. (1991), no seu sentido estrito, o monitoramento não é uma técnica de prospecção. Entretanto, é amplamente utilizada porque provê os conjuntos de dados nos quais a prospecção se baseia.

2.2.4.6. Métodos de Modelagem e Simulação

Os métodos de Modelagens e Simulação se baseiam em tentativas de identificar certas variáveis e criar modelos computacionais, jogos ou sistemas nos quais se pode visualizar a interação entre as variáveis ao longo do tempo. Computadores ou pessoas, ou ambos, podem ser envolvidos. Com os computadores, pode-se fazer o jogo do “e se...”, em que a partir de determinadas escolhas podem-se ver as consequências que se seguem (COELHO et al., 2004).

2.2.4.7. Métodos de Cenários

Os métodos de construção de Cenários buscam construir representações do futuro, assim como rotas que levam até essas representações. Essas representações procuram destacar as tendências dominantes e as possibilidades de eventos disruptivos no ambiente em que estão localizadas as organizações. Cenários constituem uma forma de integração com outras informações úteis e são excelentes para comunicar resultados aos usuários em geral (COELHO et al., 2004).

2.2.4.8. Métodos de Análise de Tendências

Segundo SKUMANICH e SILBERNAGEL (1997), a Análise de Tendências é a forma mais simples de prospecção. Esse método é baseado na hipótese de que os padrões do passado serão mantidos no futuro, ou seja, parte do pressuposto de que o futuro é a continuação do passado. A análise de tendências, em geral, utiliza técnicas matemáticas e estatísticas para extrapolar séries temporais para o futuro. Coleta-se informação sobre uma determinada variável ao longo do tempo e, em seguida, essa informação é extrapolada para um ponto no futuro.

2.2.4.9. Métodos de Avaliação e Decisão

O processo de tomada de decisão busca reduzir a incerteza sobre determinadas alternativas e permitir uma escolha razoável entre o que se encontra disponível. Métodos de Avaliação e Decisão incluem o tratamento de múltiplos pontos de vista. Esses métodos permitem priorizar ou reduzir os vários fatores que devem ser levados em consideração na análise, diminuindo a incerteza sobre as alternativas disponíveis (COELHO et al., 2004).

2.3. Métodos de FTA

Variadas metodologias sistemáticas podem ser aplicadas em uma Prospecção Tecnológica. De fato, qualquer metodologia utilizada para fornecer alguma análise útil para um estudo de FTA pode ser aplicada irrestritamente. Assim, não existem limitações em relação aos métodos de FTA – somente fatores como prazo, custo e recursos disponíveis (MARTINO, 2003).

Diversos autores apontam para a importância de se incluir mais de um método ou técnica na estrutura metodológica de um exercício prospectivo, de modo a reduzir os níveis de incerteza inerentes a esse tipo de atividade, integrando diferentes abordagens e resultados, além da constatação de que nenhum método ou técnica pode atender a todas as questões envolvidas em um estudo (COELHO et al., 2004).

O FTA incorpora uma grande variedade de métodos de prospecção tecnológica que estão listados pelo *Technology Futures Analysis Methods Working Group* (PORTER et al., 2004) e representados, de forma agrupada pela classificação de famílias, na Tabela 1.

Diante dos métodos de FTA, o foco é dado para o *Roadmapping* Tecnológico, técnica que provê, como resultado do seu processo, o *roadmap*, que é o objeto principal deste trabalho. No próximo capítulo apresentaremos uma fundamentação teórica deste método prospectivo.

Tabela 1. Métodos de FTA agrupados pela classificação de famílias. Adaptado de PORTER et al. (2004).

Famílias de Métodos	Métodos
Criatividade	<i>Brainstorming (Brainwriting; NGP - Nominal Group Process), Creativity Workshops (Future Workshops), Science Fiction Analysis, TRIZ e Vision Generation,</i>
Descritivos e Matrizes	<i>Analogies, Backcasting, Checklists for Impact Identification, Innovation System Modeling, Institutional Analysis, Mitigation Analyses, Morphological Analysis, Multicriteria Decision Analyses (DEA - Data Envelopment Analysis), Multiple Perspectives Assessment, Organizational Analysis, Relevance Trees (Futures Wheel), Requirements Analysis Needs Analysis, Attribute X Technology Matrix, Risk Analysis, Roadmapping (Product-technology Roadmapping), Social Impact Assessment (Socio-Economic Impact Assessment), Stakeholder Analysis (Policy Capture, Assumptional Analysis), State of the Future Index (SOFI), Sustainability Analysis (Life Cycle Analysis) e Technology Assessment.</i>
Estatísticos	<i>Bibliometrics (Research Profiling; Patent Analysis, Text Mining), Correlation Analysis, Cross-Impact Analysis, Demographics, Risk Analysis e Trend Impact Analysis (TIA).</i>
Opinião de Especialista	<i>Delphi (iterative survey), Focus Groups (Panels; Workshops), Interviews e Participatory Techniques.</i>
Monitoramento e Inteligência	<i>Bibliometrics (Research Profiling; Patent Analysis, Text Mining) e Monitoring (Environmental Scanning, Technology Watch; Competitive Intelligence, Veille Technologique, Vigilância Tecnológica; Benchmarking).</i>
Modelagem e Simulação	<i>Agent Modeling, Causal Models CAS (Complex Adaptive System Modeling - Chaos), Cross-Impact Analysis, Diffusion Modeling, Economic Base Modeling (Input-Output Analysis), Scenario-Simulation (Gaming; Interactive Scenarios), Sustainability Analysis (Life Cycle Analysis), Systems Simulation (System Dynamics, KSIM), Technology Assessment e Technological Substitution.</i>
Cenários	<i>Field Anomaly Relaxation Methods (FAR), Scenarios (Scenarios with consistency checks; Scenario Management; La Prospective; GBN; Puma; Pítia) e Scenario-Simulation (Gaming; Interactive Scenarios).</i>
Análise de Tendências	<i>Long Wave Analysis, Precursor Analysis, Trend Extrapolation (Growth Curve Fitting & Projection) e Trend Impact Analysis.</i>
Avaliação e Decisão	<i>Action (Options) Analysis, Multicriteria Decision Analyses (DEA - Data Envelopment Analysis), Analytical Hierarchy Process (AHP), Cost-Benefit Analysis (Monetized & Other), Decision Analysis (Utility Analysis), Economic Base Modeling (Input -Output Analysis), Relevance Trees (Futures Wheel), Requirements Analysis (Needs Analysis, Attribute X Technology Matrix) Stakeholder Analysis (Policy Capture) e Benchmarking.</i>

Capítulo 3 – *Roadmapping* Tecnológico

As raízes do *Roadmapping* Tecnológico – do inglês, *Technology Roadmapping*, ou simplesmente *Roadmapping*, são identificadas na abordagem da indústria automotiva americana que, seguida pela *Motorola* e a *Corning*, adotou processos sistemáticos de *roadmaps* ao final dos anos 70 e começo dos 80. A abordagem da *Motorola* foi mais perceptível, levando o conceito ao setor de eletrônica, notadamente através da *Philips*, da *Lucent Technologies* e da *Semiconductor Industry Association* (SIA) (PHAAL; FARRUKH; PROBERT, 2004).

Roadmapping Tecnológico pode ser definido como um processo de planejamento impulsionado pela necessidade de tecnologias, que ajuda a identificar, selecionar e desenvolver tecnologias variadas para atender um determinado conjunto de carências ou de produtos já definidos. É usual se colocar juntos um grupo de especialistas para coletar informações, desenvolver, organizar e apresentar um planejamento para a melhor tomada de decisão sobre os melhores e mais rentáveis investimentos (COELHO et al., 2005).

De acordo com PHAAL et al. (2004), *Roadmapping* é um método flexível e extensivamente utilizado para apoiar a gestão e o planejamento tecnológico de uma organização. O seu uso mais habitual é para apoiar o planejamento de produtos, onde é utilizado para conectar as novas tecnologias com os produtos reais.

Roadmapping Tecnológico também pode ser considerado como uma poderosa técnica para apoiar a gestão tecnológica e o planejamento, especialmente para explorar e comunicar os elos dinâmicos que existem entre recursos tecnológicos, objetivos organizacionais e mudanças no ambiente externo (RINNE, 2004).

Este método tem sido mais impulsionado pelo mercado (*market pull*), isto é, pelas inovações tecnológicas necessárias para as empresas atenderem a mercados futuros, do que impulsionado pela tecnologia em si mesma (*technology push*). O que se busca construir é uma visão de futuro (onde a empresa pretende chegar) e quais são as tecnologias necessárias para se chegar até lá. Por outro lado, fornece roteiros, caminhos para se atingir a visão de futuro, etapa por etapa, auxiliando as empresas e organizações a identificar, selecionar e desenvolver as alternativas tecnológicas corretas e necessárias e criar os produtos adequados para os mercados futuros (COELHO et al., 2005).

O produto gerado ao fim de um processo de TRM é o *roadmap*, que busca auxiliar as empresas a subsistir em situações subversivas, fornecendo um foco para monitorar o ambiente e meios de acompanhar o desempenho de tecnologias, incluindo aquelas

potencialmente disruptivas. Em sua forma mais básica, pode ser visto como uma análise temporal – geralmente envolvendo elementos gráficos com múltiplas dimensões de ligação: com o mundo externo (mercado), com o portfólio da organização (produto), e com a evolução tecnológica (tecnologia), para garantir o sincronismo entre essas partes com os objetivos estratégicos da organização (PHAAL; FARRUKH; PROBERT, 2004).

Conforme GARCIA e BRAY (1997), *roadmaps* são relevantes para a comunicação entre os integrantes de uma equipe, apoiando a análise dos custos e facilitando o relacionamento de questões em nível estratégico com o planejamento e a execução em nível operacional.

3.1. Por que Elaborar *Roadmaps*?

O processo de elaboração de um *roadmap* é complicado quando a decisão sobre o investimento é complexa. O êxito do desenvolvimento de produtos e a gestão desse processo demandam equipes capazes de gerenciar as complexidades e de se produzir uma série de produtos, com custos e qualidade convenientes, usando as tecnologias mais favoráveis.

Roadmaps propiciam a criação de um plano que agrupa: as necessidades do mercado e do consumidor; o avanço do produto; e a introdução de novas tecnologias logo no início do processo. Dessa forma, auxiliam na identificação e solução das lacunas antevistas, atuando como um guia durante todo o caminho e orientando os tomadores de decisão, os fornecedores, os parceiros e os consumidores (COELHO et al., 2005).

Segundo o Albright Strategy Group (ALBRIGHT, 2005), há basicamente três grandes tópicos sobre os quais se realizam estudos de *roadmaps*: ciência e tecnologia; indústrias e governos; e, produtos tecnológicos e plataformas. O estudo de ALBRIGHT (2005) ainda apresenta dez razões para se elaborar um *roadmap*:

1. *Roadmaps* são bons exercícios de **planejamento**. São processos que propiciam o estudo completo do potencial das estratégias competitivas e apontam caminhos para sua implementação. As decisões tecnológicas são incorporadas como parte integral do plano, desde seu início e não apenas como elemento posterior;
2. *Roadmaps* incorporam o **tempo** de maneira evidente. Isso auxilia na identificação das tecnologias e capacidade para utilizar elas em um estipulado período;

3. *Roadmaps relacionam* estratégias de negócios e dados de mercado com decisões sobre produtos tecnológicos;
4. *Roadmaps* revelam **lacunas** nos planos para desenvolvimento de produtos e tecnologias. Reconhecem áreas onde se evidencia a necessidade de uma atuação ágil, antes que se constituam problemas reais, de forma a alcançar os objetivos e soluções desejadas;
5. *Roadmaps* auxiliam na **priorização** dos investimentos com base em tendências fortes. A cada estágio do processo de *Roadmapping*, o foco se torna mais delineado em torno de elementos mais pertinentes. Os mapas, uma vez elaborados, são expostos aos tomadores de decisão que, por sua vez, estarão equipados para realizar suas escolhas entre os objetivos da corporação;
6. *Roadmaps organizam* um conjunto mais realista de objetivos, considerando a natureza da competitividade do setor ou indústria;
7. *Roadmaps* podem ser considerados como **guias** ou **manuais**, permitindo à equipe identificar e atuar em eventos que requerem mudanças de direção. Parte do processo de desenvolvimento de um *roadmap* é a criação de um mapa de riscos, identificando eventos ou mudanças em condições críticas que sinalizam a necessidade de reavaliar ou rever o plano durante sua execução;
8. O **compartilhamento** de *roadmaps* possibilita o uso estratégico das tecnologias através de variadas linhas de produtos. *Roadmaps* cruzados podem ressaltar necessidades comuns, capacidades que precisam ser mediadas, custos de desenvolvimento que podem ser compartilhados, e ainda podem constituir para a organização uma base de dados contendo tecnologias disponíveis e necessidades tecnológicas;
9. *Roadmaps* proporcionam a **comunicação** entre negócios, planos e produtos tecnológicos a toda a comunidade interessada;
10. Por último, *roadmaps* **constroem** equipes de desenvolvimento. O processo de *Roadmapping* permite e requer um entendimento comum entre os financiadores, gestores e responsáveis pela implementação do plano, agregando ideias e pensamentos dos envolvidos no processo, caracterizando-o como altamente participativo, contudo, estruturado.

3.2. Metodologias de *Roadmaps*

Na literatura existem diversos modelos e metodologias para se elaborar um *roadmap*. Contudo, há um entendimento de que algumas questões básicas devem ser levadas em consideração para, efetivamente, dar início a um processo de TRM (COELHO et al., 2005):

- A necessidade de se realizar a integração a partir do *market pull* (necessidades observadas do mercado), *technology push* (incentivo tecnológico) e *competitive clash* (disputa entre concorrentes) nos *roadmaps* e nas atividades de planejamento;
- A avaliação dos fatores de impacto perceptíveis e imperceptíveis que podem afigir o sucesso dos *roadmaps* em longo prazo;
- A análise do impacto dos eventos externos na implementação do plano estratégico;
- As estratégias chave para registrar o comportamento das partes externas e integrá-las ao processo;
- A garantia do apoio de especialistas internos e externos à organização e a definição dos papéis e atividades no processo, de forma a garantir o bom desenvolvimento do TRM.

De acordo com PHAAL et al. (2004), um *roadmap* pode ser elaborado sob duas perspectivas:

1. **Empresa:** permite que o desenvolvimento tecnológico seja incorporado ao plano de negócios e que o impacto de novas tecnologias e desenvolvimentos de mercado seja analisado;
2. **Multiorganizacional:** descobre tendências do ambiente, ameaças e oportunidades para um grupo específico de *stakeholders* em relação a uma tecnologia ou área de aplicação. *Roadmaps* incorporam o tempo de maneira evidente. Isso auxilia na identificação das tecnologias e capacidade para utilizar elas em um estipulado período.

PHAAL et al. (2004) também define, como ilustrado na

Figura 2, que o modelo mais comum e mais utilizado de *roadmap* é aquele estruturado em três níveis ou camadas (mercado, produto e tecnologia) – proposta genérica formulada pela *European Industrial Research Management Association*

(EIRMA, 1997). Esses níveis contribuem para a diferenciação dos elementos constantes do tema que está sendo estudado. Além disso, o horizonte temporal é essencial para o desenvolvimento de atividades e para o acompanhamento do processo.

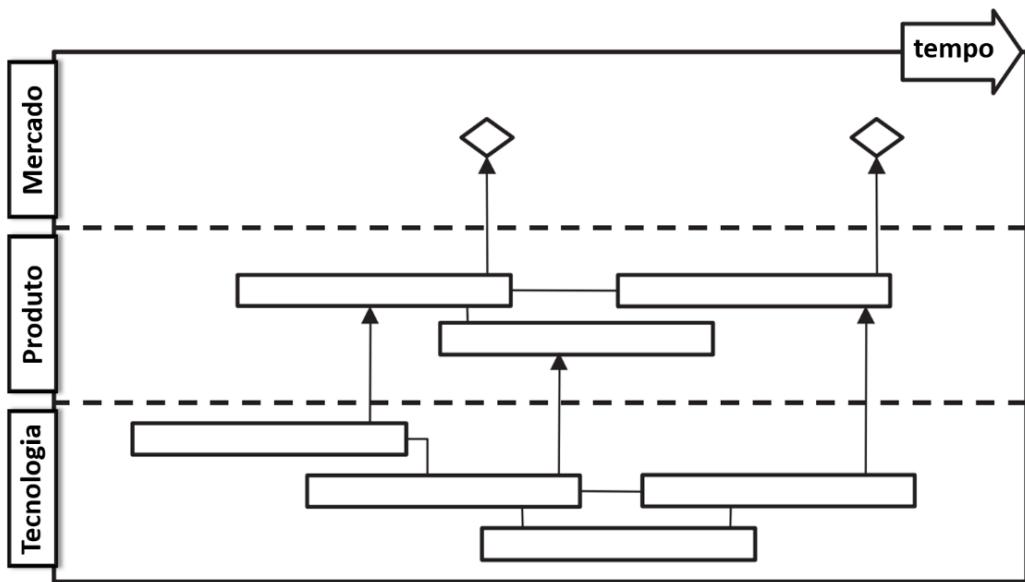


Figura 2. Modelo de um *roadmap*, adaptado de PHAAL et al. (2004).

Segundo outro ponto de vista, o processo de *Roadmapping* pode, também, ser elaborado em três fases distintas, conforme o estudo de GARCIA e BRAY (1997).

1. A primeira fase consta de atividades preliminares, o que inclui **satisfazer as condições essenciais**. *Roadmaps* devem ser elaborados a partir de um conjunto de necessidades. É o objetivo de uso que define o horizonte temporal e o nível de detalhes que se deseja. É um processo que necessita ser impulsionado pela necessidade percebida (*needs driven*) e não pela solução proposta (*solution-driven*). Essa fase garante que o contexto para o *roadmap* seja especificado. Deve-se identificar porque um *roadmap* é útil e como será utilizado. Também é fundamental se definir com nitidez o escopo, o limite ou fronteira do tema de estudo;
2. A segunda fase está direcionada para o **desenvolvimento efetivo do *roadmap*** e compreende a identificação do que será o foco do estudo; a identificação dos requisitos cruciais do sistema e seus alvos; a especificação das principais áreas tecnológicas; a especificação das forças motrizes da tecnologia (*technology drivers*) e suas metas; o reconhecimento de alternativas tecnológicas e prazos de implementação; a sugestão de tecnologias optativas que deverão ser buscadas; e, a criação do *roadmap* em si.

3. A terceira fase inclui as atividades de **continuidade** que buscam as críticas e a validação do *roadmap*; apresentam uma avaliação das tecnologias alternativas referenciadas e os objetivos a serem logrados; o desenvolvimento do plano de implementação; e a revisão e atualização do estudo.

Sob ainda outra perspectiva, para o Albright Strategy Group (ALBRIGHT, 2005), *roadmaps* precisos seguem uma característica em comum e visam guiar os desenvolvedores e seguidores a pontos de decisão cruciais. Esse formato deve obter a clareza do objetivo e das perguntas críticas: “por quê” (*why*), “o quê” (*what*), “como” (*how*) e “quando” (*when*), que devem esclarecer claramente o plano de ação para alcançar os objetivos almejados além de listar as ações a serem realizadas. A Figura 3 representa esse modelo de uma arquitetura comum de TRM.

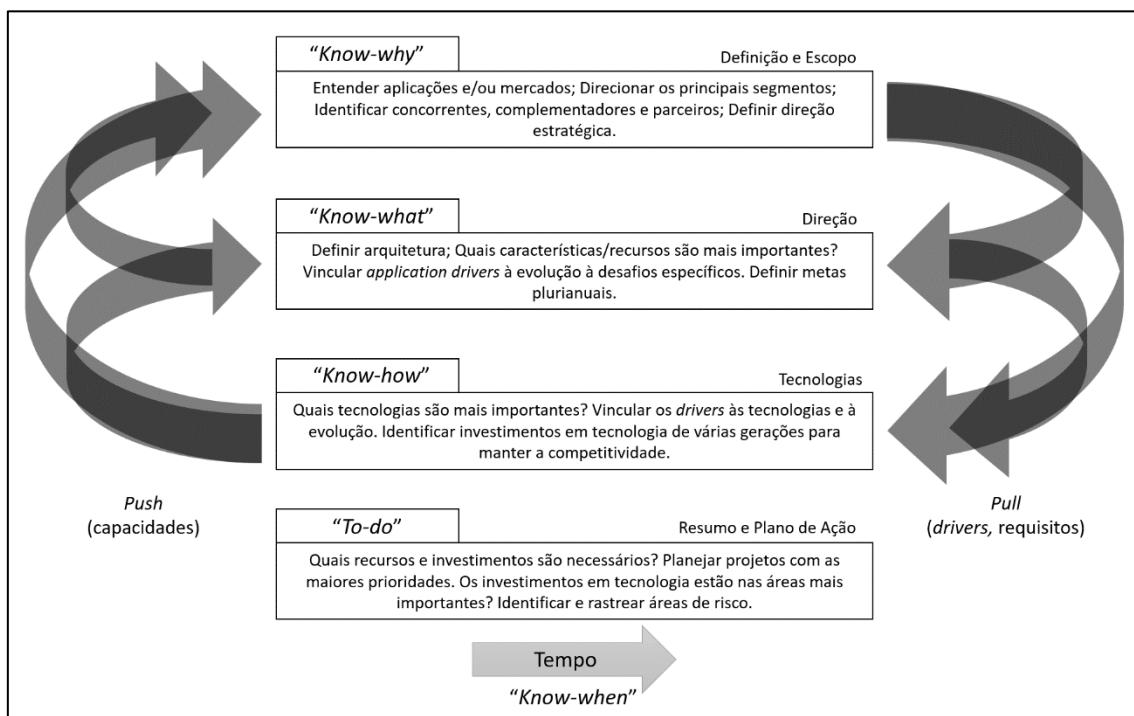


Figura 3. Formato característico de um *roadmap*. Adaptado de (ALBRIGHT, 2005).

- **“Know-why”**: A primeira etapa descreve o domínio do *roadmap*, o escopo, os objetivos, a estratégia para atingir os objetivos – este é o “por quê” do *roadmap*;
- **“Know-what”**: A segunda é “o quê” do *roadmap*, onde o direcionamento é definido. Incluindo desafios, a arquitetura do processo e sua evolução e metas a serem obtidas;

- “**Know-how**”: Essa etapa define a evolução das tecnologias que serão utilizadas para alcançar os objetivos – é o “como” do *roadmap*.
- “**To-do**”: A última etapa define o plano de ação e os riscos – lista as ações do que fazer. O plano de ação identifica os desenvolvimentos, os recursos fundamentais, os riscos relacionados e as estratégias de desenvolvimento tecnológico.

Roadmaps, em geral, possuem uma dimensão temporal, e é esse fator que constitui o “quando” (*when*) – evidenciando a sua dependência do tempo. Um *roadmap* pode ser construído iniciando com as necessidades chave do mercado e dos consumidores – uma perspectiva de *market pull*. Por outro lado, é possível, também, iniciar-se com uma tecnologia chave e buscar definir as necessidades de mercado que podem utilizar a nova tecnologia – em uma perspectiva *technology push*.

Na Tabela 2, listamos os três grandes tópicos sobre quais se dividem os *roadmaps*, conforme ALBRIGHT (2005), sob a perspectiva “*why-what-how-when*”, contendo os tópicos, conteúdos e objetivos dos *roadmaps*.

Tabela 2. Segmentos dos *roadmaps* sob a perspectiva “*why-what-how-when*”, listando seus objetivos. Adaptado de ALBRIGHT (2005).

	Definição e Escopo “Know-why”	Direção “Know-what”	Tecnologias “Know-how”	Plano de Ação “To-do”
Ciência e Tecnologia	<ul style="list-style-type: none"> • Escopo • Aplicações tecnológicas 	<ul style="list-style-type: none"> • Desafios técnicos • Arquitetura • Tendências, descontinuidades e objetivos 	<ul style="list-style-type: none"> • Elementos tecnológicos e evolução • Competitividade tecnológica e custos 	<ul style="list-style-type: none"> • Programas de ação • Investimento tecnológico • Riscos
Indústria e Governo	<ul style="list-style-type: none"> • Estrutura e direção da indústria • <i>Drivers</i> dos clientes 	<ul style="list-style-type: none"> • Desafios técnicos • Arquitetura • Tendências, disruptividades • Aprendizado e alvo 	<ul style="list-style-type: none"> • Elementos tecnológicos e evolução • Alternativas tecnológicas • Custos futuros 	<ul style="list-style-type: none"> • Programas de ação • Investimento tecnológico • Riscos
Produtos Tecnológicos e Plataformas	<ul style="list-style-type: none"> • Estrutura e tamanho do mercado • <i>Drivers</i> dos clientes • Estratégia de competitividade 	<ul style="list-style-type: none"> • <i>Roadmap</i> do produto • Arquitetura • <i>Drivers</i> e alvos do produto • Evolução de funcionalidades 	<ul style="list-style-type: none"> • Elementos tecnológicos e evolução • Posição de competitividade • Custo do alvo 	<ul style="list-style-type: none"> • Programas de ação • Investimento tecnológico • Riscos

3.3. Tipologia de *Roadmaps*

É de suma importância ter definido o objetivo principal da prospecção para a melhor escolha do tipo de *roadmap* a ser buscado. Diversos tipos e abordagens sobre *roadmaps* são encontrados na literatura, podendo assumir diferentes propósitos e formatos – sendo o mais comum apresentado na Figura 2.

O estudo de PHAAL et al. (2004) propõe uma caracterização de TRM genérico (Figura 4) após a realização de uma análise de, aproximadamente, 40 *roadmaps* que forneceu uma gama de tipos, sendo esses agrupados em 16 grandes áreas. Essas áreas se referem a propósitos e formatos gráficos, conforme as observações em suas estruturas e conteúdo.

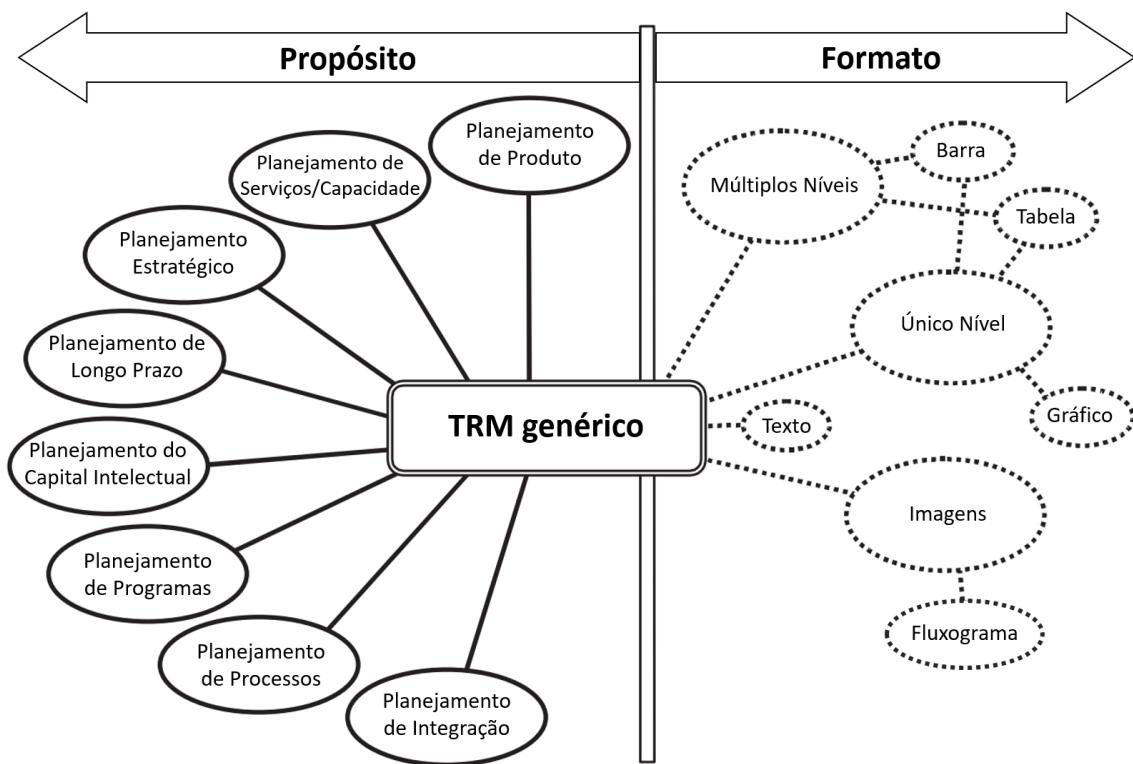


Figura 4. Caracterização de um TRM genérico – propósitos e formatos. Adaptado de PHAAL et al. (2004).

3.3.1. Propósitos

Nessa caracterização, oito propósitos de *roadmaps* que foram identificados e são detalhados a seguir, de acordo com PHAAL et al. (2004):

3.3.1.1. Planejamento de Produto

É o tipo mais usual de *roadmap*, referente à inclusão da tecnologia em produtos manufaturados, em sua maior parte incluindo mais de uma geração de produtos. A Figura

5 ilustra um *roadmap* da *Philips* onde este modelo foi adotado⁵. O exemplo demonstra como *roadmaps* são utilizados para unir planejamento tecnológico e desenvolvimento de produtos.

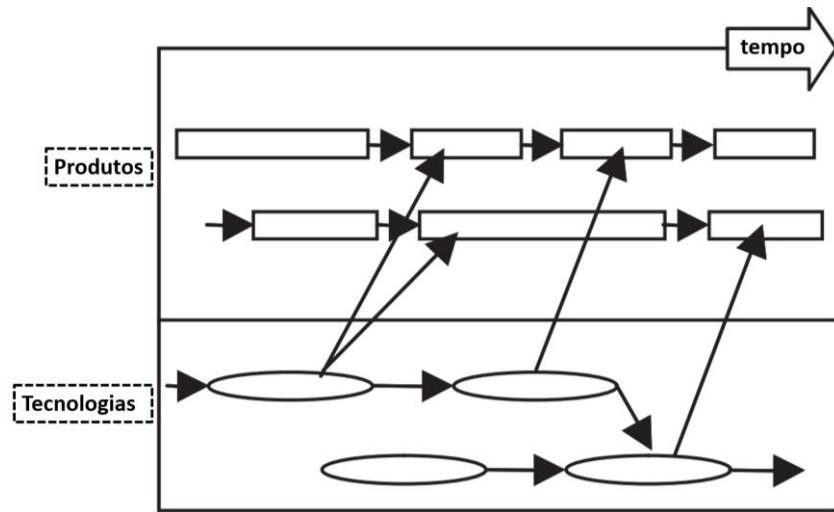


Figura 5. *Roadmap* de planejamento de produto – realizado pela *Philips*. Adaptado de PHAAL et al. (2004).

3.3.1.2. Planejamento de Serviços/Capacidades

Tem seu foco em como a tecnologia dá suporte a capacidade organizacional, sendo assim, indicado para empresas que prestam serviços. A Figura 6 ilustra o *Royal Mail roadmap*⁶, baseado em uma aplicação inicial de planejamento tecnológico, utilizado para investigar o impacto do desenvolvimento tecnológico nos negócios. Se notabiliza mais por enfatizar as capacidades organizacionais como sendo a ponte entre a tecnologia e os negócios, mais do que os produtos em si.

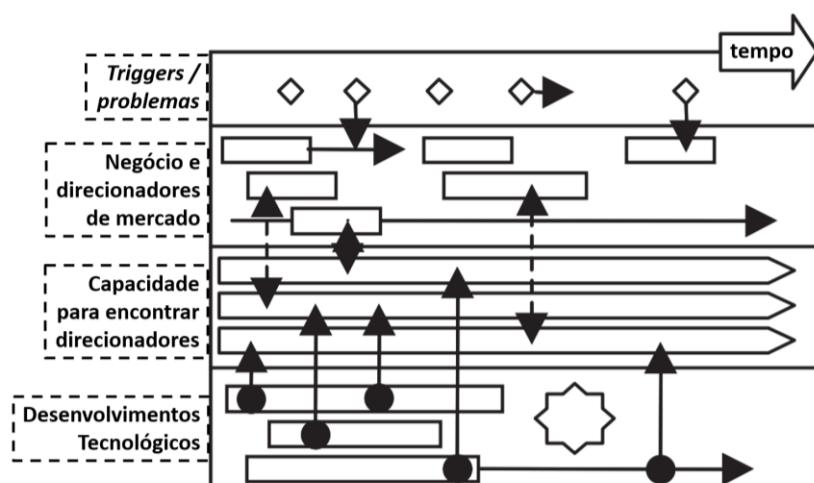


Figura 6. *Roadmap* de planejamento de serviços/capacidade – *Royal Mail*. Adaptado de PHAAL et al. (2004).

⁵ Extraído de “*Roadmapping integrates business and technology*” (GROENVELD, 2007).

⁶ Extraído da IMechE Mail Technology Conference (MTC), de BROWN e PHAAL (2001).

3.3.1.3. Planejamento Estratégico

Esse tipo de *roadmap* é apropriado para avaliação da estratégia global, em relação à análise de diversas oportunidades ou ameaças, particularmente no nível dos negócios. A Figura 7 apresenta o *roadmap* com a visão dos negócios futuros, em termos de mercado, negócios, produtos, tecnologias, competências, cultura, entre outros. Com isso, *gaps* são identificados pela comparação da visão de futuro com a posição atual e as opções estratégicas exploram as pontes entre os *gaps*.

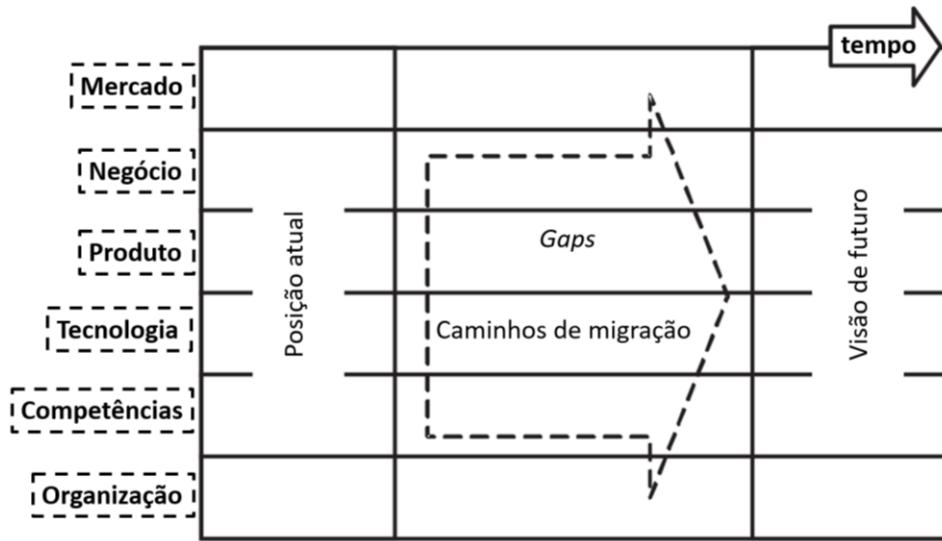


Figura 7. *Roadmap* de planejamento estratégico. Adaptado de PHAAL et al. (2004).

3.3.1.4. Planejamento de Longo Prazo

Expande o horizonte de planejamento, apoiando o planejamento de longo prazo. Este tipo de *roadmap* é, frequentemente, praticado em nível setorial ou nacional (*foresight*) e pode constituir um radar para a organização descobrir tecnologias e mercados potencialmente disruptivos. A Figura 8 ilustra um, de uma série de *roadmaps* desenvolvidos pela *U.S. Integrated Manufacturing Technology Roadmapping Initiative*⁷. Esse exemplo destaca sistemas de informação, apresentando como os desenvolvimentos tecnológicos devem convergir para uma empresa perfeita orientada a informações (“*information-driven seamless enterprise*” – um “*Nugget*”).

⁷ IMTR - *Integrated Manufacturing Technology Roadmapping Project* (1999).

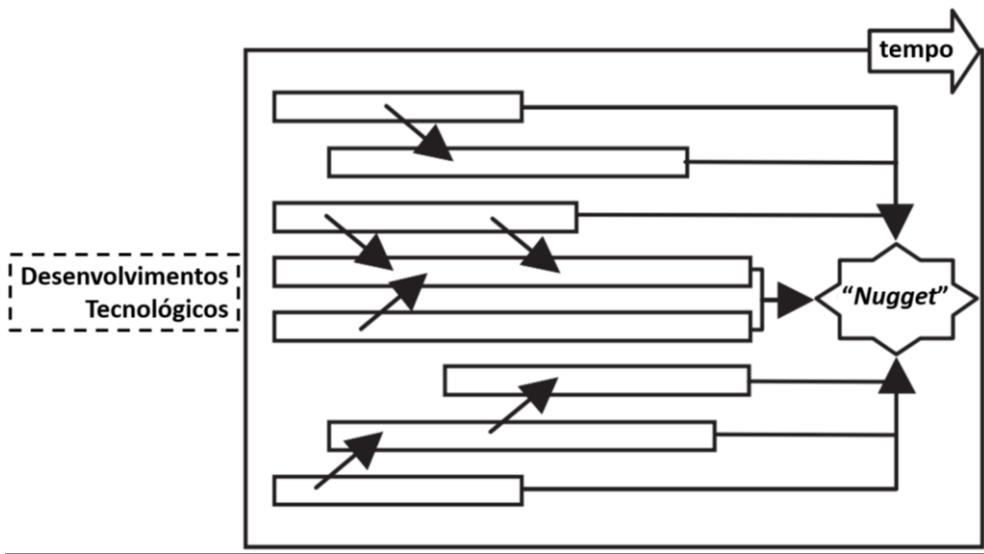


Figura 8. *Roadmap* de planejamento de longo prazo – realizado pela *U.S. Integrated Manufacturing Technology Roadmapping Initiative*. Adaptado de PHAAL et al. (2004).

3.3.1.5. Planejamento de Capital Intelectual

Esse tipo de *roadmap* estrutura iniciativas relativas ao capital intelectual e gestão do conhecimento com os objetivos do negócio. A Figura 9 mostra um exemplo desenvolvido pela *Artificial Intelligence Applications Unit - University of Edinburgh*⁸, possibilitando às organizações visualizarem os seus conhecimentos críticos e as relações entre competências, tecnológicas e habilidades necessárias para atender às demandas dos mercados futuros.

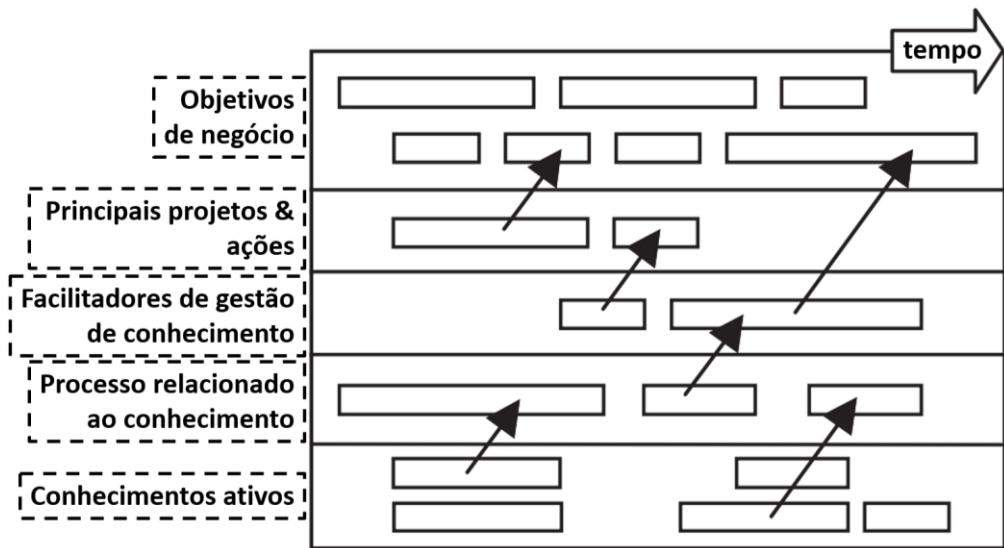


Figura 9. *Roadmap* de planejamento do capital intelectual – realizado pela *Artificial Intelligence Applications Unit - University of Edinburgh*. Adaptado de PHAAL et al. (2004).

⁸ Extraído de “*Knowledge asset roadmaps*” (MACINTOSH; FILBY; TATE, 1998).

3.3.1.6. Planejamento de Programas

Esse tipo de *roadmap* enfoca a implementação de estratégias e, de modo mais direto, o planejamento de projetos (programas de P&D). A Figura 10 ilustra um *roadmap* da *National Aeronautics and Space Administration* (NASA) para o programa *Origins*, utilizado para estudar como o universo e a vida se desenvolveram ao longo dos anos.

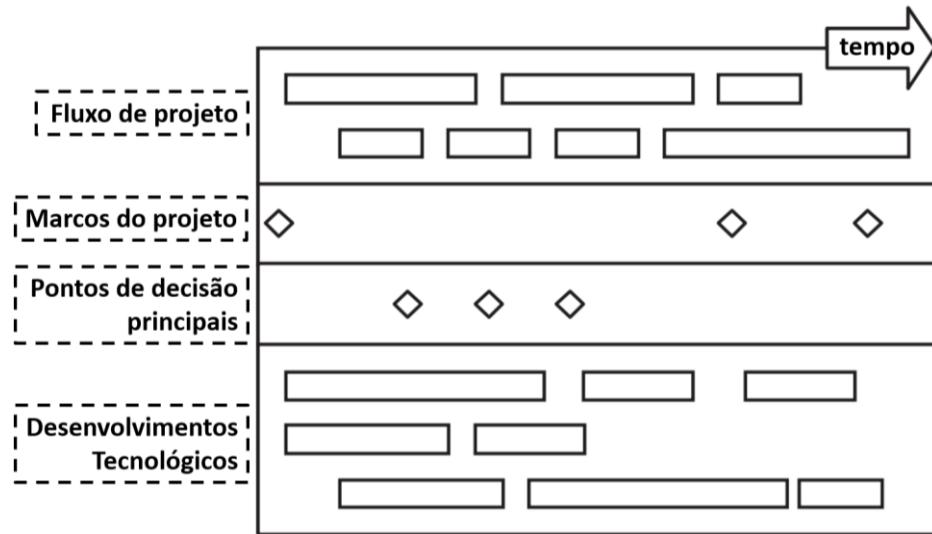


Figura 10. *Roadmap* de planejamento de programas – realizado pela NASA para o projeto *Origins*. Adaptado de PHAAL et al. (2004).

3.3.1.7. Planejamento de Processos

Esse tipo visa apoiar a Gestão do Conhecimento (GC), destacando um processo em particular (por exemplo, o desenvolvimento de um novo produto). Um *roadmap* desenvolvido para apoiar o planejamento de produto é ilustrado na Figura 11, este enfoca os fluxos de conhecimento necessários para simplificar o desenvolvimento e introdução do novo produto, agregando perspectivas comerciais e técnicas.

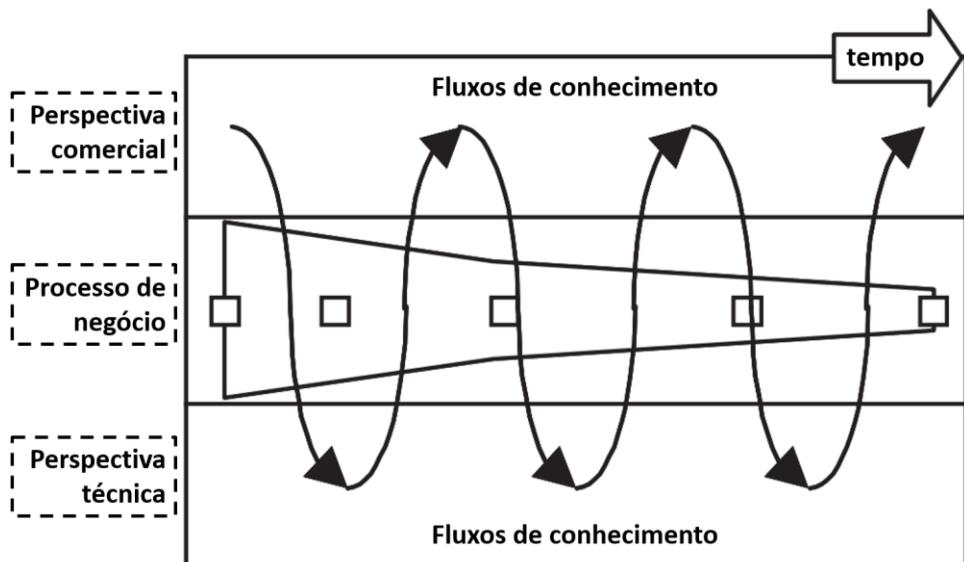


Figura 11. *Roadmap* de planejamento de processos. Adaptado de PHAAL et al. (2004).

3.3.1.8. Planejamento de Integração

Destaca a integração e/ou evolução da tecnologia no que se diz respeito a como diferentes tecnologias combinam com produtos e sistemas, ou como formam novas tecnologias (usualmente sem mostrar, de forma explícita, a dimensão tempo). A Figura 12 mostra um *roadmap* da NASA, relacionado à gestão de um programa de desenvolvimento, ilustrando o fluxo da tecnologia, como ela se enquadra em sistemas de teste e de demonstração, para dar apoio a missões científicas a serem realizadas.

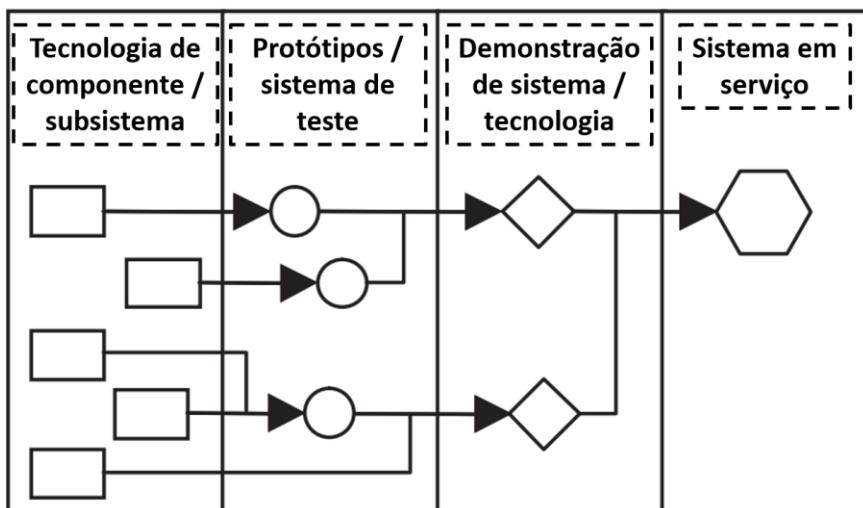


Figura 12. *Roadmap* de planejamento de integração – realizado pela NASA. Adaptado de PHAAL et al. (2004).

3.3.2. Formatos

Quanto aos formatos, foram identificados oito diferentes em uso pelas empresas que serão descritos a seguir, de acordo com PHAAL et al. (2004).

3.3.2.1. Múltiplos Níveis

Esse é o formato mais comum de roteiro de um *roadmap*, compreendendo diversos níveis (e subníveis), no âmbito de tecnologia, produto e mercado. Permite que a evolução em cada nível – ou camada – seja explorado, juntamente com as dependências entre eles, facilitando a integração da tecnologia em produtos, serviços e sistemas de negócios. A Figura 13 mostra um *roadmap* desenvolvido pela Philips, usado para apoiar a integração de tecnologias de produtos e processos no desenvolvimento de funcionalidades de produtos futuros⁹.

⁹ Extraído de “Roadmapping integrates business and technology” (GROENVELD, 2007).

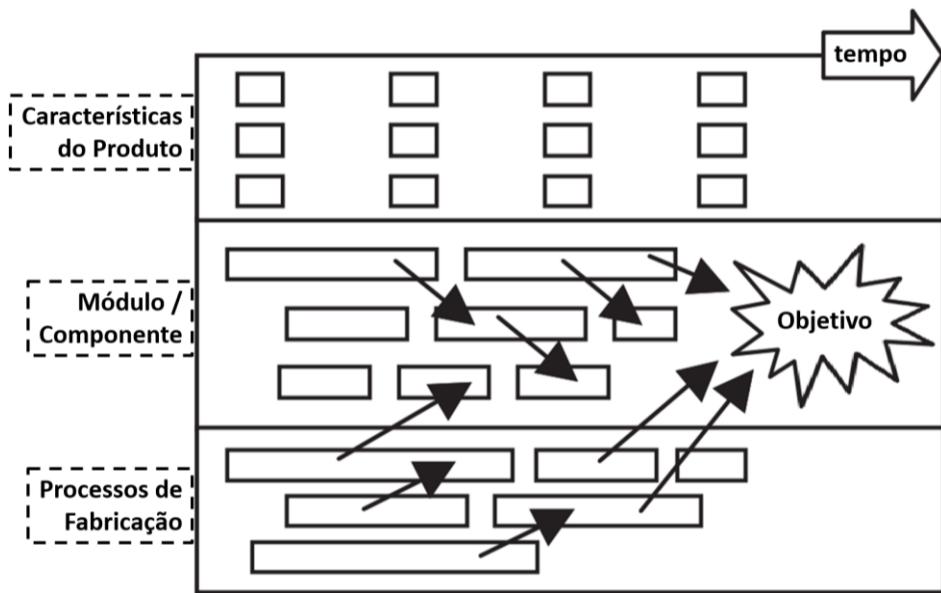


Figura 13. *Roadmap* de múltiplos níveis – realizado pela *Philips*. Adaptado de PHAAL et al. (2004).

3.3.2.2. Barra

Muitos *roadmaps* são expressos na forma de um conjunto de barras, para cada nível ou subnível. Isso tem a vantagem de simplificar e unificar os resultados, o que facilita a comunicação, a integração de *roadmaps* e o desenvolvimento de software para dar suporte ao TRM. A Figura 14 ilustra um *roadmap* da *Motorola*, relacionado à evolução dos recursos e tecnologias de aparelhos de som automotivos¹⁰.

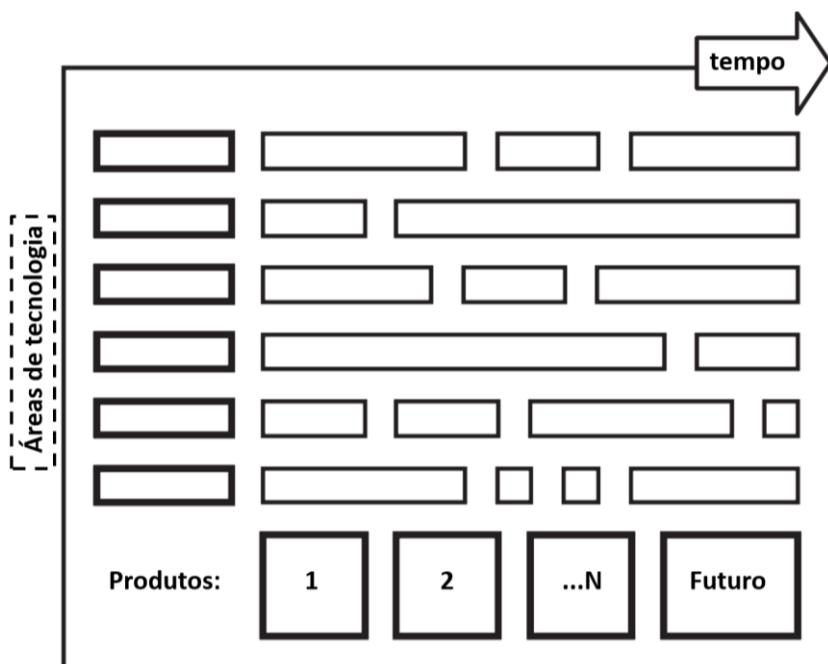


Figura 14. *Roadmap* em forma de barra – realizado pela *Motorola*. Adaptado de PHAAL et al. (2004).

¹⁰ Extraído de “*Motorola's technology roadmap process*” (WILLYARD; MCCLEES, 1987).

3.3.2.3. Tabelas

Em alguns casos, *roadmaps* completos, ou níveis do *roadmap*, são expressos em tabelas (tempo versus desempenho ou requisitos). Esse tipo de abordagem é adequado para situações em que o desempenho pode ser facilmente quantificado ou se as atividades são agrupadas em períodos específicos. A Figura 15 mostra um *roadmap* tabular, apresentando as dimensões de desempenho do produto e da tecnologia¹¹.

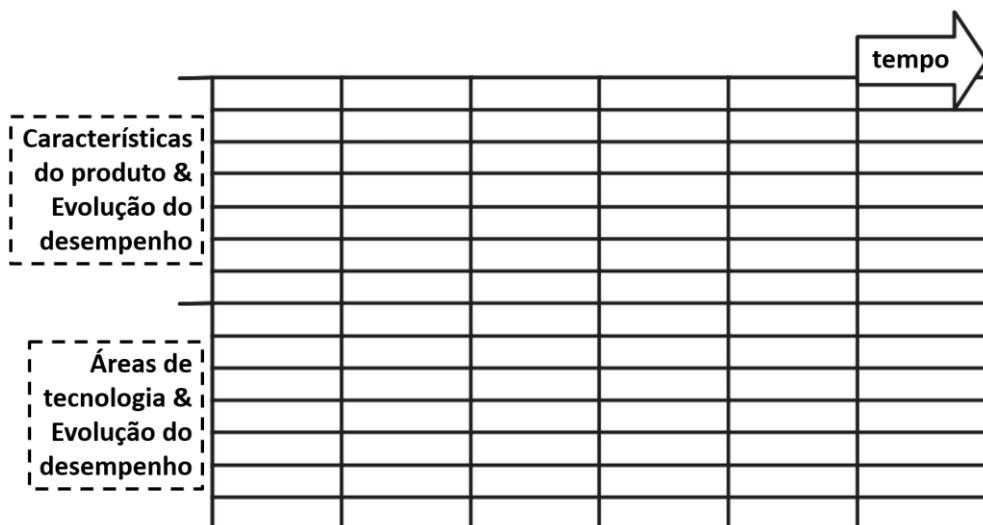


Figura 15. *Roadmap* em forma de tabela. Adaptado de PHAAL et al. (2004).

3.3.2.4. Gráficos

Um *roadmap* pode ser representado como um simples gráfico ou plotagem – geralmente um para cada subnível – de modo que o desempenho do produto ou da tecnologia possa ser quantificado. Ocasionalmente, esse tipo de gráfico é chamado de curva de experiência e está relacionado às curvas S¹² da tecnologia. A Figura 16 apresenta um *roadmap*, em forma de gráfico, mostrando como os produtos e as tecnologias coevoluem¹³.

¹¹ Extraído de “Technology roadmapping – delivering business vision” (EIRMA, 1997).

¹² A curva em S, ou curva S, como também é conhecida, é uma forma muito eficiente de apresentar informações gerenciais para o acompanhamento de projetos. A sua representação gráfica propicia o claro contraste entre os desvios daquilo que foi planejado em comparação com o realizado, de forma simultânea.

¹³ Extraído de “Technology roadmapping – delivering business vision” (EIRMA, 1997).

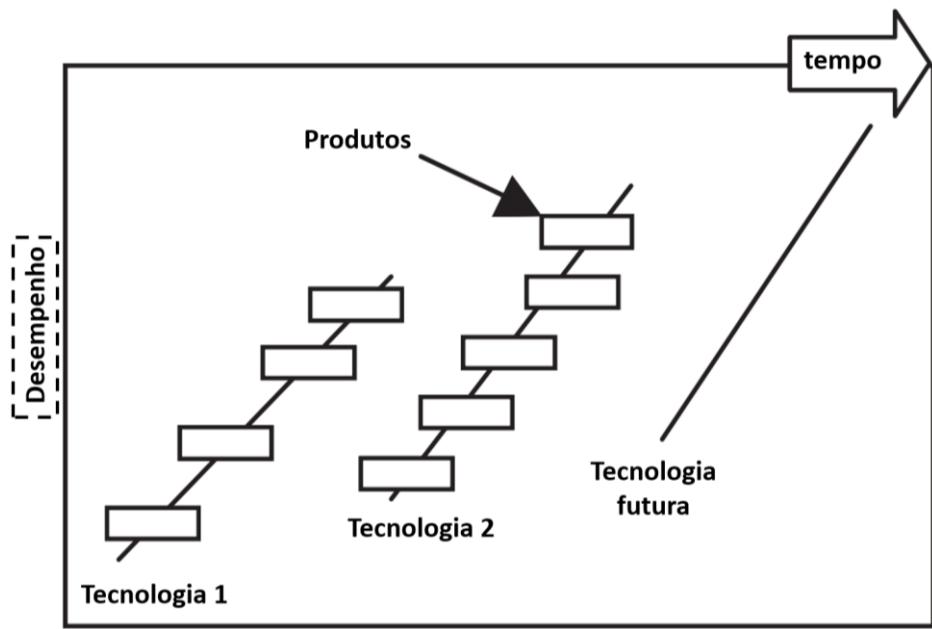


Figura 16. *Roadmap* em forma de gráfico. Adaptado de PHAAL et al. (2004).

3.3.2.5. Imagens

Alguns *roadmaps* usam representações por imagens – ou pictóricas – mais criativas para comunicar planos e integração de tecnologia. Metáforas podem ser usadas para apoiar o objetivo principal (por exemplo, uma árvore). A Figura 17 mostra um *roadmap* da *Sharp*, relacionado ao desenvolvimento de produtos e famílias de produtos, com base em um conjunto de tecnologias de *display* de cristal líquido¹⁴.

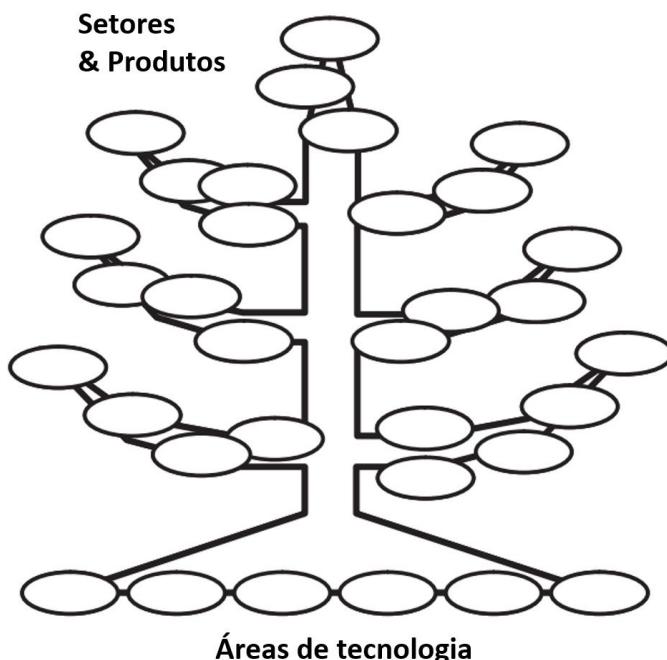


Figura 17. *Roadmap* em forma imagens – realizado pela *Sharp*. Adaptado de PHAAL et al. (2004).

¹⁴ Extraído de “*Electronic Manufacturing and Packaging in Japan*” (ITRI, 1995).

3.3.2.6. Fluxograma

Um tipo específico de representação pictórica é o fluxograma, que normalmente é usado para relacionar objetivos, ações e resultados. A Figura 18 ilustra um *roadmap* da NASA, ilustrando como a visão da organização pode estar relacionada à sua missão, questões científicas fundamentais, áreas de negócios principais, objetivos de curto, médio e longo prazo e contribuição para as prioridades nacionais dos EUA.

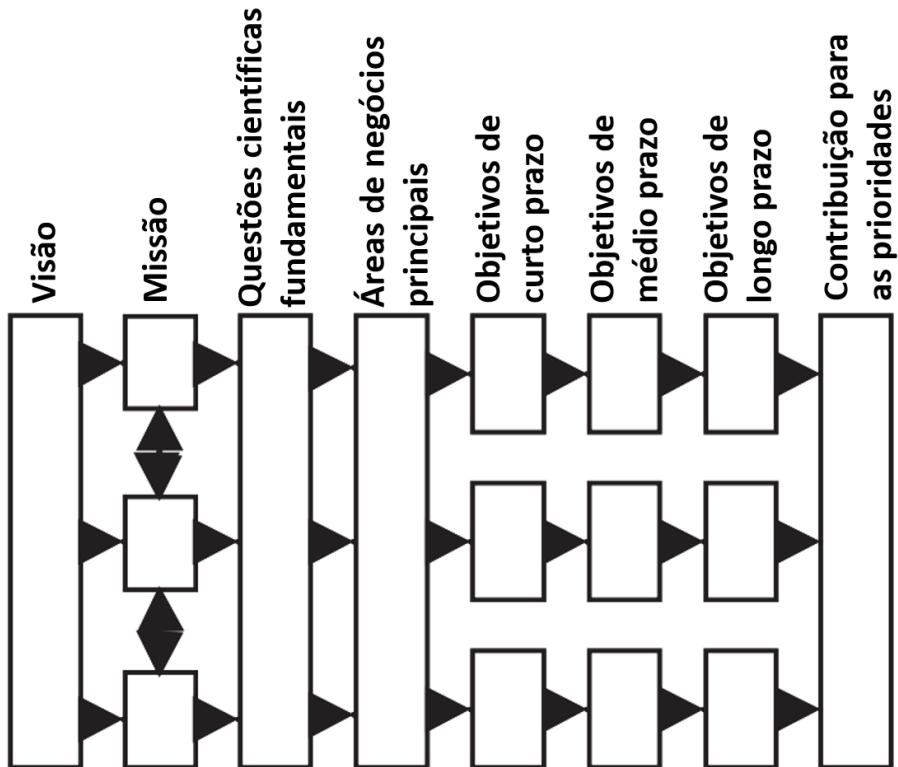


Figura 18. *Roadmap* em forma de fluxograma – realizado pela NASA. Adaptado de PHAAL et al. (2004).

3.3.2.7. Único Nível

Esse formato é um subconjunto do Múltiplos Níveis, focado em apenas um único nível do *roadmap* multinível. Embora menos complexa, a desvantagem desse tipo é que as ligações entre os níveis geralmente não são mostradas. O *roadmap* da *Motorola* (Figura 14) é um exemplo desse formato, com foco na evolução tecnológica associada a um produto e seus recursos.

3.3.2.8. Texto

Alguns *roadmaps* são totais ou principalmente baseados em texto, descrevendo os mesmos problemas incluídos em *roadmaps* gráficos mais convencionais (que, normalmente, têm relatórios associados a eles).

3.3.3. Aplicações, Ênfase e Propósito

Em virtude da diversidade de aplicações, KAPPEL (2001) propôs um modelo que representa a taxonomia dos diferentes tipos de *roadmaps* quanto à sua ênfase e ao seu propósito, como ilustra a Figura 19. O eixo horizontal representa o propósito do *roadmap*, que pode ser desempenhado através de informações do setor industrial ou internas à empresa. O eixo vertical trata-se da ênfase do *roadmap*, normalmente tendo foco em posicionamento ou tendências e trajetórias.

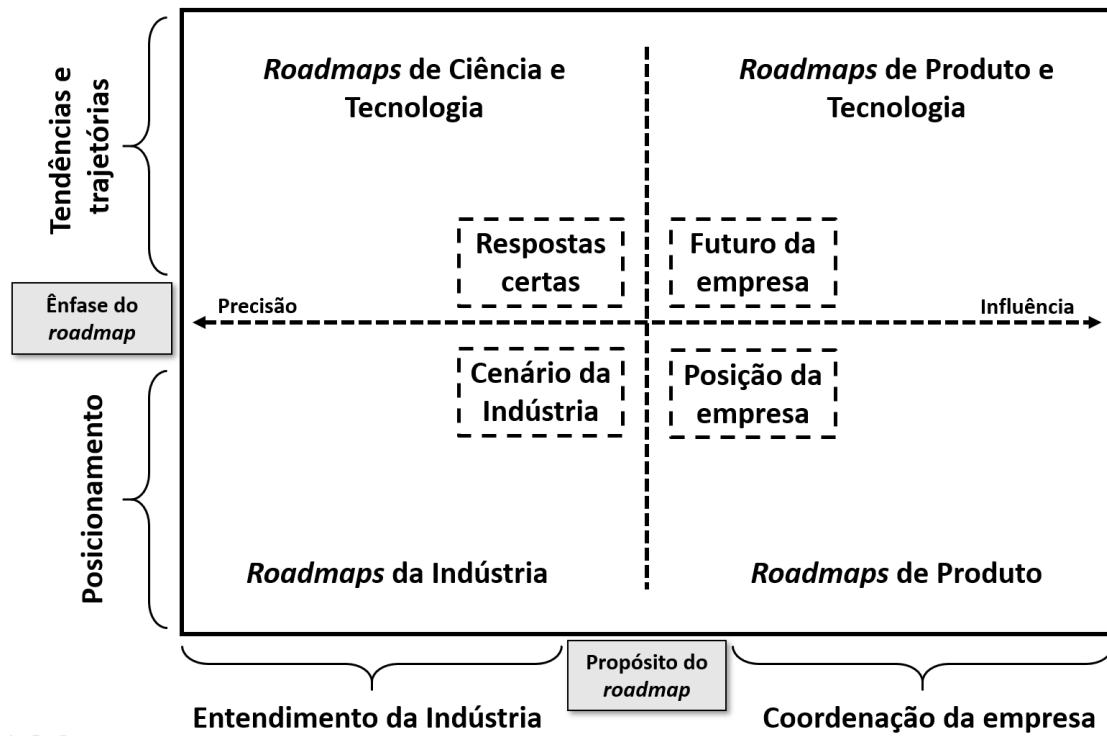


Figura 19. Taxonomias de *roadmaps*. Adaptado de KAPPEL (2001).

De acordo com KAPPEL (2001), a peculiaridade que distingue os *roadmaps* de todos os demais tipos de gráficos estratégicos é a explícita dimensão de tempo para cada elemento representado. O horizonte temporal deve estar presente em todos os *roadmaps*. Os quatro grupos que delineiam tipos distintos de *roadmaps*, são descritos a seguir:

3.3.3.1. *Roadmap de Ciência e Tecnologia*

Ilustrado no quadrante superior esquerdo da Figura 19, tem como propósito melhor entender o futuro da CT&I, predizendo os desenvolvimentos futuros de um campo técnico ou científico. Também, segundo (COELHO et al., 2005) constituem pontes entre

os conceitos de *technology foresight* e de *technology planning*¹⁵ por meio dos incontáveis relacionamentos entre aplicações tecnológicas, avanços tecnológicos potenciais e planos de investimentos para realizar esses avanços. A Tabela 3 apresenta alguns elementos essenciais que um *roadmap* voltado para a CT&I deve conter.

Tabela 3. Componentes de um *roadmap* em CT&I. Adaptado de ALBRIGHT (2005).

Definição e Estratégia	Elementos de ciência e tecnologia	Definição do escopo da área
	Aplicações tecnológicas	Onde e quando as tecnologias se transformarão em inovação (uso) – “os porquês”
Direção	Arquitetura	O modo como os elementos e integram e interagem
	Desafios	Objetivos e metas de desempenho para os elementos tecnológicos – “o que”
	Tendências e descontinuidades	Desempenho e tendências de crescimento, curvas de experiência e possíveis rupturas
Roadmap	Elemento de evolução tecnológica	O <i>Roadmap (Technology Roadmap)</i> – “como”
	Posição técnica competitiva	Enfoque competitivo para os desafios – Tecnologias competitivas
Plano de Ação	Plano de ação	Estratégia tecnológica, recursos e prazos de investimentos em tecnologias – “o que fazer”
	Propriedade intelectual e padrões	Necessidades, barreiras, ações para ter acesso, proteção e influência
	Mapa de investimentos tecnológicos	Prioridades de investimentos em tecnologia
	Mapa de risco	Indicadores chave de risco para os planos; Rastreamento de necessidades de mudança

3.3.3.2. *Roadmap* da Indústria

Quando as previsões de desempenho, adoção e custos das tecnologias são combinadas com informações da indústria, o resultado é o *roadmap* da indústria (ilustrado no quadrante inferior esquerdo da Figura 19) (KAPPEL, 2001). Seu objetivo é desenvolver uma agenda comum entre as empresas do setor, relacionando pontos em comum, como: cadeia de suprimentos, investimentos em larga escala e análise de

¹⁵ *Technology planning* é um processo que demanda tempo e recursos para entender o que é apropriado para a equipe e para a organização. É um processo que identifica o que é ensinado, como é ensinado e qual tecnologia será usada durante uma prospecção (GÜLBAHAR, 2007).

competitividade. O ITRS¹⁶ é um dos exemplos mais conhecidos deste tipo de *roadmap*. (KOSTOFF; SCHALLER, 2001)

3.3.3.3. *Roadmap* de Produto e Tecnologia

Disposto no quadrante superior direito da Figura 19, visam combinar planos específicos de produtos com informações de mercado e tendências de tecnologia. Este *roadmap* engloba as sucessivas famílias de produtos as gerações de tecnologia (KAPPEL, 2001).

3.3.3.4. *Roadmap* de Produto

Seu objetivo é demonstrar a evolução e o cronograma dos produtos como ferramenta de comunicação com clientes e interessados, representado no quadrante inferior direito da Figura 19. A avaliação e entendimento da abordagem de TRM precisa que a discussão seja restrita a um dos quatro domínios apresentados (KAPPEL, 2001).

3.3.3.5. *Roadmaps* para Tecnologias Disruptivas

Tecnologias disruptivas geram riqueza para os setores industriais onde são usadas, ou criam indústrias totalmente novas por meio da introdução de produtos ou serviços, que são muito mais baratos, melhores e mais adequados à demanda (COELHO et al., 2005). Tecnologias disruptivas podem ser geradas a partir de uma nova combinação das tecnologias existentes ou de novas tecnologias (WALSH; LINTON, 2000).

Existem duas grandes perspectivas na geração de uma tecnologia disruptiva emergente. A primeira foca no mercado, onde empresas identificam uma necessidade no mercado e criam as tecnologias necessárias para atender a essa necessidade. Este enfoque *top-down* se inicia com uma mudança no cenário operacional e gera os requisitos para uma tecnologia que resultará em operações disruptivas. A segunda perspectiva é a análise das competências tecnológicas da empresa, analisando se existe um nicho de mercado onde as competências principais podem ser aplicadas. Esse modelo *bottom-up* começa com as tecnologias que já estão em desenvolvimento e identifica o cenário operacional e os mercados nos quais essas tecnologias poderiam entrar de forma disruptiva (COELHO et al., 2005).

¹⁶ ITRS - *International Technology Roadmap for Semiconductors*. Semiconductor Industry Association *Roadmap* (SIA). Disponível em: www.semiconductors.org/. Acesso em setembro/2019.

3.3.3.6. Roadmaps para Tecnologias Convergentes

Existem muitas questões a serem respondidas acerca das tecnologias convergentes que os *roadmaps* podem auxiliar a responder (COELHO et al., 2005):

- Que invenções terão utilidade prática para se tornar inovações de grande uso, e quando estas estarão disponíveis para a sociedade?
- Como os diferentes campos do conhecimento irão interagir para produzir inovações?
- Quais as tendências de mercado e as demandas que vão orientar o desenvolvimento de ações que serão requeridas para uma efetiva comercialização?
- Quais os fatores chave para as inovações e como podem ser satisfeitos?
- Quais os riscos envolvidos nas inovações?

O escopo das tecnologias convergentes é amplo e é preciso definir subáreas para aplicar os métodos de TRM para compreender e expressar em gráficos suas direções.

Um exemplo arquitetura do *roadmap* para tecnologias convergentes, ilustrado na Figura 20, de acordo com ALBRIGHT (2005), define como colocar agrupados os diferentes elementos do problema e auxilia na determinação de prioridades para alcançar os objetivos principais.

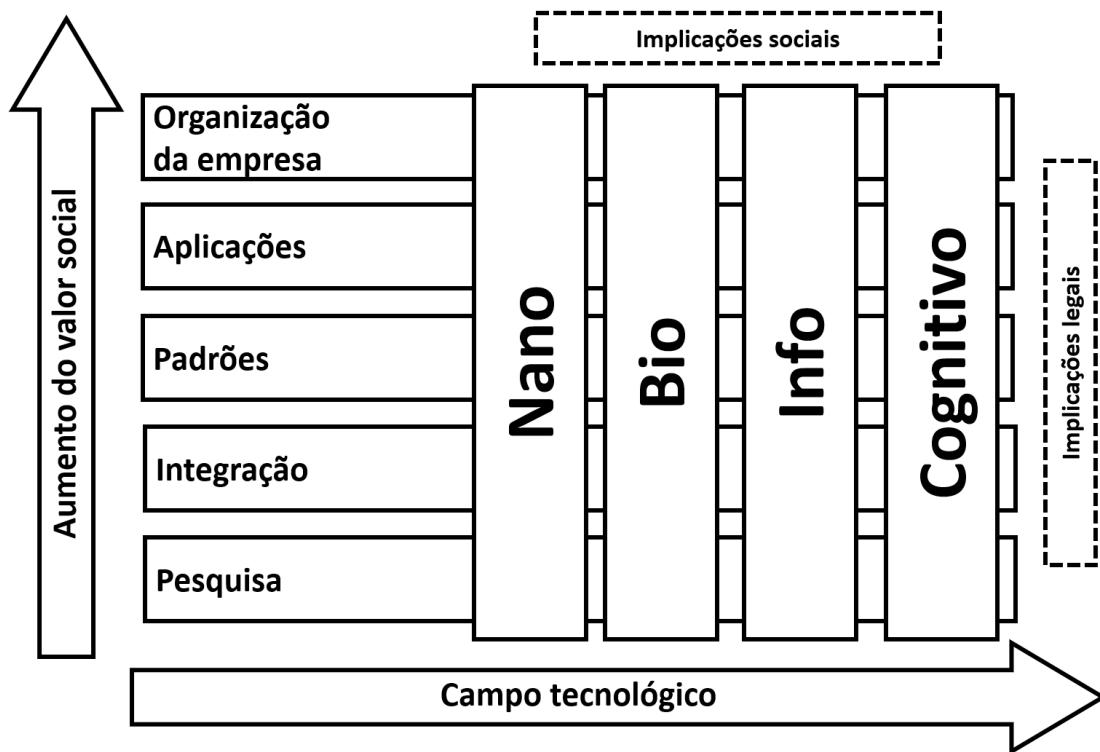


Figura 20. Arquitetura do *roadmap* para tecnologias convergentes. Adaptado de ALBRIGHT (2005).

3.4. Abordagens para Implantações de TRM

Diversos autores analisam as etapas chave para o processo de TRM. Porém, em sua maioria, não incluem a etapa de aplicação do *roadmap*. Nos primeiros estudos publicados na área, os autores se preocupavam em apresentar os diferentes tipos de *roadmaps*, tanto do ponto de vista da aplicação (KAPPEL, 2001), quanto das formas visuais de representação (BROWN; PHAAL, 2001). Mais pra frente, os esforços se passaram a se concentrar em métodos e guias para inicialização rápida da abordagem, de onde surgiu o guia *Fast-Start* (PHAAL; FARRUKH; PROBERT, 2001). As publicações mais recentes falam sobre estratégias de implementação da abordagem e fatores que influenciam a sua aceitação e utilização pelas empresas (LEE et al., 2011).

KAPPEL (2001) lista duas estratégias para implantação do TRM nas empresas:

- **Difusão:** Faz uso de uma abordagem de TRM padronizada e comum para toda a organização;
- **Introdução seletiva:** O TRM é implantado localmente, propiciando uma customização para atender as características do negócio ou segmento da empresa.

Já KOSTOFF e SCHALLER (2001) citam três estratégias de implantação:

- **Abordagem baseada em especialistas:** Nesta abordagem, equipes de especialistas são convidados a identificar e gerar atributos para os nós e ligações do *roadmap*. Os especialistas podem ser externos ou internos a organização.
- **Abordagem computacional:** Neste tipo, bancos de dados textuais que retratam ciência, tecnologia, engenharia e produtos são propensos a estudos computacionais.
- **Abordagem híbrida:** Engloba tanto a análise computacional quanto a avaliação de especialistas, tendo como o objetivo eliminar as fraquezas de cada abordagem anterior.

O *T-Plan Fast-Start* foi desenvolvido como parte de projeto de pesquisa, em que mais de 20 *roadmaps* foram desenvolvidos em cooperação com várias empresas em diversos setores industriais. Segundo PHAAL et al. (2004), este processo engloba duas abordagens: uma considerada padrão e outra que é flexível e capaz de ser adaptada a

diferentes contextos. A experiência e o aprendizado resultantes do *T-Plan* foram convertidos em um guia prático para aplicação do TRM nas organizações.

O processo *T-Plan* compreende duas partes principais (PHAAL; FARRUKH; PROBERT, 2004):

1. Abordagem padrão, para apoiar o planejamento de produtos.
2. Abordagem personalizada, que incluem orientações mais amplas sobre a aplicação do método.

A abordagem padrão, como ilustra na Figura 21, engloba quatro workshops de facilitação: os primeiros três destacando os três níveis do *roadmap* (mercado/negócios, produtos/serviços; e tecnologia). Posteriormente, colocam-se esses três níveis em conjunto, conforme o horizonte temporal, para que se defina o *roadmap*, e esse é realizado após a identificação e compreensão das relações entre os diversos níveis e subníveis. Aspectos de coordenação e planejamento são também vistos como indispensáveis para o êxito do projeto.

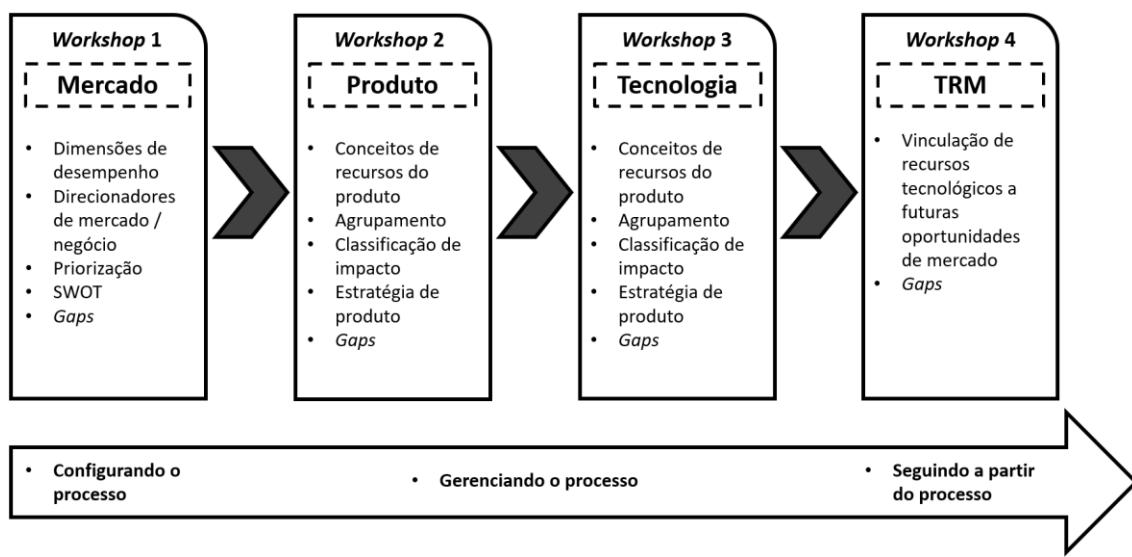


Figura 21. Passos da abordagem *T-Plan* padrão. Adaptado de PHAAL et al. (2004).

A abordagem customizada permite criar níveis e subníveis adaptados a cada caso em particular devido à grande flexibilidade do *T-Plan*. Um TRM é capaz de suportar uma gama de objetivos comerciais diferentes, incluindo planejamento de produtos, exploração de novas oportunidades, alocação e gerenciamento de recursos estratégicos e planejamento aprimorados de negócios. Além disso, cada organização é diferente em termos de contexto comercial, cultura organizacional, processos comerciais, recursos disponíveis, tipos de tecnologia, entre outros.

Desse modo, para obter todos os benefícios do *roadmap*, deve-se esperar que a abordagem seja personalizada para se adequar ao ambiente que está sendo estudado. Conforme PHAAL et al. (2004), o *roadmap* de múltiplos níveis é a forma mais comum e a aplicação mais flexível para essa abordagem, incluindo as seguintes dimensões:

- **Tempo:** essa dimensão pode ser adaptada para se adequar à situação específica, em termos de horizonte temporal; escalas; e intervalos.
- **Níveis:** corresponde ao eixo vertical do *roadmap* e é crítico, pois ele precisa ser projetado para atender à organização e ao problema que está sendo tratado de maneira específica. Devido à isso, um esforço considerável deve ser empregado para definir os níveis e subníveis que irão compor o *roadmap*.
- **Anotação:** Além das informações contidas nos níveis, com base no horizonte temporal, outras informações podem ser armazenadas no *roadmap*, incluindo:
 - Ligações entre objetos em níveis e subníveis (de vários tipos possíveis);
 - Informações suplementares, como chave, declaração da estratégia comercial ou direcionadores do mercado, pessoas envolvidas no desenvolvimento do *roadmap* e premissas;
 - Outros dispositivos gráficos, incluindo objetos, notas e códigos de cores, para indicar os principais pontos de decisão, lacunas, caminhos críticos, oportunidades e ameaças (incluindo mercados e tecnologias disruptivas).
- **Processo:** as etapas necessárias para concluir o primeiro *roadmap* e levar o processo adiante posteriormente serão tipicamente diferentes para cada organização (e geralmente também em processos dentro da mesma organização). O processo mais adequado depende de muitos fatores, incluindo o nível de recursos disponíveis (pessoas, tempo, orçamento), natureza do problema abordado (objetivo e escopo), informações disponíveis (mercado e tecnologia) e outros processos e métodos de gerenciamento que são relevantes (estratégia, orçamento, desenvolvimento de novos produtos, gerenciamento de projetos e pesquisa de mercado).

No âmbito de uma arquitetura multinível, PHAAL et al. (2004) julgam essencial que os aspectos-chave do conhecimento sobre negócios sejam capturados, estruturados, compartilhados e que se consiga acordo sobre questões estratégicas identificadas e ações concretas. O alinhamento do propósito (*know-why*), com os objetivos e estratégias (*know-what*), e os recursos tecnológicos (*know-how*) e o tempo (*know-when*) são os fatores que propiciam atingir um bom equilíbrio entre o *market pull* e *technology push*. A Figura 22 apresenta uma arquitetura de *roadmap* generalizada, com base em muitos *roadmaps* que foram desenvolvidos e observados no estudo de PHAAL et al. (2004). Os diferentes tipos de níveis nos *roadmaps* são listados, destacando a flexibilidade da abordagem em termos de fornecer uma estrutura para apoiar o planejamento estratégico.

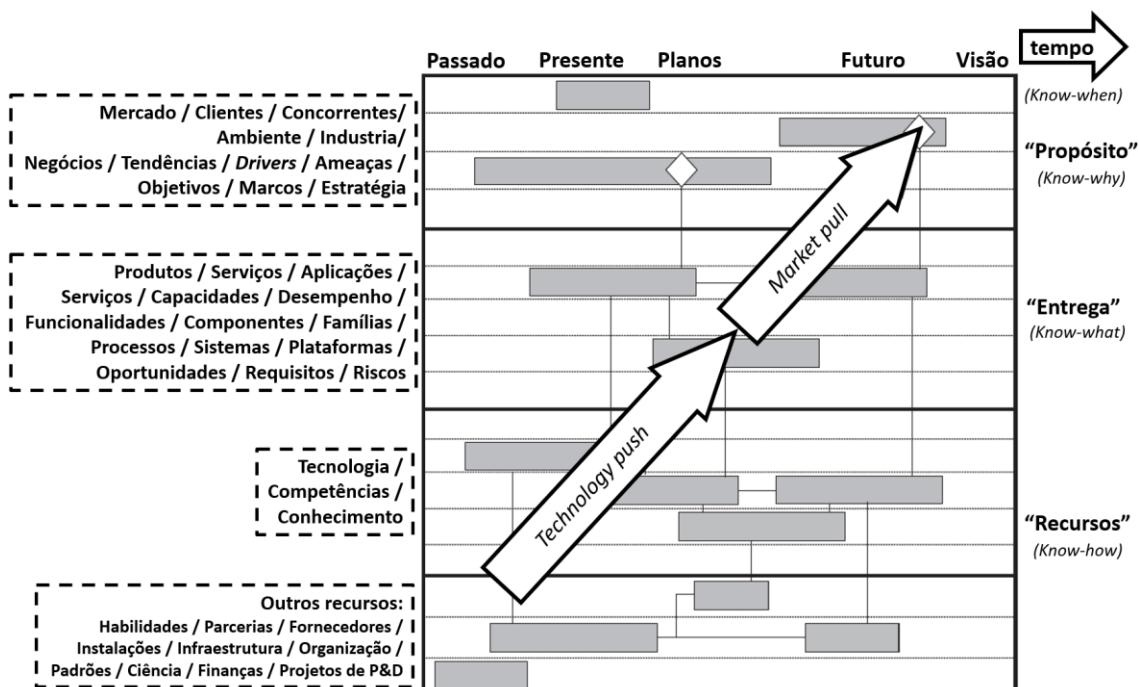


Figura 22. Arquitetura de *roadmap* generalizada. Adaptado de PHAAL et al. (2004).

Os níveis superiores estão relacionados ao propósito (objetivo organizacional) que está “guiando” o *roadmap* (*know-why*). Os níveis inferiores estão relacionados aos recursos que serão implantados para atender à demanda dos níveis superiores do *roadmap* (*know-how*). Os níveis intermediários do *roadmap* são cruciais, fornecendo um mecanismo de ponte – ou entrega – entre o propósito e os recursos (*know-what*).

A maioria dos *roadmaps* incluem aspectos de *market pull* e *technology push* combinando perspectivas externas e internas (oportunidades, ameaças, pontos fortes/fracos) para identificar e avaliar um conjunto de opções de tecnologias, como mostrado na Figura 23. Uma abordagem de *technology push* é geralmente mais divergente e complexa em comparação com *market pull*, pois uma tecnologia específica pode ter

muitas aplicações em domínios em que a empresa não tem experiência (PHAAL; FARRUKH; PROBERT, 2004).

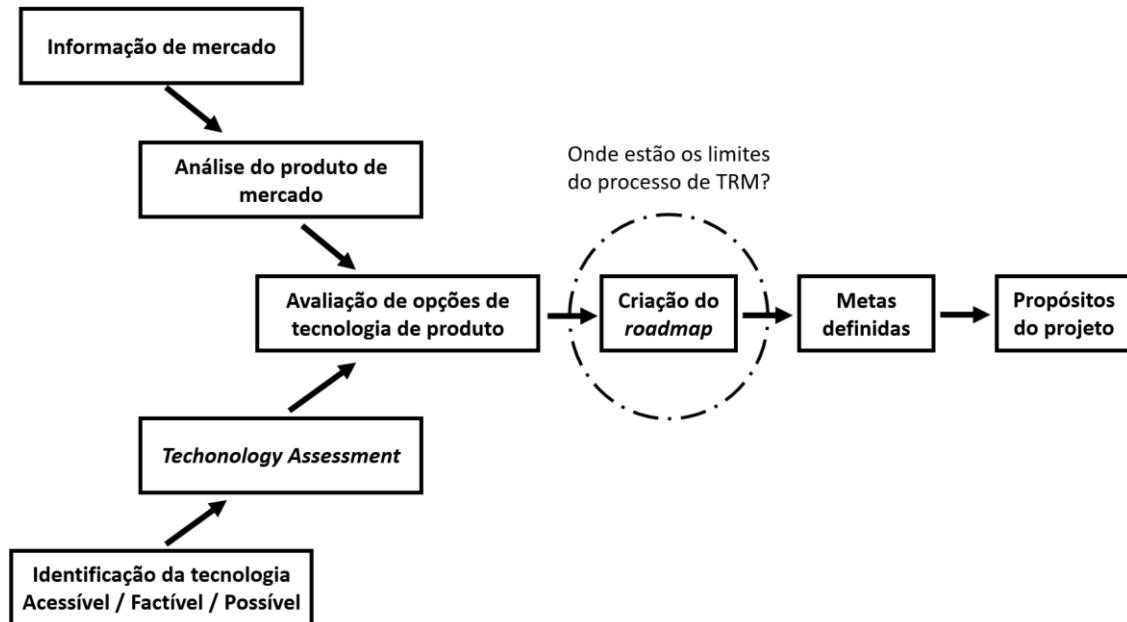


Figura 23. *Roadmaps* integram conhecimento comercial e tecnológico. Adaptado de PHAAL et al. (2004).

Segundo PHAAL et al. (2010), a abordagem *S-Plan*, assim como a *T-Plan*, também se baseia em *workshops* interativos para juntar diversos grupos de participantes para, assim, capturar e discutir perspectivas, explorar opções e oportunidades, tomar decisões, conciliar ações e, por fim, gerar *roadmaps* preliminares. O *S-Plan* tem a capacidade de convergir de forma rápida para um escopo muito amplo (tipicamente negócios, corporação ou níveis setoriais), capturando perspectivas, identificando e priorizando tópicos chaves para exploração e planejamento da ação. Se um dos tópicos enfoca o nível de inovação do produto, serviço ou processo, então o *T-Plan* fornece um método mais detalhado para integrar informações de mercado, produto, tecnologia e recursos.

Uma outra abordagem de implantação de TRM em uma organização foi proposta por, VATANANAN e DANSAMASADIT (2009), que pode ser dividida em três estágios: Iniciação, Desenvolvimento e Integração (conforme a Figura 24).

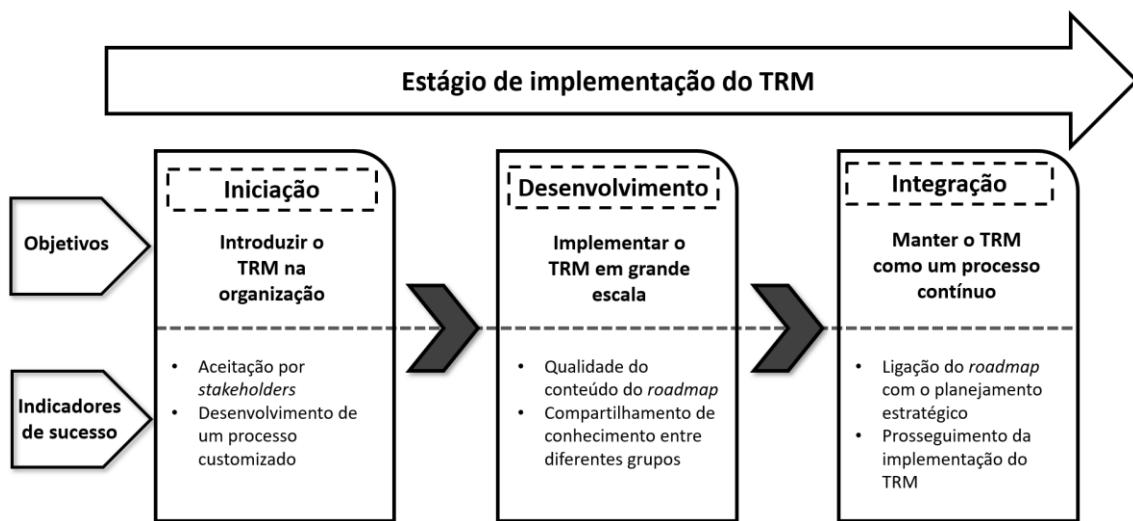


Figura 24. Estágios de implantação de um TRM. Adaptado de VATANANAN e DANSAMASADIT (2009).

O estágio de iniciação tem como objetivo preparar a organização para iniciar o processo de TRM. Ele compreende a formação de equipes para o entendimento básico sobre o TRM visando a customização do processo para adaptá-lo ao processo de planejamento estratégico e cultura da organização.

O estágio de desenvolvimento engloba o desenvolvimento do primeiro *roadmap* por meio do envolvimento de pessoas adequadas, coleta de informações e a condução de uma análise do tipo passo-a-passo. O principal objetivo desta fase é a coleta e análise de informações, onde *workshops* são realizados para a análise das informações e a construção do *roadmap*.

O último estágio visa integrar o TRM aos processos operacionais da empresa, de modo que o *roadmap* seja revisado e atualizado de forma estruturada.

Complementando toda a revisão apresentada sobre as abordagens de implantação de TRM, PHAAL et al. (2010) recomendam uma abordagem evolutiva, compreendendo três fases: a fase piloto, a fase de aprendizado e a fase de desenvolvimento de um sistema amplo de TRM para a organização.

A fase piloto compreende um primeiro teste exploratório de como o TRM pode dar suporte às necessidades da empresa. A fase de aprendizado, onde toda a experiência de implantação da fase piloto é coletada e compartilhada, tem como objetivo alinhar as práticas em um método único para a organização. A última fase recomendada, de desenvolver um sistema amplo de TRM para a organização, precisa de experiência e conhecimento suficiente por parte das equipes. A implantação depende da maturidade do negócio e dos sistemas de gestão. Nesta fase, há necessidade de relacionar o TRM a outros

processos, como planejamento estratégico, orçamentos e processo de desenvolvimento de produtos. A chave do sucesso para este processo é a padronização para toda a organização.

3.5. Benefícios de Realizar TRM

O principal benefício do produto resultante do TRM – o *roadmap* – é que este provê informações que permitem tomar melhores decisões sobre investimentos em P&D, pela identificação das tecnologias críticas e dos *gaps* existentes e identificar formas de alavancar investimentos.

Segundo GARCIA e BRAY (1997), o benefício mais importante do TRM são as informações para que as decisões sobre os investimentos tecnológicos sejam feitas pela identificação de tecnologias críticas e de gargalos tecnológicos, além de identificar caminhos para conseguir financiamentos para P&D. Pode ser usado também como um instrumento de *marketing*. *Roadmap* é crítico quando a decisão sobre o investimento em tecnologia não é clara. Identifica qual é a melhor alternativa, quão rapidamente a tecnologia é necessária ou onde é preciso coordenar o desenvolvimento de múltiplas tecnologias.

Tanto para as empresas individualmente, quanto setorialmente, *roadmaps* têm vários usos potenciais com benefícios visíveis. De acordo GARCIA e BRAY (1997), os três principais usos são:

1. Auxiliar a desenvolver um consenso sobre um conjunto de necessidades e as tecnologias necessárias para atendê-las;
2. Proporcionar um mecanismo para ajudar a antecipar desenvolvimentos tecnológicos em áreas selecionadas;
3. Gerar uma estrutura para ajuda no planejamento e na coordenação de desenvolvimentos tecnológicos, tanto para uma empresa quanto para um segmento industrial.

GARCIA e BRAY (1997) também citam que o principal benefício de um *roadmap* é fornecer informações para ajudar a organização a tomar melhores decisões sobre investimentos em tecnologia, por meio de:

- Identificação de tecnologias críticas ou *gaps* tecnológicos que precisam ser preenchidos para auxiliar o desempenho dos produtos desejados;

- Reconhecimento de meios de alavancar investimentos em P&D através da coordenação das atividades de pesquisa, seja em uma única empresa ou por parcerias.

Ainda de acordo com GARCIA e BRAY (1997), o processo de TRM é potencialmente capaz de:

- Mostrar visões e aproximar recursos da indústria e do governo;
- Estimular pesquisas e monitorar progressos;
- Encorajar a formação de redes interdisciplinares e trabalho em equipes;
- Tornar-se um inventário de possibilidades para um campo específico, estimulando pesquisas com um maior aprofundamento.

Capítulo 4 – Identificando Prospecções em Textos

Este capítulo descreve a metodologia aplicada para a identificação de prospecções, que podem ser vistas como acontecimentos em um horizonte temporal, para a formação de *roadmaps*. Em seguida, apresentamos uma visão geral de Mineração de Textos, detalhando os conceitos de Processamento de Linguagem Natural e Extração da Informação necessários para o contexto deste trabalho, que faz uso da técnica de Reconhecimento de Entidades Nomeadas.

Em seguida, a ênfase é direcionada para a técnica de Reconhecimento de Entidades Nomeadas, descrevendo conceitos e métodos relevantes ao desenvolvimento do *framework* proposto neste trabalho.

4.1. Conceitos Fundamentais

A identificação de prospecções em textos pode ser realizada por diferentes metodologias. A metodologia aplicada neste trabalho divide os termos que compõe um texto prospectivo em 3 (três) tipos:

- **Espaço Temporal:** Termo que designa uma data, podendo ser um ano ou uma expressão de tempo;
- **Termos indicativos de futuro:** Palavras ou expressões que apontam que um acontecimento vai acontecer num futuro, sempre associado a um espaço temporal;
- **Acontecimento:** Trecho que indica a prospecção em si, ou seja, algo que foi previsto para acontecer no futuro.

Na Tabela 4 é apresentado um exemplo destes termos sendo identificados em um excerto de texto de uma publicação que trata sobre o futuro. As prospecções ordenadas de forma cronológica para o exemplo da Tabela 4 são mostradas na Tabela 5, de modo a formar um *roadmap* para o assunto tratado na publicação.

Tabela 4. Exemplo de identificação de prospecções em textos.

Tipo	Entrada de Dados
Espaço Temporal	<p><i>In 2030, China will lead the world in Artificial Intelligence. According to Eric Schmidt “it’s pretty simple. By 2020, China will have caught up. By 2025, China will be better than USA. For the year of 2030, China will dominate the industries of AI.”</i></p> <p><i>And the numbers don’t lie, says Peter Diamandis: “PricewaterhouseCoopers recently projected AI’s deployment will add \$15.7 trillion to the global GDP by 2030, with China taking home \$7 trillion of that total, dwarfing North America’s \$3.7 trillion in gains.”</i></p>
Termos Indicativos de Futuro	<p><i>In 2030, China will lead the world in Artificial Intelligence. According to Eric Schmidt “it’s pretty simple. By 2020, China will have caught up. By 2025, China will be better than USA. For the year of 2030, China will dominate the industries of AI.”</i></p> <p><i>And the numbers don’t lie, says Peter Diamandis: “PricewaterhouseCoopers recently projected AI’s deployment will add \$15.7 trillion to the global GDP by 2030, with China taking home \$7 trillion of that total, dwarfing North America’s \$3.7 trillion in gains.”</i></p>
Acontecimento	<p><i>In 2030, China will lead the world in Artificial Intelligence. According to Eric Schmidt “it’s pretty simple. By 2020, China will have caught up. By 2025, China will be better than USA. For the year of 2030, China will dominate the industries of AI.”</i></p> <p><i>And the numbers don’t lie, says Peter Diamandis: “PricewaterhouseCoopers recently projected AI’s deployment will add \$15.7 trillion to the global GDP by 2030, with China taking home \$7 trillion of that total, dwarfing North America’s \$3.7 trillion in gains.”</i></p>

Tabela 5. Exemplo de *roadmap* a ser formado.

Data	Prospecções
2020	<i>By 2020, China will have caught up.</i>
2025	<i>By 2025, China will be better than USA.</i>
2030	<i>In 2030, China will lead the world in Artificial Intelligence.</i>
2030	<i>For the year of 2030, China will dominate the industries of AI.</i>
2030	<i>PricewaterhouseCoopers recently projected AI’s deployment will add \$15.7 trillion to the global GDP by 2030, with China taking home \$7 trillion of that total, dwarfing North America’s \$3.7 trillion in gains.</i>

Pela necessidade de identificar esses termos, foi empregada a técnica de Reconhecimento de Entidades Nomeadas, que está inserida dentro da área de Mineração de Textos. Na próxima subseção serão detalhados os conceitos desta área necessários para o entendimento deste trabalho e que precedem o aprofundamento em REN, detalhado na seção 4.3.

4.2. Uma Visão Geral de Mineração de Textos

Mineração de Textos (MT) relaciona-se ao processo não trivial de extração de padrões úteis e relevantes – conhecimento – a partir de uma coleção de documentos textuais não estruturados (FELDMAN; HIRSH, 1997).

De acordo com WEISS et al. (2010), as técnicas e os métodos aplicados na Mineração de Textos são análogos aos utilizados na Mineração de Dados (MD), isto é, fazem uso dos mesmos métodos de aprendizagem, independente se uma técnica se utiliza de dados textuais (MT) e outra de dados numéricos (MD).

MORAIS e AMBRÓSIO (2007) apresentam que a Mineração de Textos é um processo de Descoberta de Conhecimento que emprega ferramentas computacionais e algoritmos específicos a fim de processar e identificar textos, frases ou apenas palavras. Dessa forma, pode-se extrair dados que não poderiam ser acessados de outra maneira, pois normalmente é utilizado em textos que estão armazenados em formato não estruturado, isto é, formatos de dados em linguagem natural ou sem nenhum tipo de padronização (WIVES, 2002).

O estudo de ARANHA (2007) relata em seu trabalho a metodologia de Mineração de Textos dividida em cinco etapas diferentes, como mostrado na Figura 25. A primeira se baseia na coleta dos dados, a segunda no pré-processamento dos mesmos, em prol de elaborar o primeiro nível de estruturação, a terceira etapa confere a criação dos índices que permitem uma melhora na recuperação dos dados, a quarta na aquisição de conhecimento e, por fim, uma quinta fase para interpretação dos resultados obtidos.

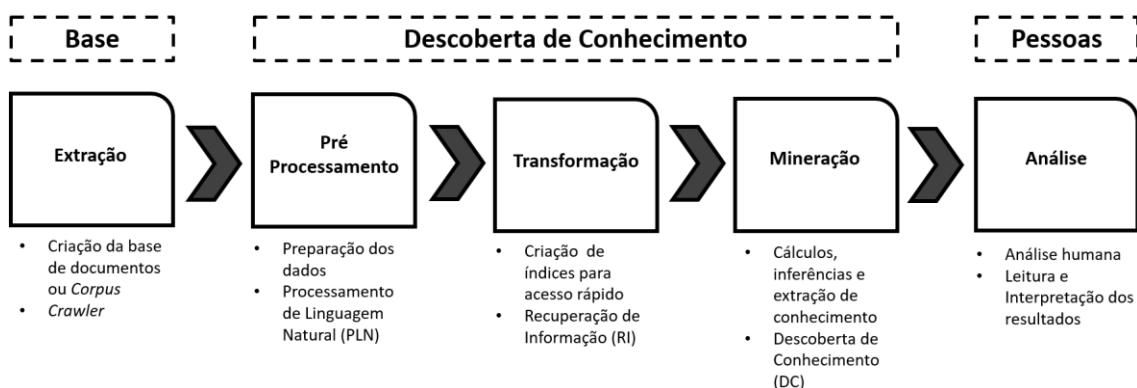


Figura 25. Diagrama da metodologia de Mineração de Textos. Adaptado de ARANHA (2007).

Ainda de acordo com a abordagem de MORAIS e AMBRÓSIO (2007), basicamente a Descoberta de Conhecimento pode ser dividida em duas áreas a saber: Descoberta de Conhecimento em Dados Estruturados (*Knowledge Discovery in*

Databases - KDD) e a Descoberta de Conhecimento em Dados não Estruturados (*Knowledge Discovery in Text* - KDT), de acordo com a Figura 26.

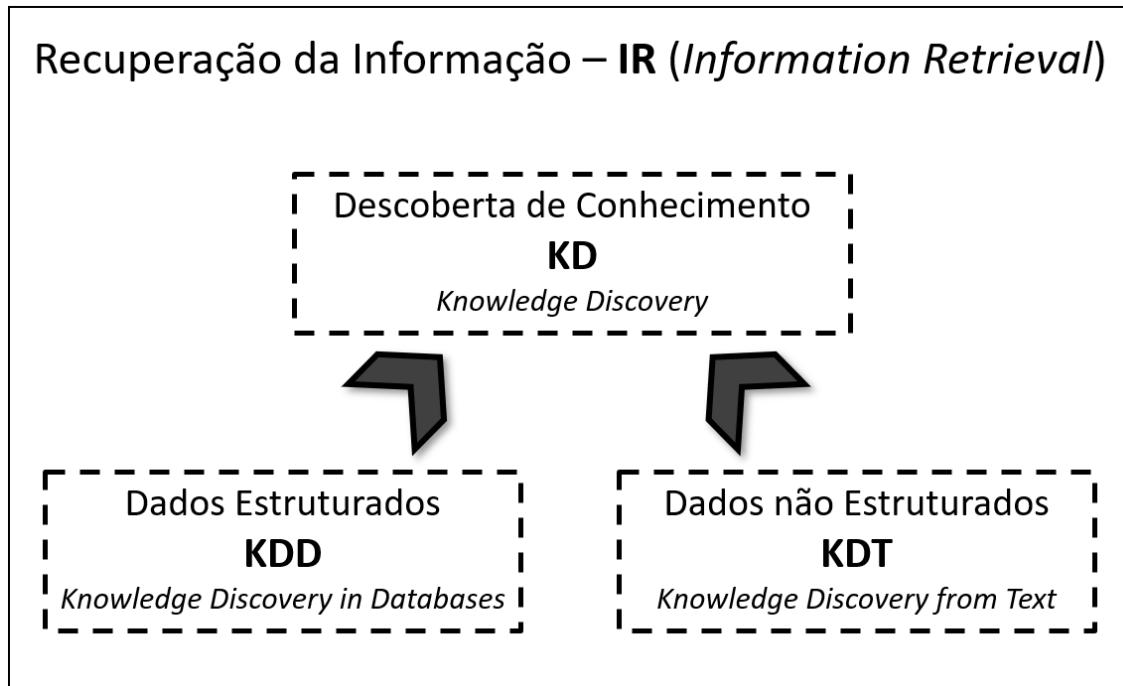


Figura 26. Tipos de Descoberta de Conhecimento. Adaptado de MORAIS e AMBRÓSIO (2007).

Uma visão prática da aplicação de KDD, de acordo com de FAYYAD et al. (1996), enfatizando a interatividade do processo, é mostrada na Figura 27.

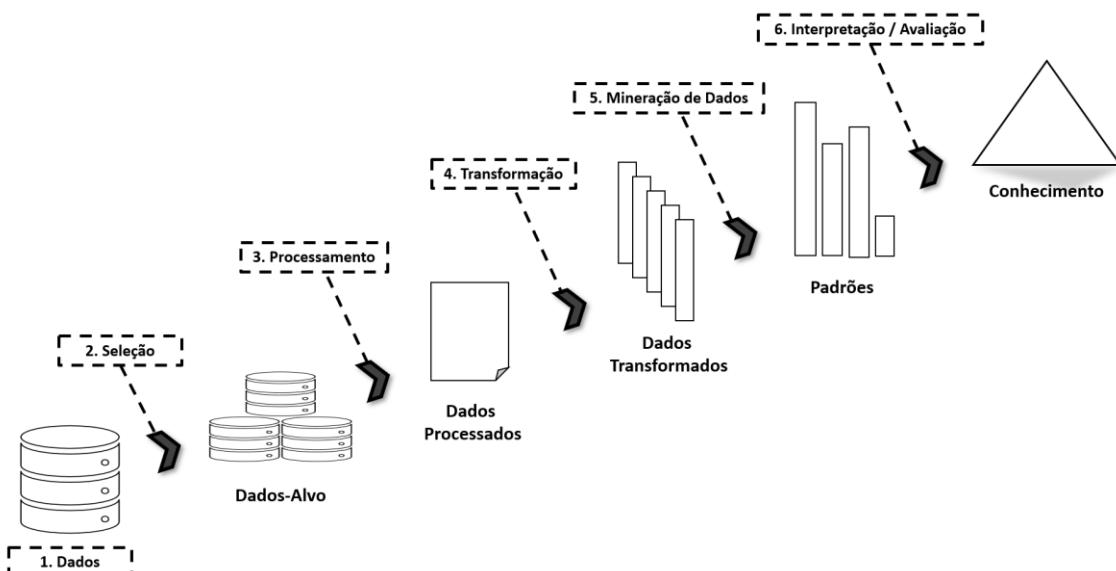


Figura 27. Etapas de Mineração de Textos (KDD). Adaptado de FAYYAD et al. (1996).

As etapas que compõem o processo do KDD, apresentadas na Figura 27, são descritas brevemente, de acordo com FAYYAD et al. (1996):

- 1. Dados:** A Descoberta de Conhecimento em Banco de Dados se fundamenta nos dados armazenados de forma estruturada;
- 2. Seleção dos Dados:** A etapa de seleção é onde um conjunto de dados é colhido de forma a atender aos objetivos finais. Para isso, é preciso entender o domínio da aplicação e conhecer esses objetivos;
- 3. Processamento:** Nesta etapa os dados são processados (ou seja, é feita a “limpeza” dos dados) visando adequá-los aos algoritmos. Isso se faz através da integração de dados heterogêneos, eliminação de dados incompletos, eliminação de tuplas repetidas, problemas de definição de tipos, entre outras;
- 4. Transformação:** A etapa de transformação armazena os dados de forma adequada no banco de dados, de forma a facilitar a aplicação dos algoritmos de mineração.
- 5. Mineração de Dados:** Esta etapa começa com a escolha do algoritmo a ser aplicado (classificação, clusterização, regras associativas, summarização, entre outros). Dependendo do objetivo que se deseja alcançar com o processo de mineração, um algoritmo é aplicado nos dados obtendo-se padrões;
- 6. Interpretação / Avaliação:** A etapa final do processo resume-se em interpretar os resultados obtidos e elaborar uma apresentação de modo que o usuário possa entender.

A KDT emergiu visando o tratamento de informações textuais, surgindo da necessidade de análises automáticas em textos, já que a grande quantidade de textos disponível, juntamente com a falta de padronização e organização destes, ocasionou uma sobrecarga de informações, dificultando sua análise manual, localização e acesso. (FELDMAN; DAGAN, 1995)

Conforme BEPPLER et al. (2005), KDT abrange técnicas e ferramentas inteligentes e automáticas que contribuem para a análise de grandes massas de dados com o objetivo de “garimpar” conhecimento útil e, dado que a forma mais comum de armazenamento de informação é através de texto, a KDT, teoricamente, tem um potencial maior de aplicação do que KDD, visto que cerca de 80% das informações contidas nas organizações estão mantidas em documentos textuais.

De acordo com MINER et al. (2012), a Mineração de Textos é constituída pela intersecção entre sete áreas de atuação (Extração da Informação, Processamento de Linguagem Natural, Extração de Conceitos, Mineração Web, Recuperação da Informação, Agrupamento de Documentos e Classificação de Documentos) e seis

principais campos de trabalho (Mineração de Dados, Estatísticas, Inteligência Artificial e Aprendizagem de Máquina, Linguística Computacional, Biblioteca e Ciências da Informação e Bancos de Dados), como ilustrado na Figura 28.

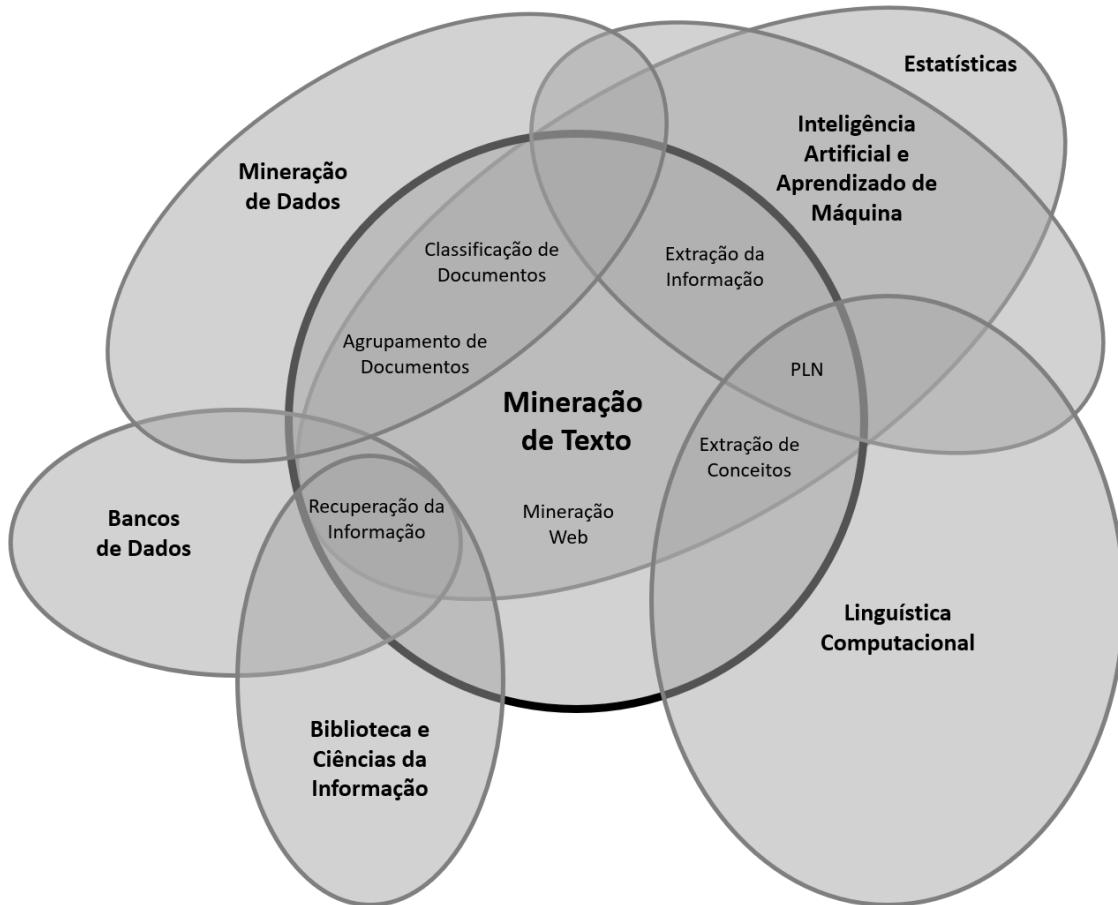


Figura 28. Intersecção das áreas de atuação e campos de trabalho de Mineração de Textos. Adaptado de MINER et al. (2012) e CARVALHO (2017).

Duas áreas de atuação são as mais importantes para o desenvolvimento deste trabalho, Processamento de Linguagem Natural e Extração da Informação, ambos presentes nos campos de Mineração de Texto e Inteligência Artificial e Aprendizado de Máquina. Destas duas áreas, extraímos a técnica de Reconhecimento de Entidades Nomeadas, principal método utilizado na execução deste trabalho e que será discutido nas próximas subseções.

4.2.1. Processamento de Linguagem Natural

Processamento de Linguagem Natural consiste no desenvolvimento de métodos computacionais para a análise e manipulação ou codificação de informações representadas em língua natural (ALLEN, 1995).

O principal objetivo do PLN é extrair um significado mais amplo da representação do texto livre, obtendo informações como “quem fez o quê”, “a quem”, “quando”, “onde”, “como” e “porquê”. O PLN normalmente faz uso de conceitos linguísticos, tais como discurso (substantivo, verbo, adjetivo, entre outros) e estrutura gramatical (seja representada como frases, como sintagma nominal ou frase preposicional, ou relações de dependência como sujeito e objeto). Também precisa lidar com anáforas e ambiguidades (tanto de palavras quanto de estrutura gramatical), fazendo uso de diversas representações de conhecimento, tais como, um conjunto de regras gramaticais, além de outros recursos, como uma ontologia de entidades e ações, ou até mesmo, uma enciclopédia de sinônimos ou abreviações (KAO; POTEET, 2007).

Algumas das principais tarefas de PLN são reconhecimento de contexto, análise sintática, semântica, léxica e morfológica, sumarização e tradução de textos (MANNING; SCHÜTZE, 1999).

O Processamento Automático da Linguagem Natural visa aproximar o computador da realidade do homem, através do desenvolvimento de ferramentas que permitam uma comunicação mais natural entre homem e máquina, além de ferramentas para a extração de informações de grandes bases textuais e tradução automática de textos (GASPERIN; LIMA, 2001).

O Processamento Estatístico da Linguagem Natural consiste no uso de abordagens quantitativas para o processamento automático de textos. Uma coleta de informações é realizada sobre uma grande base de textos, denominada corpus, para o levantamento das probabilidades de ocorrência de palavras, sequências de palavras, categorias de palavras, entre outros (GASPERIN; LIMA, 2001).

4.2.1.1. Técnicas de PLN

Nesta seção veremos algumas técnicas geralmente empregadas no processamento de textos e que são úteis para o entendimento deste trabalho.

4.2.1.1.1. Tokenização

Conforme o estudo de GASPERIN e LIMA (2001), o processo de tokenização consiste em uma das primeiras etapas do processamento de texto, tendo como objetivo dividir o texto de entrada em unidades chamadas *tokens*, onde cada unidade é uma palavra, ou ainda um número ou uma marca de pontuação.

De acordo com SEIDEL (2012), o resultado desse processo é uma sequência de palavras intercaladas por espaços ou por símbolos delimitadores. Por exemplo, na língua

portuguesa, após o processo de tokenização, a frase “Disseram-me que hoje poderia chover, mas não está parecendo.” poderia ser representada como a seguinte sequência de *tokens*: [Disseram] [-me] [que] [hoje] [poderia] [chover] [,] [mas] [não] [está] [parecendo] [.].

GASPERIN e LIMA (2001) definem que a especificação do que deve ser considerado como uma palavra é complexa. A principal informação utilizada é a ocorrência de um espaço em branco – espaço, tabulação ou o início de uma nova linha – mas mesmo este sinal não é necessariamente confiável. Alguns desses problemas inerentes ao processo de tokenização são mostrados a seguir:

- **Marcas de pontuação:** As palavras nem sempre vêm cercadas por espaços em branco. Comumente as marcas de pontuação vêm junto com as palavras, a exemplo da vírgula, do ponto e vírgula e do ponto final. A remoção dos pontos finais que se encontram com as palavras é um dos problemas, já que um ponto pode indicar, por exemplo, uma abreviação;
- **Hífen:** Dentre todas as suas aplicações, o hífen pode comumente ser utilizado como separador de sílabas de uma palavra; entre as palavras de um substantivo composto; ou ainda, principalmente na língua inglesa, para auxiliar a qualificação de uma palavra. Nos dois primeiros casos tende-se a considerar uma palavra apenas, porém no último caso descrito, pode-se reconhecer as palavras separadamente.

4.2.1.1.2. Separação de Sentenças

Pode-se imaginar que “uma sentença é alguma sequência de caracteres que termina com “?”, “.” ou “!”. Entretanto, uma marca de pontuação pode não indicar o fim de uma sentença, mas sim uma abreviação, ou ambas as funções simultaneamente, que neste último caso representa uma haploglia. Além do problema de ambiguidade destes delimitadores, existem frases que contêm outras frases dentro delas, a exemplo das ocorrências de frases com aspas e parênteses (GASPERIN; LIMA, 2001; SEIDEL, 2012).

Apesar dessas dificuldades, conforme RILEY (1989), 90% dos pontos são indicadores do fim de uma sentença. Em seu estudo, MANNING e SCHÜTZE (1999), afirmam que, na prática, a maior parte das soluções para a separação de sentenças englobam métodos heurísticos.

Contudo, estas soluções demandam marcação manual e conhecimento do domínio por parte do desenvolvedor do sistema de “tokenização”. Encontram-se ainda diversas

outras abordagens para a automatização do processo de separação de sentenças. Dentre elas há propostas de técnicas que empregam árvores de classificação estatísticas, redes neurais baseadas na distribuição das categorias das palavras, ou ainda abordagem baseada em maximização de entropia (GASPERIN; LIMA, 2001; SEIDEL, 2012).

4.2.1.1.3. Análise Morfossintática

A Análise Morfossintática pode ser definida como a tarefa de PLN que reconhece corretamente a classe de cada uma das palavras que compõem uma sentença. Estas classes de palavras são consideradas como classes gramaticais ou categorias morfossintáticas. Em geral, estas classes são caracterizadas por um conjunto de códigos, e estes são empregados na etiquetagem das palavras (SEIDEL, 2012).

Ainda conforme SEIDEL (2012), um problema inerente à esta tarefa é a existência de muitas palavras com diversas classificações possíveis. Tais palavras, se estiverem fora de contexto, provocam a ambiguidade sobre sua interpretação para a correta classificação gramatical. Por exemplo, na frase “Vamos assistir ao jogo”, a palavra “jogo” é um substantivo que pode significar, dentre outras, uma partida de futebol. Entretanto, a mesma palavra empregada na frase “Eu jogo videogame”, trata-se de uma flexão na primeira pessoa do singular do presente do indicativo do verbo “jogar”. A palavra “jogo” é um exemplo de ambiguidade existente na língua portuguesa.

De acordo com MANNING e SCHÜTZE (1999), a coleção de códigos mais difundida é utilizada pelo corpus *American Brown (Brown tagset)*¹⁷ e as séries de coleções desenvolvidas na Universidade Lancaster. O *Penn Treebank tagset*¹⁸ é uma versão mais simplificada do *Brown tag set* e tem sido uma coleção de códigos largamente utilizada computacionalmente.

As informações das classes gramaticais das palavras presentes no texto, permitem resolver, de forma mais simples, problemas mais complexos como a análise sintática, tradução ou até mesmo a extração da informação de textos (ALVES, 2003).

¹⁷ *American Brown (Brown tagset)*: Corpus do inglês americano.

Pode ser acessado em: <https://www.sketchengine.eu/brown-corpus>.

¹⁸ *Penn Treebank tagset*: Uma versão simplificada do corpus *Brown corpus*.

Pode ser acessado em: <https://www.sketchengine.eu/penn-treebank-tagset>.

4.2.2. Corpus

Um corpus pode ser definido como uma coleção especial de textos coletados conforme critérios específicos e, de acordo com MANNING e SCHÜTZE (1999), é um dos principais requisitos para o processamento estatístico da linguagem natural. Tal recurso é utilizado para treinamento e teste de modelos estatísticos de linguagem natural escrita e falada, bem como para avaliação de componentes de sistemas de linguagem natural.

Na seleção ou elaboração de um corpus deve-se dar importância aos tipos de informações que são alvo de um determinado problema. A título de exemplo, se um corpus foi desenvolvido como uma amostra representativa de etiquetagem morfológica de textos históricos, as estimativas obtidas deste podem não ser adequadas para textos contemporâneos. Além do mais, deve-se ter prudência com a validade dos resultados da análise estatística obtida, observando os critérios de anotação adotados na coleção (GASPERIN; LIMA, 2001).

Segundo KRENN e SAMUELSSON (1997), as características fundamentais e observáveis de um corpus são:

- Ocorrência de fenômenos linguísticos no contexto;
- Agregação de aspectos de competência e performance: os dados do corpus refletem o uso da linguagem – todos os dados, até mesmo os que são falsos em termos de competência gramatical, devem ser julgados como úteis;
- Informações de frequência de uma diversidade de fenômenos em vez de fenômenos selecionados;
- Consideração de todos os dados existentes pelo esquema de anotação, sem discernimento entre termos competentes ou não.

Também conforme KRENN e SAMUELSSON (1997), corpora¹⁹ podem ser classificados de acordo com o tipo de texto que os mesmos contêm, de acordo com o tipo de anotação que possuem, e segundo o uso a que se destinam (GASPERIN; LIMA, 2001).

De acordo com tipo de texto:

- **Balanceado:** baseiam-se em gêneros distintos de textos de tamanho ou quantidade proporcional à relevância de certo tipo de texto na linguagem em questão;

¹⁹ Corpora: plural de corpus.

- **Piramidal:** contém desde grandes amostras de tipos de texto pouco representativos, até pequenas amostras de uma grande diversidade de tipos;
- **Oportunista:** a maior parte dos corpora disponíveis pode ser caracterizada como oportunista, ou seja, contêm o que se deseja;
- **Paralelo:** apresenta textos que possuem o mesmo conteúdo, porém em duas ou mais línguas diferentes. Estes corpora são comumente utilizados em sistemas de tradução automática.

De acordo com tipo de anotação:

- **Original:** o texto é “tokenizado” e limpo, ou seja, caracteres de controle são eliminados. O tipo do texto, títulos e parágrafos são possivelmente marcados;
- **Marcado com categorias das palavras:** o texto original é anotado com a categoria sintática de cada palavra;
- **Treebanks:** o texto marcado com as categorias das palavras é anotado com uma estrutura sintática. Usualmente, uma gramática é definida, e os corpora são automaticamente analisados. As árvores de derivação são selecionadas e, se necessário, corrigidas pelos anotadores humanos. Sequências de palavras para as quais nenhuma árvore de derivação é encontrada são omitidas ou manualmente anotadas;
- **Corpora interpretados linguisticamente:** Ao contrário dos *treebanks*, onde somente a categoria sintática e a estrutura da sentença são anotadas, corpora interpretados linguisticamente buscam abranger anotações de outros tipos de informação linguística como, por exemplo, anotações semânticas.

De acordo com tipo de uso:

- **Treinamento:** modelos estatísticos para processamento da linguagem natural são treinados, em outras palavras, aprendem a partir de grandes corpora, tipicamente anotados;
- **Teste:** Corpora de teste são utilizados para avaliar modelos estatísticos após o treinamento;
- **Avaliação:** Corpora anotados são utilizados para avaliar componentes de sistemas de processamento da linguagem natural.

4.2.3. Texto Marcado

De acordo com SEIDEL (2012), um corpus de texto puro (texto sem marcações) por si só, pode ser valioso em informações, sendo possível extrair bastante conhecimento. Diversas técnicas fazem uso de um tipo especial de corpus, conhecido como *texto marcado* ou *corpus anotado*. Com textos marcados é possível obter informações sobre o domínio desejado e esse corpus é desenvolvido através de um processo conhecido como *anotação*, onde informações estruturais são inseridas no texto.

A marcação em textos – *Part-of-Speech* (POS) – visa atribuir marcas (*tags*) de discurso (ou seja, categorias morfossintáticas, tais como substantivo, verbo, adjetivo, entre outros) para palavras no contexto. A principal dificuldade é a ambiguidade léxica que existe na linguagem natural. Por exemplo: “process” e “programs” poderiam ser tanto substantivos como verbos (RAJMAN; BESANÇON, 1998).

Ainda conforme SEIDEL (2012) os textos marcados – ou tagueados – são empregados em técnicas computacionais que demandam processamento de propriedades não-observáveis em textos puros. Em alguns deles, somente as estruturas básicas são marcadas, tais como as fronteiras de sentenças e parágrafos. Porém outros compreendem uma carga de informação maior assim como toda a estrutura sintática. A marcação mais usual é a codificação das categorias das palavras.

A marcação da estrutura de um texto dispõe de vários esquemas. Um esquema bastante usado na anotação gramatical de um texto é a utilização de um caractere específico entre cada palavra e o código referente à sua categoria. O caractere específico geralmente empregado é uma barra “/” ou um sublinhado “_”.

Na Tabela 6 é apresentado um pequeno extrato de texto livre e em seguida o mesmo texto marcado com etiquetas morfológicas, extraído do corpus *Tycho Brahe* (GALVES; BRITTO, 2002). Já na Tabela 7 é mostrado um pequeno extrato de texto do corpora *Penn Parsed* (KROCH; TAYLOR, 2000), que abrange textos do idioma inglês, também marcados com etiquetas morfológicas (*Part-of-Speech* – POS – *tagging*, termo utilizado em inglês).

Tabela 6. Excerto de texto do corpus *Tycho Brahe* (GALVES; BRITTO, 2002).

Texto puro	Eu disponho para vós o Reino, como meu Pai o dispôs para mim;
Texto marcado	Eu/PRO disponho/VB-P para/P vós/PRO o/D Reino/NPR ./, como/CONJS meu/PRO Pai/NPR o/CL dispôs/VB-D para/P mim/PRO ;/.

Tabela 7. Excerto de texto do corpora *Penn Parsed* (KROCH; TAYLOR, 2000).

Texto puro	It befell in the days of Uther Pendragon, when he was king of all England, and so reigned, that there was a mighty duke in Cornwall that held war against him long time.
Texto marcado (POS tagged)	It_PRO befell_VBD in_P the_D days_NS of_P Uther_NPR Pendragon_NPR _, when_P he_PRO was_BED king_N of_P all_Q England_NPR _, and_CONJ so_ADV reigned_VBD _, that_C there_EX was_BED a_D mighty_ADJ duke_N in_P Cornwall_NPR that_C held_VBD war_N against_P him_PRO long_ADJ time_N ..

4.2.4. Extração da Informação

A área de Extração de Informação (EI) é uma importante tarefa na Mineração de Textos que visa descobrir e estruturar informações dispostas em documentos semiestruturados ou desestruturados. A EI tem sido amplamente estudada em várias áreas de pesquisa, incluindo o Processamento de Linguagem Natural, Recuperação de Informação (RI) e Mineração de Dados.

Segundo KUSHMERIK (1999), a Extração da Informação pode ser vista como a tarefa de identificar os fragmentos específicos de um documento que constituem o núcleo de seu conteúdo semântico.

O campo de Extração de Informação busca identificar e extraír, de modo automático, informações importantes em um documento ou coleção de documentos expressos em língua natural e estruturar tais informações para os padrões de saída, por exemplo em banco de dados ou textos em língua natural, com o propósito de facilitar sua manipulação e análise (GRISHMAN, 1997).

Os sistemas de EI contribuem para coletar dados em textos com relevância e valor para a temática abordada. A EI inicia suas atividades com a coleção de textos coletados, em seguida, transforma esses dados em informação de compreensão (COWIE; LEHNERT, 1996).

Extração de Informação não deve ser confundida com a área de Recuperação de Informação, a qual elege, de uma grande coleção, um subconjunto de documentos significativos requisitados pelo usuário através de uma consulta. A divergência entre os

objetivos dos sistemas de EI e RI pode ser declarado da seguinte forma: RI recupera documentos relevantes de uma coleção, ao passo que EI extrai informações relevantes dos documentos. Desse modo, as duas técnicas são complementares e, quando combinadas, podem elaborar ferramentas importantes para processamento de texto (GAIZAUSKAS; WILKS, 1998).

Ainda conforme GAIZAUSKAS e WILKS (1998), a área de Extração de Informação, no sentido contrário de Recuperação de Informação, teve um crescimento acelerado nas duas últimas décadas. Dois fatores foram fundamentais para esse crescimento: o aumento exponencial da quantidade de dados textuais oferecidos na forma digital, e a atenção dada a área através das Conferências de Entendimento de Mensagens (*Message Understanding Conference – MUC*) durante a década de 80.

A área de EI, dentre outras prerrogativas, está empenhada em identificar conteúdo a partir de textos não estruturados ou totalmente organizados (SILVA; BARROS; PRUDÊNCIO, 2005). Essa atribuição abrange a identificação de entidades, relacionamentos e contextos específicos, com o intuito de lapidar a coleção total de textos para representações mais compreensíveis de acordo com as necessidades humanas. A EI pode ser empregada em diversos domínios de aplicação, sendo que cada um possui particularidades. No caso da extração de entidades nomeadas específicas, palavras chave e de articulação devem ser introduzidas com participação humana para produzir resultados mais adequados para cada cenário (ELLOUMI et al., 2013).

De acordo com SAWARAGI (2007), a EI compreende a atividade de extração automática de informações em textos – como entidades, relações e atributos. Sua estrutura possui, por essência, o Processamento de Linguagem Natural – PLN e engloba os temas como a Aprendizagem de Máquina, Recuperação de Informação, Banco de Dados, Web, Análise de Documentos, sendo possível ser visto na Figura 28, possibilitando que consultas complexas possam ser executadas em dados não estruturados.

DODDINGTON et al. (2004) afirma que um sistema de EI, devido à grande quantidade de informação e à maneira como faz uso dos mais diferentes sistemas de informação, é capaz de transformar os diversos dados intrínsecos em informação com algum valor agregado para determinado cenário.

Na maior parte dos sistemas de EI, o usuário define *templates*²⁰ de extração, ou seja, modelos estruturados, como campos de uma base de dados, a serem preenchidos a partir do texto original pelo processo de extração, como apresentado na Figura 29. A representação da informação em modelos tem a vantagem de ser fortemente estruturada. Isto possibilita que, posteriormente, o processamento da informação possa ser feito por outras aplicações, como pacotes de bancos de dados tradicionais, aplicações comerciais e sistemas especialistas ou de redes neurais (COSTANTINO; MORGAN; COLLINGHAM, 1996).

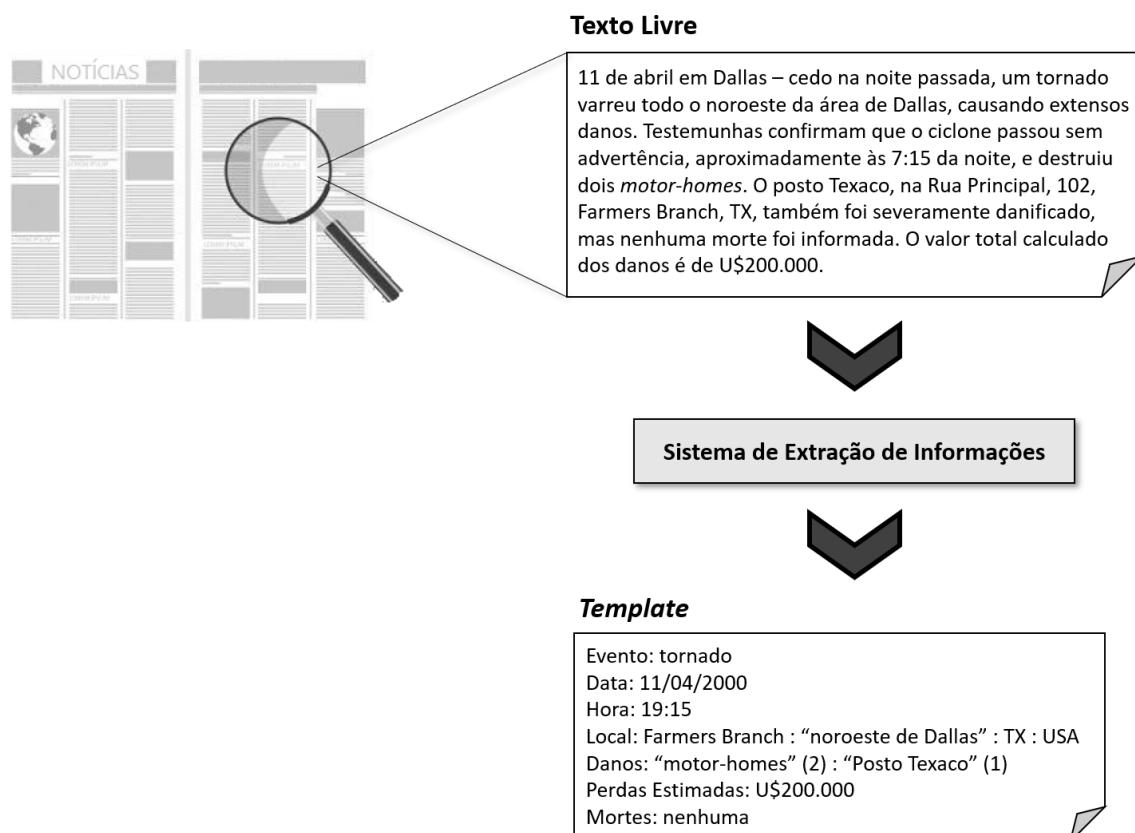


Figura 29. *Template* de um sistema de extração de informações. Adaptado de ZAMBENEDETTI (2002).

De modo a atender os requisitos e os objetivos indispensáveis na elaboração deste trabalho, a EI atua essencialmente na subárea de REN, na identificação e extração de entidades com base em fontes de dados textuais, em particular, elementos textuais das notícias em linguagem natural e na tarefa de transformação desses dados em uma informação mais adequada para gerar informações de conteúdo relevante a determinados contextos (ELLOUMI et al., 2013; WEIKUM et al., 2009).

²⁰ *Template* – ou "modelo de documento" – é um documento de conteúdo.

4.3. Reconhecimento de Entidades Nomeadas

Um componente básico na concepção de um sistema de EI é o REN ou Reconhecimento de Entidade Mencionada (REM), tradução usada pela comunidade de língua portuguesa para o original em inglês, *Named Entity Recognition* (NER) (BUNESCU, 2007). O termo Reconhecimento de Entidade Nomeada foi historicamente mais empregado no mercado a partir da década de 90, essencialmente pelo crescimento e progresso de uma quantidade de componentes de EI para os documentos publicados na internet (RIZZO; TRONCY, 2011).

O REN se baseia em tarefas de detecção e rastreamento com o intuito de coletar as entidades contidas no texto e classificá-las de acordo com a estrutura vigente e suas regras (DODDINGTON et al., 2004; ELLOUMI et al., 2013).

O REN, assim como a Extração de Relações (ER), são tarefas de suma importância em EI (JUNG, 2012). A primeira visa encontrar entidades nomeadas, incluindo nome de pessoas, instituições, lugares, entre outros; enquanto a segunda, consiste na detecção e caracterização de relações que envolvem as entidades nomeadas presentes no texto. Segundo NADEAU e SEKINE (2007), os termos que apresentam um ou mais designadores rígidos, num determinado texto, por exemplo, substantivos próprios, tais como nomes de pessoas, organizações e entidades locais definem as Entidades Nomeadas (EN).

Devido ao crescimento na ênfase nas técnicas de reconhecimento de linguagem natural, o REN se consolidou como um componente primordial para o campo de EI (JIANG, 2012). Atualmente, o REN é muito usado para extração de conhecimento específico em fontes de dados textuais e nas tarefas de classificação de tipos de categorias pré-definidas, como pessoa, organização e localização (MAGALHÃES, 2016).

De acordo com MOTA et al. (2007), o REN consiste na tarefa de identificar, em sua maior parte, nomes próprios, a partir de textos de forma livre e classificar essas ENs dentro de um conjunto de tipos de categorias pré-definidas, tais como pessoa, organização e local, as quais remetem a um referente específico.

Segundo SUREKA et al. (2009), o REN e a posterior classificação de tais entidades é uma técnica bastante empregada no PLN e baseia-se na identificação de nomes de entidades-chave presentes na forma livre de dados textuais. A entrada para o sistema de extração de EN é o texto de forma livre, e a saída é um conjunto das anotações, ou seja, grupo de caracteres extraídos de trechos do texto de entrada. A saída do sistema

de extração de EN é, basicamente, uma representação estruturada a partir da entrada de um texto não estruturado.

As soluções para o REN baseiam-se em padrões de regras trabalhadas de forma manual, pois exigem experiência humana além de um trabalho intenso para a criação de tais padrões (APPELT et al., 1993). O desenvolvimento de sistemas de REN tem o objetivo de aprender automaticamente esses padrões de dados etiquetados (CIRAVEGNA, 2001). Os modelos HMM (*Hidden Markov Models*) (BIKEL et al., 1997), ME (*Maximum Entropy*) (CHIEU; NG, 2003) e MEMM (*Maximum-Entropy Markov Model*) (BENDER; OCH; NEY, 2003; CURRAN; CLARK, 2003; FINKEL et al., 2005) são formalismos estatísticos de aprendizagem de máquina que desempenham o REN. Outros trabalhos sobre REN utilizam modelos matemáticos probabilísticos, como o CRF (*Conditional Random Field*) (AMARAL, 2013; CHATZIS; DEMIRIS, 2012; DA SILVA; CASELI, 2015; LAFFERTY; MCCALLUM; PEREIRA, 2001; MAGALHÃES, 2016; SETTLES, 2004), o qual é o modelo aplicado neste presente trabalho.

4.3.1. Conferências de REN

Visando aprimorar o estado da arte e estimular novas iniciativas de reconhecimento de entidades, conferências de avaliação de resultados foram criadas, como exemplo: *Message Understanding Conference* (MUC), *Automatic Content Extraction* (ACE), *Conference on Computational Natural Language Learning* (CoNLL) e, para o português, a conferência de reconhecimento de entidades denominada de avaliação conjunta na área do reconhecimento de entidades mencionadas em português (HAREM).

As conferências buscam avaliar os resultados obtidos entre os diversos sistemas, fazendo uso de métricas como precisão, abrangência e medida-F. Essas conferências proporcionaram grande avanço nas tarefas de reconhecimento de entidades ao longo dos anos (MAGALHÃES, 2016).

Na seção a seguir estão apresentadas as principais conferências e suas características, estudos e resultados relevantes.

4.3.1.1. MUC

O MUC, foi uma série de eventos que ocorreram entre 1987 (MUC-1) e 1998 (MUC-7), e tinha como propósito avaliar e promover o progresso nas pesquisas sobre extração de informações e padronizar a avaliação das tarefas dessa área (SEIDEL, 2012).

A segunda edição da conferência (MUC-2) aconteceu em 1989 e foi designada a extrair informações em mensagens navais. Nas duas edições seguintes, MUC-3 e MUC-4, o trabalho de extração foi concentrado em utilizar as notícias coletadas sobre os incidentes terroristas em países latino-americanos. O MUC-5 apresentou uma nova melhoria baseada na taxa erro e evoluções na medida de desempenho (GRISHMAN; SUNDHEIM, 1996).

O MUC-6, realizado em novembro de 1995, foi o sexto desta série e foi nessa versão da conferência onde o termo “entidade” foi cunhado e foram introduzidas as tarefas relacionadas ao reconhecimento de entidades mencionadas para o idioma inglês (GRISHMAN; SUNDHEIM, 1996). Nessa versão da conferência também, a tarefa de REM era vista como tendo o objetivo de reconhecer nome de pessoas, organizações, nome de lugares, expressões temporais e certos tipos de expressões numéricas (SEIDEL, 2012), fazendo uso da entidades apresentadas a seguir (MIKHEEV; GROVER; MOENS, 1998):

- **ENAMEX:** composta por nomes próprios, organizados nas categorias pré-definidas como pessoa (PERSON), organização (ORGANIZATION) e local (LOCATION);
- **TIMEX:** formada pelas representações temporais de data (DATE) e hora (TIME);
- **NUMEX:** constituída por expressões monetárias representadas em valores monetários (MONEY) ou percentuais (PERCENT).

Para o processo de avaliação era necessário seguir algumas regras de anotação do texto. A seguir temos um exemplo, extraído de MIKHEEV, GROVER e MOENS (1998), de como um texto deveria ser anotado para o MUC:

... <ENAMEX TYPE="ORGANIZATION">Grupo Televisa</ENAMEX> ...

Do exemplo temos que o trecho “Grupo Televisa” é uma entidade mencionada do tipo ENAMEX subcategorizada como uma organização (ORGANIZATION).

Em 1998, ocorreu a sétima e última versão da MUC (MUC-7), com a participação de 200 artigos e com resultado expressivo (mais de 90%) na métrica denominada “medida-F” para as atividades de reconhecimento de entidades mencionadas na língua inglesa (MIKHEEV; GROVER; MOENS, 1998).

4.3.1.2. ACE

A ACE teve início no ano de 1999, alcançando maior notoriedade durante os anos de 2001 e 2002 com a adição das tarefas de identificação e classificação das entidades, tarefa denominada de em *Entity Detection and Tracking* (EDT), além de englobar categorias não mencionadas anteriormente pela MUC (DODDINGTON et al., 2004).

A etapa seguinte da conferência aconteceu nos anos de 2002-2003, no estabelecimento de relação entre pares de entidades, ou relacionamento entre as entidades, tarefa denominada como *Relation Detection and Characterization* (RDC) (DODDINGTON et al., 2004).

No ano de 2004 as tarefas relacionadas a extração e caracterização da presença de eventos, denominadas *event extraction*, foram apresentadas. Outro avanço dessa versão da conferência foi o desenvolvimento da capacidade de extrair o significado de fontes de diversos tipos de fontes de dados, incluindo texto, áudio e imagens (DODDINGTON et al., 2004).

4.3.1.3. CoNLL

A CoNLL é uma série de conferências, com início em 1997, que tem como objetivo promover a pesquisa e avaliação em diversas áreas de PLN (SEIDEL, 2012).

A conferências de 2002 (CoNLL-2002) e 2002 (CoNLL-2003), tinha como tarefa compartilhada a avaliação de sistemas de REN independente de linguagem. Nestes eventos, quatro tipos de EN foram consideradas: pessoas, lugares, organizações e nomes de entidades diversas que não se encaixam em nenhuma das categorias anteriores (SEIDEL, 2012).

Uma das principais tarefas dos participantes da CoNLL-2003 era descobrir como melhorar o desempenho dos seus sistemas de REN fazendo uso de alguns recursos adicionais providos pelo evento, tais como listas de EN e textos não-anotados. A avaliação dos sistemas é baseada na comparação da Medida-F, onde uma EN é considerada correta apenas se o texto exato é identificado e a sua classificação for correta. A precisão é o percentual de EN corretas que são encontradas pelo sistema. A cobertura é o percentual de entidades presentes no texto solução que são encontradas pelo sistema (SEIDEL, 2012).

4.3.1.4. HAREM

O HAREM é uma avaliação conjunta de sistemas de Reconhecimento de Entidades Mencionadas organizada pela Linguateca²¹, iniciado em 2005 e foi a primeira e mais importante conferência de reconhecimento de entidade em língua portuguesa (MAGALHÃES, 2016; SEIDEL, 2012).

Segundo CARDOSO (2006), o HAREM teve como motivação o fato de sentirem que os eventos de avaliação de REN anteriores não haviam apresentado a tarefa com profundidade suficiente, e o objetivo de reunir a comunidade científica em torno de outro evento de avaliação dentro do processamento da língua portuguesa. Nesse intuito, os participantes obtiveram um papel ativo na organização da conferência, tendo colaborado na criação das diretrivas e na anotação das coleções.

O HAREM difere da MUC principalmente pela “liberdade” na tarefa de encontrar entidades. Enquanto em MUC as entidades se mantinham restritas aos tipos: pessoa, localização e organização, no HAREM os trabalhos apresentados estavam interessado em todos os nomes próprios (SANTOS, 2007).

A primeira conferência HAREM ocorreu no ano de 2004 e apresentou dois aspectos principais (MAGALHÃES, 2016):

- A classificação e identificação de um termo;
- A viabilidade de que um determinado termo possa ser classificado em mais de uma categoria, dado que seu contexto seja indeterminado de ser solucionado.

Em 2008, aconteceu a segunda conferência HAREM, apresentando um total de 10 categorias e 42 tipos de elementos identificáveis, conforme apresentadas na Tabela 8.

As categorias retratadas na segunda HAREM são de alta granularidade, isto é, nível macro dos elementos, com bastantes subcategorias. A Figura 30 ilustra uma representação das categorias do HAREM, sendo que: as categorias, tipos e subtipos representados nas caixas com contorno sólido preto só existem no Segundo HAREM, já as categorias, tipos e subtipos representados nas caixas com contorno pontilhado só existem no Primeiro HAREM.

²¹ A Linguateca é um centro de recursos (distribuídos) para o processamento computacional da língua portuguesa. Tem como objetivos facilitar o acesso aos recursos já existentes tais como córpus, enciclopédias e textos em português; desenvolver em colaboração com os interessados, os recursos mais prementes; e organizar avaliações conjuntas que envolvam a comunidade científica de interesse em torno do PLN. Pode ser acessado em: <https://www.linguateca.pt/>.

Tabela 8. Categorias do segundo HAREM. Adaptado de (OLIVEIRA; MOTA; FREITAS, 2008).

Categoría	Tipos
Abstração	Disciplina, Estado, Ideia, Nome, Outro.
Acontecimento	Efemeridade, Evento, Organizado, Outro.
Coisa	Classe, MembroClasse, Objeto, Substancia, Outro.
Local	Físico Humano, Virtual.
Obra	Arte, Plano, Reproduzida, Outro.
Organização	Administração, Empresa, Instituição, Outro.
Pessoa	Cargo, GrupoCargo, GrupoInd, GrupoMembro, Individual, Membro, Povo, Outro.
Tempo	Duração, Frequência, Genérico, TempoCalend, Outro.
Valor	Classificação, Moeda, Quantidade, Outro.
Outro	

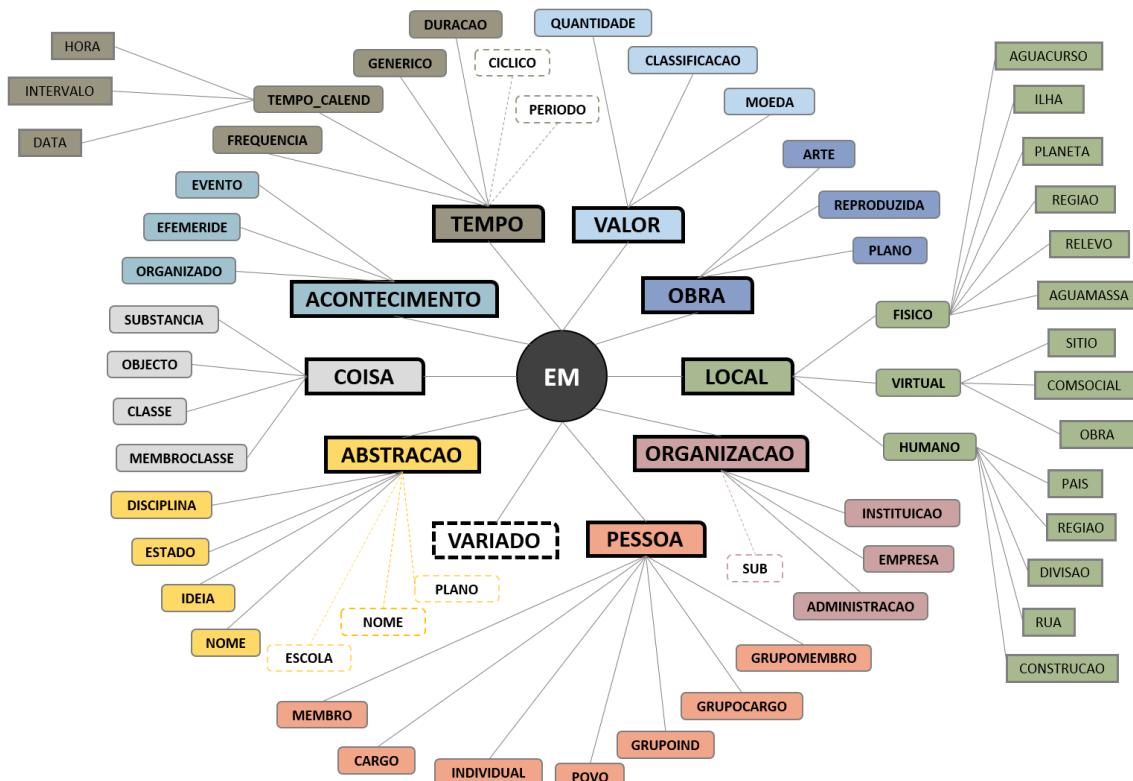


Figura 30. Árvore de categorias do HAREM. Adaptado de (MOTA; SANTOS, 2008).

4.3.2. Resolução e Anotação de Homônimos

A tarefa de resolução de homônimos se baseia em métodos para encontrar determinado homônimo em um conjunto finito de texto, lidando com subjetividades e fatores de ambiguidade de determinação de classificação do termo (MAGALHÃES, 2016).

Diversas soluções foram propostas para a categorização e anotação exata das entidades, para evitar que uma entidade possa ser referenciada como outra – ocorrência de ambiguidade – evitando que a máquina de aprendizado lide com problemas de dualidade de categorias (DODDINGTON et al., 2004; RATINOV; ROTH, 2009; ZHANG; TSAI, 2009).

Um exemplo frequente são os problemas relacionados a categorização ambígua das fontes de dados, onde uma entidade em uma determinada fonte é categorizada como “Categoria X” e, em outra fonte, como “Categoria Y”, conforme a Tabela 9 ilustrando, como exemplo, a frase “*Dr. Alan Grant*”.

Tabela 9. Exemplo de Categorização Ambígua de Entidades.

Conjunto de Dados I		Conjunto de Dados II	
Palavras	Categorias pré-definidas	Palavras	Categorias pré-definidas
<i>Dr.</i>	Título	<i>Dr.</i>	Título
<i>Alan</i>	Nome	<i>Alan</i>	Nome
<i>Grant</i>	Sobrenome	<i>Grant</i>	Verbo

Nesse caso a palavra “*Grant*” teria uma classificação ambígua, resultando em problema para a categorização. RATINOV e ROTH (2009) apresentam que uma forma de auxiliar na resolução dos problemas de categorização de homônimos é analisar contextos distintos, atribuindo especificidade ao contexto, ou seja, analisar os grupos de entidades em separado.

O tagueamento (em inglês: *tagging*) – marcação POS (*Part-of-Speech*) – de homônimos são necessários para classificar um termo, constituindo parte importante para o sucesso da tarefa de REN. De modo geral, essa é uma tarefa manual e que exige grande esforço e dedicação, sendo tradicionalmente feita por especialistas, como na HAREM, onde sua coleção de treinamento para a entidade LOCAL foi realizada por quatro pessoas (MAGALHÃES, 2016).

Para o problema da ambiguidade, segundo o estudo de RATINOV e ROTH (2009), após a conclusão da etiquetagem POS e da segmentação das sentenças, determinou-se como as ENs poderiam ser identificadas, fazendo uso de três tipos de notações: IO, BIO e BILOU. A primeira é a mais popular para a tarefa, principalmente pela significativa rapidez em marcar as entidades nos textos, tendo o seguinte significado: I (*Inside*) uma ou mais palavras que se definem como uma EN; e O (*Outside*) a palavra não é uma EM. A segunda possui uma classificação a mais que a primeira, tendo o seguinte significado: B (*Begin*) significa a primeira palavra da EN; I (*Inside*) uma ou mais palavras que se localizam entre as entidades; e O (*Outside*) a palavra não é uma EN. Já a segunda notação tem a mesma descrição do BIO, acrescentando-se as seguintes particularidades: L (*Last*) a última palavra reconhecida como EN e U (*Unit*) quando a EN for uma única palavra.

4.3.3. Métodos Probabilísticos

Este tópico descreve os dois métodos mais relevantes, baseados em modelos probabilísticos, descritos na literatura. O primeiro método, o HMM (*Hidden Markov Models*), é descrito brevemente e em seguida o CRF (*Conditional Random Field*) é explicado de forma mais detalhada.

Nesse trabalho, o método utilizado para o Reconhecimento de Entidades Nomeadas foi o CRF, visto que a metodologia utilizada nesse trabalho é a supervisionada, permitindo a transferência de conhecimento por meio da elaboração de uma base de treinamento. Além disso, o CRF possui uma capacidade de resposta mais rápida do que outras técnicas estudadas (MOTA; SANTOS, 2008) e provê um maior grau de sucesso nas tarefas de REN (FINKEL; GRENGER; MANNING, 2005).

4.3.3.1. *Hidden Markov Models*

O *Hidden Markov Models* se trata de uma extensão da cadeia de Markov, adicionando mais um processo estocástico ao modelo, isto é, o HMM é um modelo estocástico que modela dois processos estocásticos, sendo um desses processos, oculto. Diferentemente da cadeia de Markov, as observações do modelo seguem uma função probabilística (RABINER, 1989).

O HMM tem como objetivo simular uma fonte de sinal a partir de um conjunto de observações. Essa fonte de sinal é um sistema X que, dado um conjunto de observações

$O = \{O_1, O_2, \dots, O_n\}$, produz os sinais $Q = \{Q_1, Q_2, \dots, Q_n\}$. O modelo é composto, essencialmente: pelas observações (de diversos tipos), pelos números de estados da cadeia de markov, pela função de transição entre estados, pela função probabilística de emissão de um sinal dado um estado e pela função probabilística do estado inicial (SILVA, 2012).

O HMM tem sido amplamente empregado em tarefas de PLN tais como reconhecimento de *Part-of-Speech* (POS) e o REN. Em tarefas de PLN normalmente as observações são as sequências de palavras contidas em um texto e os sinais são as informações semânticas relativas às palavras observadas (SILVA, 2012).

4.3.3.2. Conditional Random Field

O *Conditional Random Field* é um modelo matemático probabilístico que tem o objetivo de etiquetar e segmentar dados sequenciais, baseados na abordagem de uma distribuição condicional $p(Y|X)$ com um modelo gráfico associado. A variável X é um vetor de variáveis aleatórias de entrada e Y é um vetor de variáveis aleatórias de saída. Então $p(Y|X)$ é a probabilidade de dado como entrada o vetor X obter a saída Y (LAFFERTY; MCCALLUM; PEREIRA, 2001; SUTTON; MCCALLUM, 2007).

O CRF tem aplicações em diversas áreas tais como, processamento de texto (PENG; MCCALLUM, 2004; TASKAR; ABBEEL; KOLLER, 2012), bioinformática (LIU et al., 2005; SATO; SAKAKIBARA, 2005) e visão computacional (KUMAR; HEBERT, 2004; XUMING HE; ZEMEL; CARREIRA-PERPINAN, 2004).

Exemplos da importância do CRF podem ser encontrados nos trabalhos de BATISTA et al. (2010), SILVA (2012), AMARAL e VIEIRA (2014), DA SILVA e CASELI (2015) e MAGALHAES (2016).

4.3.3.2.1. Definição

O problema de REN pode ser formulado como um problema de encontrar uma distribuição condicional. Seja $X = [x_1, x_2, \dots, x_{n-1}, x_n]$ uma sequência de palavras em um texto, onde x_j é uma palavra localizada na posição j do texto. Seja $Y = [y_1, y_2, \dots, y_{n-1}, y_n]$ uma sequência contendo um rótulo para cada palavra do vetor X . Então temos que $p(Y|X)$ diz a probabilidade do texto representado por X ter a rotulação Y (SILVA, 2012).

Em seu trabalho, SILVA (2012) cita o seguinte exemplo. Seja o texto “*O Brasil investe R\$ 2.000.000.000 no estado do Rio de Janeiro*”, então temos:

$$X = \{O, Brasil, investe, R\$, 2.000.000.000, no, estado, do, Rio, de, Janeiro\}$$

$$Y_1 = \{O, LOCAL, O, O, O, O, O, LOCAL, LOCAL, LOCAL\}$$

$$Y_2 = \{O, ORGANIZACAO, O, O, O, O, O, LOCAL, LOCAL, LOCAL\}$$

$$Y_3 = \{O, O, O, O, O, O, O, LOCAL, LOCAL, LOCAL\}$$

O vetor X representa o texto e os vetores Y_1 , Y_2 e Y_3 são possíveis rotulações de X . No texto apresentado o termo “Brasil” é uma organização e o termo “Rio de Janeiro” é um local. Dessa forma $p(Y_2|X)$ tem que ser maior do que $p(Y_1|X)$ e $p(Y_3|X)$.

Dado X e uma distribuição condicional $p(Y|X)$ temos que a melhor rotulação de X é o vetor Y_k tal que $p(Y_k|X)$ é maior do que qualquer outro Y_j com $p(Y_j|X)$.

A dificuldade da modelagem dada acima é encontrar a distribuição $p(Y|X)$ e dada essa distribuição calcular a melhor rotulação Y dado o vetor de entrada X . Para isso, é utilizado um caso especial do *Conditional Random Field* que é o *Linear-Chain Conditional Random Field*.

SUTTON e MCCALLUM (2007) definem esse caso especial como sendo X e Y vetores de variáveis aleatórias, $\Lambda = \{\lambda_k\} \in \mathbb{R}^k$ um vetor de pesos, $\{f_k(y, y', x_t)\}_{k=1}^k$ um conjunto de funções características e t uma posição do vetor X . Então o *Linear-Chain Conditional Random Field* é uma distribuição $p(Y|X)$ que tem a seguinte forma:

$$p(y|x) = \frac{1}{Z(x)} \exp \left\{ \sum_{k=1}^K \lambda_k f_k(y_t, y_{t-1}, x_t) \right\}$$

Onde $Z(x)$ é uma função de normalização dada por:

$$Z(x) = \sum_y \exp \left\{ \sum_{k=1}^K \lambda_k f_k(y_t, y_{t-1}, x_t) \right\}$$

De acordo com BIKEL et al. (1997), uma importante característica deste modelo, bem como todos os derivados do *Conditional Random Field*, é que não há necessidade de se conhecer a distribuição $p(X)$, que pode ser uma distribuição demasiadamente complexa.

Dada a definição, ainda conforme SUTTON e MCCALLUM (2007), é preciso definir o vetor de funções características $f = (f_1, f_2, f_3, f_4, \dots, f_k)$ e o vetor de pesos $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \dots, \lambda_k)$. O vetor Λ aplica pesos ao vetor de funções de características,

dessa forma algumas funções têm maior impacto do que outras na contabilização final da probabilidade estimada.

As funções características possuem o formato $f_k(y_t, y_{t-1}, x_t)$, onde y_t é a rotulação da palavra na posição t do vetor de entrada X , y_{t-1} é a rotulação da palavra na posição $t-1$ do vetor de entrada X e x_t é o vetor formado pelas palavras de X que são relevantes para o cálculo do rótulo y_t .

4.3.3.2.2. Estimação de Pesos

Segundo SUTTON e MCCALLUM (2007), a distribuição $p(Y|X)$ depende dos valores dados ao vetor de pesos $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \dots, \lambda_k)$ como visto acima. Este vetor tem o objetivo de ponderar o impacto de cada função característica. Maiores valores de λ_j implicam em maior importância dada à função característica f_j .

O cálculo do vetor de pesos é realizado por meio de um treino supervisionado sendo $\theta = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \dots, \lambda_k)$ e $D = \{x^{(i)}, y^{(i)}\}^N$ o conjunto de treinamento com N exemplos de rotulação. Dessa forma, temos que a escolha ótima de Λ é a escolha de θ que maximize o somatório dos logaritmos das probabilidades $p(y^{(i)}|x^{(i)})$ conforme explicitado abaixo:

$$l(\theta) = \sum_{i=1}^N \log p(y^{(i)}|x^{(i)})$$

Uma vez que as funções características estejam definidas e que o vetor de pesos $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \dots, \lambda_k)$ tenha sido calculado, a distribuição $p(Y|X)$ pode ser calculada para todo Y e X .

4.3.4. Medidas para Avaliação e Desempenho

A necessidade de medir os diversos resultados obtidos nas atividades entre os projetos participantes das conferências de REN se tornou evidente devido aos avanços dos trabalhos. Com isso, diversas medidas – ou métricas – foram desenvolvidas de modo a permitir medir os resultados, dentre elas destacam-se as medidas de Precisão (*Precision*), Abrangência (*Recall*) e Medida-F (*F-measure* ou *F-score*) (COWIE; LEHNERT, 1996; KONKOL, 2012).

De modo a obter os resultados das métricas é necessária a classificação dos termos (objetos), que podem ser denotados por dois valores, positivos e negativos, e suas relações que resultam em quatro possíveis tipos de classificação (KONKOL, 2012).

- **Verdadeiros Positivos (VP):** são itens relevantes ao contexto que corretamente são identificados como positivos ou relevantes;
- **Verdadeiros Negativos (VN):** são itens irrelevantes ao contexto que corretamente são identificadas como falsos ou irrelevantes;
- **Falsos Positivos (FP):** são itens irrelevantes ao contexto que erroneamente são identificados como positivos ou relevantes;
- **Falsos Negativos (FN):** são itens relevantes ao contexto que erroneamente são identificadas como falsos ou irrelevantes.

Uma representação da relevância da classificação dos termos é ilustrada na Figura 31. Os termos com delimitação na cor preta são denominados relevantes (VP, FP, FN), pois estes são utilizados para mensurar o modelo proposto pela máquina de aprendizado, através do cálculo baseado nas medidas (MAGALHÃES, 2016).

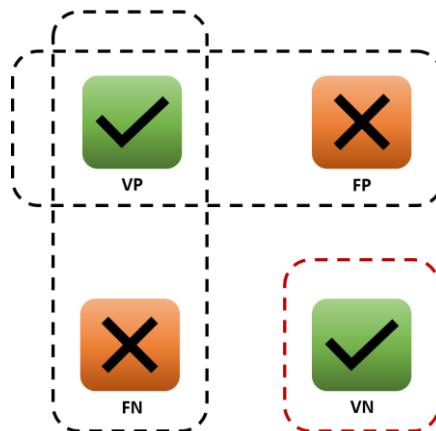


Figura 31. Relevância dos termos. Adaptado de (MAGALHÃES, 2016).

A Figura 32 apresenta essa classificação, onde as curvas representam a distribuição de objetos positivos e negativos e a linha tracejada ilustra a divisa da decisão do classificador. As áreas marcadas como FN e FP são alguns objetos marcados erroneamente (KONKOL, 2012).

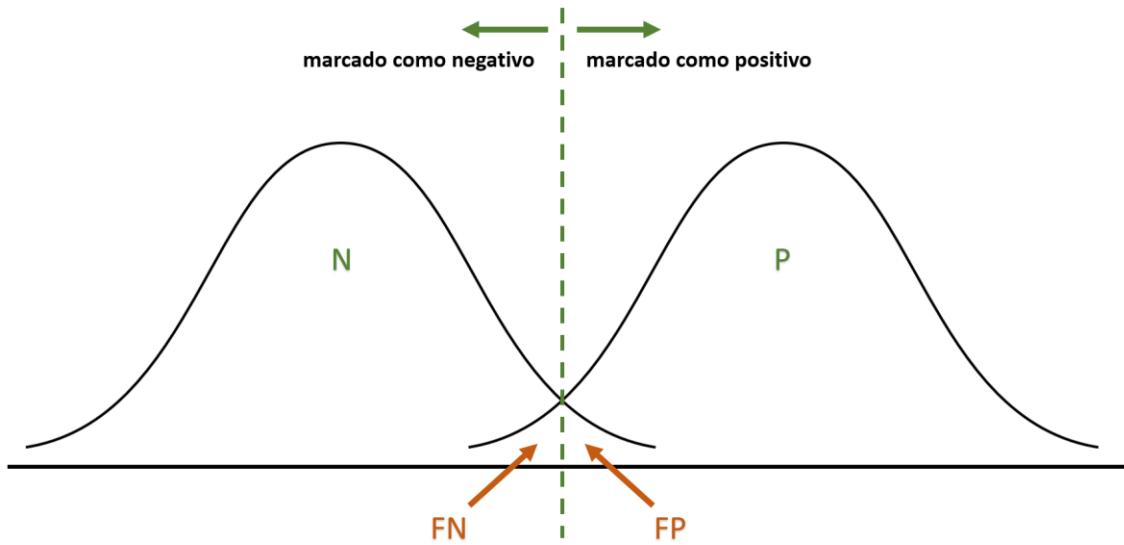


Figura 32. Fronteiras de Classificação dos Termos - Baseado de *Precision* e *Recall*.
Adaptado de (KONKOL, 2012).

4.3.4.1. Abrangência

A Abrangência (do inglês, *Recall*) – em alguns estudos pode ser vista como Cobertura – é uma medida que pode ser definida como a quantidade total de itens relevantes identificados/extraidos corretamente sobre o número total dos itens relevantes existentes no corpus (KONKOL, 2012).

$$\text{Abrangência} = \frac{\text{Verdadeiros Positivos (VP)}}{\text{Verdadeiros Positivos (VP)} + \text{Falsos Negativos (FN)}}$$

4.3.4.2. Precisão

A Precisão (do inglês, *Precision*) pode ser definida como a quantidade de itens corretamente extraídos sobre todos os itens extraídos (KONKOL, 2012).

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos (VP)}}{\text{Verdadeiros Positivos (VP)} + \text{Falsos Positivos (FP)}}$$

4.3.4.3. Medida-F

A Medida-F (do inglês, *F-measure* ou *F-score*), mais especificamente a Medida-F₁, é uma média harmônica da Abrangência e Precisão que determina a acurácia, fazendo a contagem dos objetos de acordo com a classificação de seus termos em P, N, FP, FN (KONKOL, 2012; MAGALHÃES, 2016).

$$\text{Medida-F}_1 = \frac{2 \cdot \text{Precisão} \cdot \text{Abrangência}}{\text{Precisão} + \text{Abrangência}}$$

4.3.5. Aprendizado de Máquina

Segundo NADEAU e SEKINE (2007), um fator essencial da tarefa de Reconhecimento de Entidades Nomeadas é a capacidade de reconhecer as entidades anteriormente desconhecidas. Inicialmente os algoritmos de REN eram baseados em regras escritas manualmente que apontavam a existência de uma determinada entidade dado um contexto.

Por exemplo, se um substantivo capitalizado é precedido do termo “no” então este substantivo é um Local ou se um substantivo capitalizado é precedido do termo “diz” então este substantivo é uma Pessoa (SILVA, 2012).

Todavia, a escrita manual de regras conta com a desvantagem de ser tediosa, pois necessita da escrita de muitas regras, e tais regras precisam atender aos mais diversos cenários possíveis que dificilmente podem ser descobertos manualmente. Desse modo, diversos estudos foram realizados na tentativa de desenvolver métodos de aprendizado automatizado (SILVA, 2012).

O aprendizado de máquina (do inglês, *machine learning*) pode ser definido como a área computacional, mais precisamente de Inteligência Artificial, cujo objetivo é o desenvolvimento automático de conhecimento computacional com base em outras formas de aprendizado (MONARD; BARANAUSKAS, 2003).

Os métodos empregados na elaboração de sistemas de reconhecimento de entidades são divididos em três categorias que serão detalhadas a seguir: Aprendizado Supervisionado, Aprendizado Semissupervisionado e Aprendizado Não Supervisionado.

4.3.5.1. Aprendizado Supervisionado

A técnica de aprendizado supervisionado tem sido a técnica mais utilizada na resolução da tarefa de Reconhecimento de Entidades Nomeadas (MØLLER, 1993; MONARD; BARANAUSKAS, 2003; NADEAU; SEKINE, 2007). O conceito dos métodos baseados em aprendizado supervisionado, cuja arquitetura simplificada é ilustrada na Figura 33, se baseia em extrair de uma coleção de treinamento as características necessárias para a classificação correta das entidades nomeadas em textos desconhecidos (SILVA, 2012).

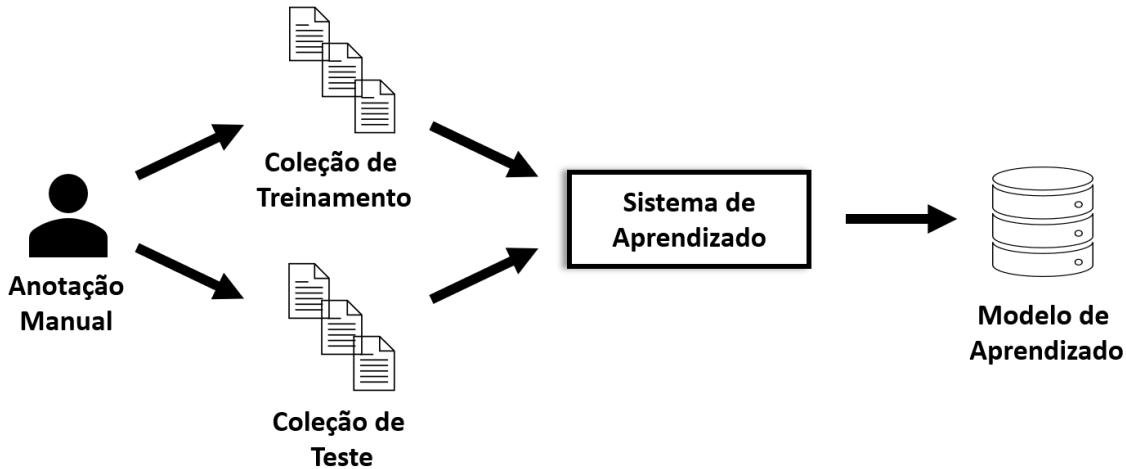


Figura 33. Arquitetura simplificada do Aprendizado Supervisionado. Adaptado de (SILVA, 2012).

Nesta técnica as seguintes etapas são realizadas: um ou mais especialistas avaliam um conjunto de textos e, para cada texto, os especialistas anotam manualmente as suas entidades nomeadas de forma a criar exemplos. No passo seguinte, o conjunto de textos anotados é dividido em um conjunto de treino e de teste. Destes, o primeiro é utilizado para transferir o conhecimento do especialista para o sistema de aprendizado quanto que o segundo é utilizado para validar a generalização do aprendizado obtido pelo sistema por meio do conjunto de treino. Para que a não seja afetada a qualidade da validação, é primordial que o conjunto de treino e treinamento sejam diferentes, evitando que o sistema memorize ao invés de generalizar (SILVA, 2012).

Dentre os métodos mais conhecidos que utilizam o aprendizado supervisionado, estão listados: Árvores de Decisão (*Decision Trees*) (SEKINE, 1998), *Hidden Markov Models* (BIKEL et al., 1997), *Maximum Entropy Models* (BORTHWICK et al., 1998) e *Conditional Random Fields* (MCCALLUM; LI, 2003).

4.3.5.2. Aprendizado Semissupervisionado

O aprendizado semissupervisionado é o segundo mais utilizado na tarefa de Reconhecimento de Entidades Nomeadas (SILVA, 2012) e é derivada do aprendizado supervisionado, diferenciando pelo fato de necessitar de um conjunto menor de dados para treinamento. As tarefas iniciais de entrada são minimizadas nesse processo, decorrente da minimização da quantidade de elementos necessários para as etapas de treinamento e testes (MAGALHÃES, 2016).



Figura 34. Arquitetura simplificada do Aprendizado Semissupervisionado. Adaptado de (SILVA, 2012).

Nesse tipo de aprendizado, cuja arquitetura simplificada é representada na Figura 34, um ou mais especialistas elaboram uma coleção de treino e teste contendo as entidades nomeadas anotadas manualmente da mesma maneira que no aprendizado supervisionado, porém com um tamanho menor. A partir do conjunto reduzido de treinamento este pode ser automaticamente expandido (SILVA, 2012).

A qualidade da coleção de treinamento está mais associada à variedade de contextos do que à variedade de nomes, desse modo, com a coleção de textos não anotada sendo somada ao sistema de aprendizado, é possível expandir o conjunto de treinamento, por conta de entidades com nomes que raramente podem ser atribuídos a mais de um tipo de entidade.

Como exemplo, a expressão “*Phil Collins*” muito provavelmente é uma entidade do tipo Pessoa, então todas as sentenças na coleção não anotada que possuam esta expressão podem constituir um novo exemplo para o treinamento. Dessa forma, expandindo o conjunto de exemplos já que novos contextos podem ser adicionados à coleção.

Alguns dos métodos que fazem uso do aprendizado semissupervisionado podem ser encontrados em: BRIN (1999), COLLINS (2002), COLLINS e SINGER (1999) e RILOFF e JONES (1999).

4.3.5.3. Aprendizado Não Supervisionado

O aprendizado não supervisionado – ou aprendizado por observação e descoberta – se baseia na extração de informação sem supervisão humana e sem a necessidade de um

corpus de treinamento para a tarefa de Reconhecimento de Entidades Nomeadas (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

De acordo com ETZIONI et al. (2005), “como os sistemas de extração não supervisionados não necessitam de intervenção humana, eles podem, recursivamente, descobrir novas entidade, relações, atributos e instâncias, de forma escalável e totalmente automatizada”.

Nesse tipo de aprendizado os parâmetros existentes no contexto em estudo são adequados ao conjunto definido de dados, de modo a melhor resumir regularidades encontradas nos dados (MAGALHÃES, 2016).

ALPAYDIN (2014) enfatiza que o objetivo desse tipo de aprendizado é encontrar regularidades com a amostra de dados, ou seja, uma estrutura dos dados de entrada, tais como certos padrões que ocorrem mais do que outros e observar o que geralmente ocorre e o que não ocorre.

Dentre os algoritmos mais conhecidos, estão listados os algoritmos de clusterização – ou agrupamento – como o *K-Means* (HARTIGAN; WONG, 1979) e de associação como o Apriori (CHEUNG et al., 1996).

Capítulo 5 – O *framework* NERMAP

O *framework* NERMAP foi pensado para atender os especialistas de Prospecção Tecnológica de forma a apoiar e facilitar o processo de *Roadmapping* Tecnológico, semiautomatizando seu uso e permitindo ao pesquisador analisar grande volume de documentos prospectivos, otimizando tempo e custos durante um estudo de FTA.

Este trabalho se utiliza do *framework* de *Design Science* (MARCH; SMITH, 1995), composto pelos processos de construção e avaliação de conceitos, modelos, métodos e instanciações para a produção de um artefato tecnológico. Os conceitos se configuram os elementos e as questões relevantes ao processo de *Roadmapping* Tecnológico, que têm estreita ligação com o contexto e os objetivos do estudo de FTA no qual se insere. O modelo representa a estrutura que compõe o processo e sua interface com a respectiva instância de prospecção tecnológica da qual faz parte. O método complementa o modelo desenvolvido, descrevendo o processo de *Roadmapping* Tecnológico em si através do detalhamento das interações necessárias à sua execução e objetivos esperados. A instanciação consiste na materialização dos conceitos, modelos e métodos em um produto, acessível e utilizável.

Os próximos tópicos definem de maneira detalhada os artefatos mencionados acima, para o objetivo final da construção do sistema NERMAP.

5.1. Conceitos e Modelo

Este tópico tem o objetivo de descrever, à luz do *Design Science*, a primeira e a segunda partes da construção do artefato (conceitos e modelo), de modo a definir os elementos necessários para o sistema NERMAP (instanciação), capaz de semiautomatizar o processo de *Roadmapping* Tecnológico, tendo como resultado o *roadmap* – seu produto.

Os Usuários do NERMAP são cadastrados através do módulo de Gerenciamento de Usuários, podendo ser orientado a grupos. Partindo deste princípio, TRMs podem ser criados ou compartilhados através do módulo de Gerenciamento de TRM, dando início ao processo para gerar o artefato final, o *roadmap*, como ilustrado na Figura 35.

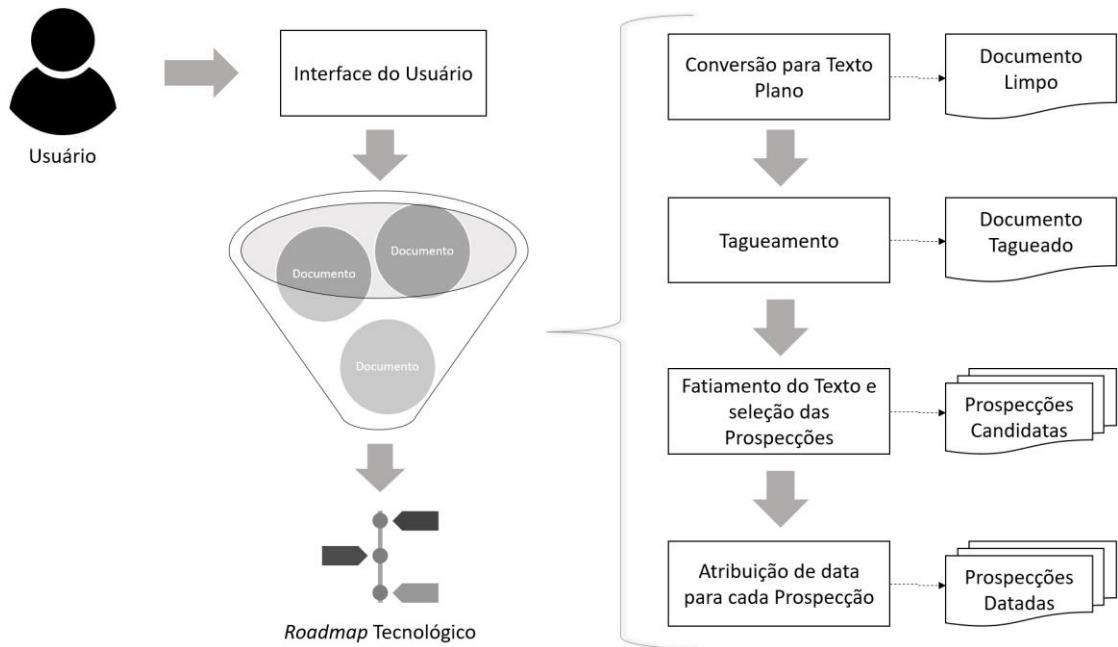


Figura 35. Modelo de geração do *roadmap*.

O usuário carrega os documentos relacionados ao contexto de prospecção tecnológica através da interface do módulo de Gerenciamento de Documentos, em consonância com as identidades e papéis (*roles*) para cada usuário definidas pelo módulo de Gerenciamento de TRM. Esses documentos são adicionados de acordo com o ano limite do TRM e sua área de pesquisa, reproduzindo a estrutura envolvida no processo de um TRM.

Os documentos, então, passam por etapas de processamento para serem convertidos para texto plano e, em seguida, serem tagueados através do Reconhecimento de Entidades Nomeadas, descrito de forma detalhada na subseção 4.3, fazendo parte do módulo de Processamento de Documentos.

Após o tagueamento, O TRM é construído, no módulo de Processamento do TRM. O texto é fatiado e as prospecções candidatas são separadas, sendo datadas de acordo com o escopo e a data atribuída a cada sentença. As prospecções candidatas e posteriormente datadas podem ser definidas em três categorias: *Completa*, ou seja, o texto selecionado representa o acontecimento previsto em sua totalidade, provendo a também data do mesmo; *Parcial*, ou seja, apresenta uma prospecção incompleta, sendo necessário refinamento manual; *Inválida*, ou seja, um trecho que não se configura como prospecção.

Cada TRM pode ser exposto e analisado como um TRM único – de cada documento – ou um composto de TRMs formando um único *roadmap*, como representado na Figura 36.

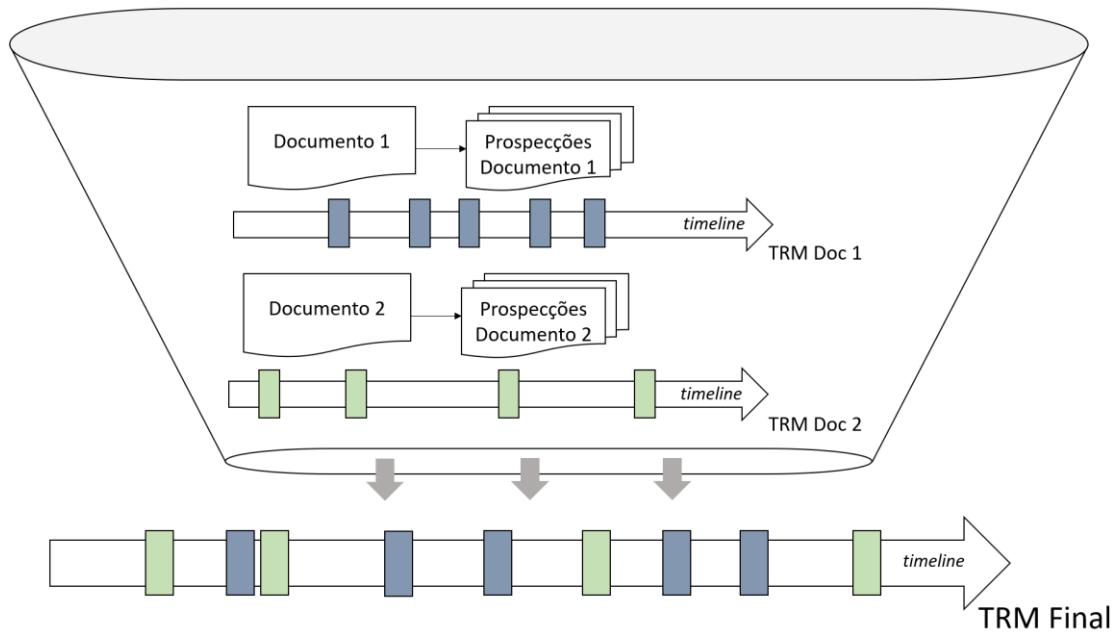


Figura 36. Composição de TRMs.

O Usuário pode refinar cada *roadmap* adicionando, editando ou excluindo prospecções. Além disso, filtros podem ser aplicados para obter *roadmaps* direcionados para um relacionado tema. Por exemplo, ao aplicar o termo “covid” como filtro em um TRM para a área macro de Medicina, somente serão adicionados ao *roadmap* prospecções que contenham o termo “covid”.

5.1.1. Processamento de Documentos

A principal etapa do modelo NERMAP é o processamento de documentos, onde os arquivos são convertidos para texto plano usando bibliotecas de *parser*²² e, em seguida, é aplicada a técnica de Reconhecimento de Entidades Nomeadas para o tagueamento do texto.

5.1.1.1. Stanford NER

O REN, utilizado neste trabalho na tarefa de aprendizado supervisionado de máquina, foi apoiado pelo Stanford NER (FINKEL; GRENAGER; MANNING, 2005). A escolha pelo Stanford NER deve-se ao fato da facilidade de adaptações e modelos pré-existentes que a ferramenta dispõe e a implementação nativa do algoritmo CRF, descrito

²² O *parser* é um analisador sintático. Sua função é ler uma entrada de dados que possuem regras específicas - em geral é um texto reconhecível por humanos - e montar uma estrutura de como é sua composição. Pode ser descrito como uma “decodificação”, “interpretação” ou, em alguns casos, “conversão”.

de forma detalhada na subseção 4.3.3.2, que teve sua escolha atrelada ao fato do tipo de aprendizado de máquina utilizado, possibilitando a extração automática de entidades a partir de um conjunto de dados com uma capacidade de resposta mais rápida do que outras técnicas já utilizadas, como a implantação de heurísticas específicas (MOTA; SANTOS, 2008).

A aplicação do CRF para as tarefas de REN apresenta bons resultados, principalmente o utilizando aplicado ao processo de aprendizado supervisionado de máquina, pois o algoritmo apresenta uma capacidade de resposta mais rápida e adequada se comparado a outros algoritmos (AMARAL; VIEIRA, 2014), além de prover um maior grau de sucesso nas tarefas de REN (FINKEL; GREAGER; MANNING, 2005).

5.1.1.2. Aprendizado de Máquina

Para o aprendizado de máquina é necessário um modelo de treino – corpus – composto da anotação manual das entidades existentes em artigos e relatórios de prospecção tecnológica, possibilitando o aprendizado geral da ferramenta de aprendizado supervisionado. A tarefa de anotação manual das prospecções tecnológicas é importante para o sucesso da tarefa de REN, pois através dessa atividade a máquina aprende padrões em linguagem natural necessário para a classificação dos novos elementos contidos nos textos.

Existem três tipos de formas para taguear as EN no texto, conforme explicado na subseção 4.3.2, IO, BIO e BILOU. Para o presente trabalho, utilizou-se a notação BILOU por dois motivos: (i) Testes aplicados sob a Coleção Dourada (CD) do Segundo HAREM, empregando ambas as notações, demonstraram que a notação BILOU se equivale a BIO, conforme os resultados analisados. Isso porque o BILOU facilita o processo de classificação feito pelo sistema desenvolvido por possuir mais duas identificações: L(*Last*) e U(*Unit*); e (ii) Segundo o estudo de RATINOV e ROTH (2009), testes foram realizados com as duas notações, concluindo também com os seus resultados obtidos que, apesar do formalismo BIO ser amplamente adotado, o BILOU o supera significativamente.

Como exemplo, a prospecção: *“For the year of 2030, China will dominate the industries of AI.”*, após o tagueamento usando a notação BILOU, seguindo as diretrizes para REN proposta neste trabalho, teria o seguinte resultado conforme a Tabela 10. O objetivo do tagueamento neste trabalho é identificar os trechos que antecedem um período

– que pode ser um ano, uma década ou expressões do tipo: “...in the next 5 years...” – e, com isso, poder identificar possíveis prospecções na frase.

Tabela 10. Exemplo tagueamento usando a notação BILOU.

Termo	Tag	Termo	Tag
<i>For</i>	B_TEMPPRED	<i>will</i>	O
<i>the</i>	I_TEMPPRED	<i>dominate</i>	O
<i>year</i>	I_TEMPPRED	<i>the</i>	O
<i>of</i>	L_TEMPPRED	<i>industries</i>	O
<i>2030,</i>	U_DATE	<i>Of</i>	O
<i>China</i>	O	<i>AI.</i>	O

5.1.1.3. Criação de Etapas de Treinamento/Validação e Testes

Para o aprendizado de máquina é necessário um conjunto de procedimentos que permita que a máquina de aprendizado desenvolva um modelo de aprendizado supervisionado com um valor significativo nas medidas de precisão, abrangência e medida-F.

A base de dados *Futures*²³ (Elsevier) foi utilizada como insumo nesse processo, em que foram coletados 1092 artigos entre os anos de 2010 e 2020 para serem divididos em conjuntos ou *folds*, e são separados em dois fluxos denominados: Treinamento/Validação e Testes, para a realização do *K-fold Validation*.

5.1.1.3.1. Treinamento/Validação

O processo de treinamento e validação se baseia em treinar a máquina para o Reconhecimento de Entidades Nomeadas e a identificação dos parâmetros que proporcionem as melhores métricas.

Nessa etapa, os *folds* correspondentes a treinamento/validação são submetidos ao Stanford NER com uma gama de parâmetros utilizados pelo algoritmo CRF. A cada mudança do valor dos parâmetros, o algoritmo calcula a média das medidas de Precisão, Abrangência e Medida-F para determinar qual o melhor parâmetro para a máquina de aprendizado.

²³ *Futures* é um *journal* internacional e multidisciplinar voltado para os estudos do futuro, estando preocupado com o futuro de culturas e sociedade a médio e longo prazo. É publicado pela Elsevier e engloba temas como ciência e tecnologia, economia e política, meio ambiente e humanidade. Disponível em: <https://www.journals.elsevier.com/futures>.

Neste trabalho foram testados os parâmetros expostos na Tabela 11, selecionados com base nas particularidades do CRF, de acordo com a literatura, de modo a identificar quais os parâmetros que entregam os melhores resultados.

Tabela 11. Parâmetros Utilizados.

Parâmetros	Descrição	Intervalo de Valores Testados
<i>maxNGramLen</i>	Define a quantidade máxima de n-gramas que será usada para determinar a entropia.	1 a 10
<i>useClassFeature</i>	Insere uma prévia sobre as classes que equivale a quantas vezes o recurso apareceu nos dados de treinamento.	Verdadeiro ou Falso
<i>useNGrams</i>	Análise baseado nas características das letras dos termos, ou seja, <i>substrings</i> do termo.	Verdadeiro ou Falso
<i>usePrev</i>	Analisa a entropia utilizando o termo anterior, semelhante ao uso da notação BILOU.	Verdadeiro ou Falso
<i>useNext</i>	Analisa a entropia utilizando o termo posterior, semelhante ao uso da notação BILOU.	Verdadeiro ou Falso
<i>useSequences</i>	Utilizar o recurso de combinação entre os grupos de entidade para analisar o termo.	Verdadeiro ou Falso

A Figura 37 apresenta o arquivo de parâmetros aplicado ao Stanford NER que foi utilizado no trabalho por apresentar os melhores resultados.

```

map = word=0,answer=1
useClassFeature=true
useWord=true
useNGrams=true
noMidGrams=true
useDisjunctive=true
maxNGramLeng=3
usePrev=true
useNext=true
useSequences=true
usePrevSequences=true
maxLeft=1
useTypeSeqs=true
useTypeSeqs2=true
useTypeySequences=true
wordShape=chris2useLCS

```

Figura 37. Melhor resultado de parâmetros.

5.1.1.3.2. Testes

A etapa de testes consiste em avaliar a capacidade da máquina de aprendizado na tarefa de classificação das EN. A Figura 38 ilustra esse processo.

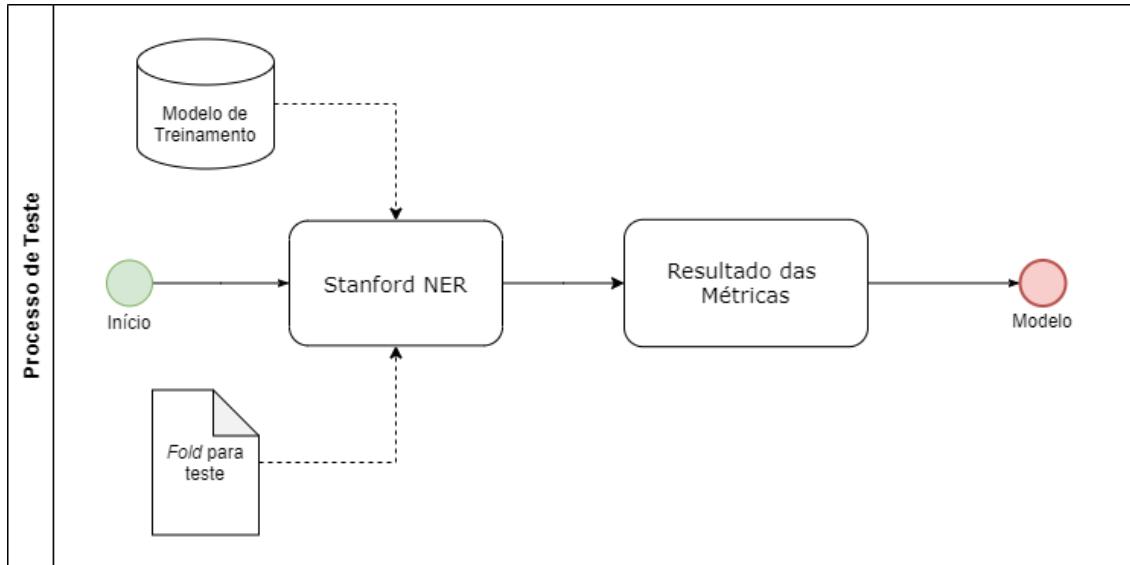


Figura 38. Processo de Testes.

Os *folds* do processo de treinamento/validação em conjunto com o *fold* de teste então, são submetidos ao processo de *K-fold Validation* utilizado no Stanford NER. Esse processo permite que a máquina avalie seu aprendizado médio para a classificação de elementos baseado nas métricas de precisão, abrangência e medida-F, além de prover a comparação dos valores médios obtidos no processo e validar o modelo de aprendizado para o módulo de Processamento de Documentos do NERMAP.

5.2. Método: Processo de *Roadmapping* Tecnológico

Este tópico visa apresentar o método que complementa o modelo apresentado no tópico anterior, de acordo com o *framework Design Science*. Desse modo, o Processo NERMAP, produto desta seção, apoia e estende o Modelo NERMAP ao descrever as etapas necessárias à realização do processo de *Roadmapping* Tecnológico.

O Processo NERMAP é capaz de representar um processo de *Roadmapping* Tecnológico de um FTA de forma simples e iterativa, semiautomatizando o processo manual e permitindo a geração de *roadmaps* de diversos documentos simultaneamente.

O processo se dá início quando os usuários, de acordo com seu papel organizacional e responsabilidades definidas no âmbito do gerenciamento de usuários, determinam os objetivos do processo de *Roadmapping* Tecnológico, alinhados com os próprios objetivos e necessidades que deram origem ao estudo de FTA correspondente.

A próxima etapa é a definição dos parâmetros do TRM (“Nome”, “Área” e “Ano Limite”) em conformidade com os objetivos anteriormente definidos. Ao criar o TRM, o usuário está apto a anexar documentos para gerar *roadmaps*, dessa forma, o usuário

submete arquivos (em formato TXT ou PDF) junto com informações sobre o mesmo (“Título da publicação”, “Ano da publicação”, “Autores” e “Confiabilidade da fonte”).

O módulo executa, de maneira automática e em segundo plano, o processamento para converter os documentos para texto plano, taguear (fazendo uso de REN) e fatiar em prospecções datadas. O produto, o *roadmap*, em formato de *timeline* é disponibilizado ao usuário para refinamento manual (adição, edição e exclusão de prospecções). O usuário pode redefinir os parâmetros previamente estabelecidos, tanto para o TRM quanto para cada documento, e obter variações tanto no *roadmap* gerado para cada arquivo, quanto no *roadmap* completo do TRM (composto dos *roadmaps* de todos os documentos do TRM). O *roadmap* pode ser exportado em formatos como PDF, DOC, CSV, JSON e TXT ou ser usado como *Application Programming Interface* (API) por outra plataforma.

O Processo NERMAP no formato *Business Process Model and Notation* (BPMN) é apresentado na Figura 39 (OBJECT MANAGEMENT GROUP, 2013).

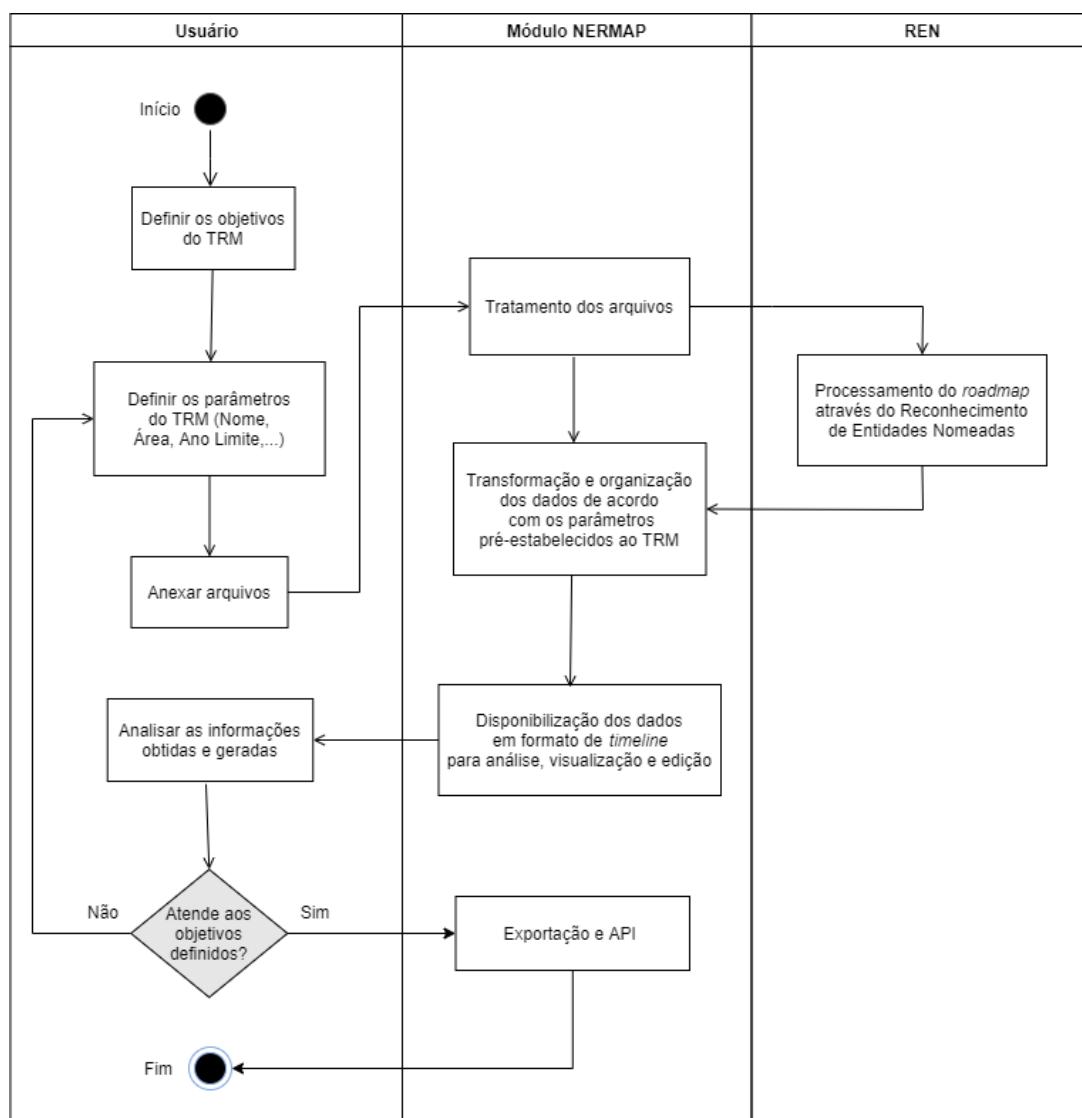


Figura 39. Processo NERMAP.

5.3. Instanciação: O sistema NERMAP

A partir da definição do Modelo NERMAP e do Processo NERMAP, elementos do *Design Science*, segue-se, então, a instanciação desses elementos em um sistema computacional, denominado NERMAP. O sistema tem a capacidade de executar um processo de *Roadmapping* Tecnológico de forma semiautomatizada (ou seja, com interação do usuário em etapas como: preparação, análise e refinamento).

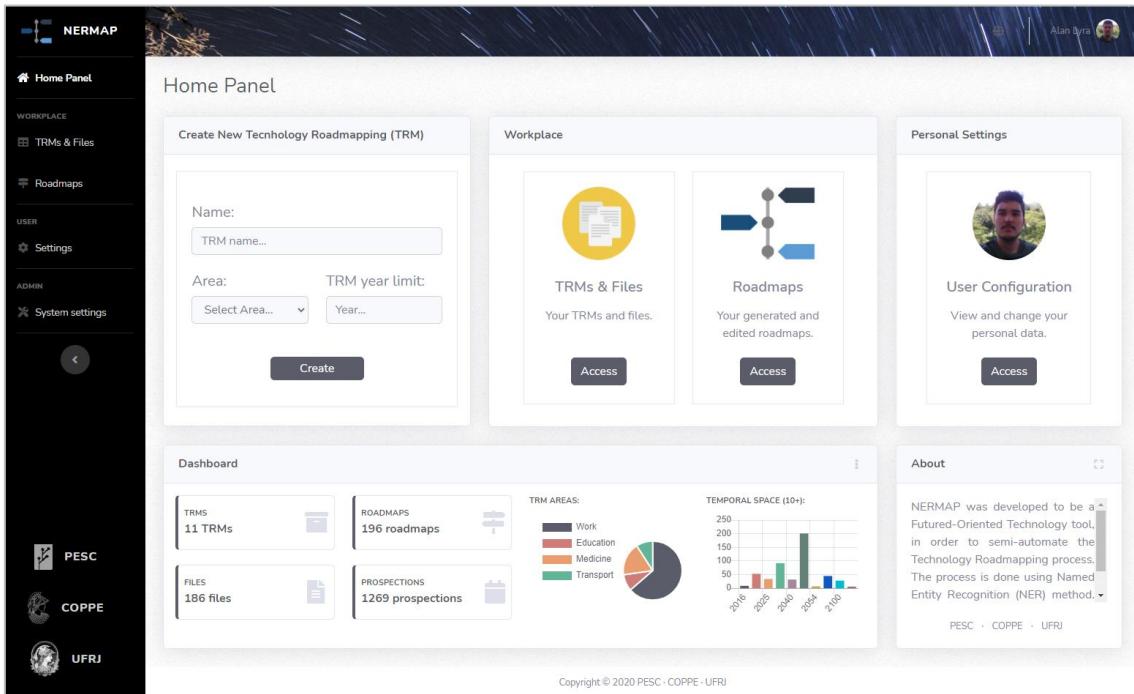


Figura 40. Painel de Início do NERMAP.

O Painel de Início do NERMAP, tela exibida após o login do usuário, dispõe dos atalhos para as funcionalidades do sistema, separando em criação de um novo TRM, *Workplace* (Gerenciamento dos arquivos e Gerenciamento dos *Roadmaps*) e configuração do usuário. Além disso, também apresenta um *dashboard*²⁴, contendo o uso do sistema mostrando a quantidade de TRMs, arquivos, *roadmaps* e prospecções junto aos gráficos de áreas e datas mais presentes no estudo.

5.3.1. Arquitetura do NERMAP

Com o objetivo de apoiar o FTA distribuído e integrado, com o foco em um de seus métodos, o *Roadmapping* Tecnológico, o NERMAP tem sua arquitetura dividida em quatro camadas: apresentação, gerenciamento, processamento e armazenamento, como

²⁴ *Dashboard* é um painel visual que contém informações, métricas e indicadores.

ilustrado na Figura 41. A camada de Apresentação é constituída pela interface do usuário, meio pelo qual os usuários interagem com o sistema. A camada de Gerenciamento é responsável por toda a gerência do sistema: nela estão implementados os módulos de Gerenciamento de Usuário (encarregado do controle de acesso, ações administrativas e o gerenciamento de pessoas e grupos), Gerenciamento do TRM (responsável por permitir a gerência e o compartilhamento dos TRMs, bem como a inclusão/edição de parâmetros) e Gerenciamento de Documentos (responsável pela gerência dos arquivos submetidos ao sistema).

A terceira camada é a de Processamento, responsável por processar os documentos inseridos no sistema e aplicar o REN para a geração das prospecções datadas visando o produto, o *roadmap*. A quarta camada é a de Armazenamento, sendo o banco de dados do NERMAP o local de armazenamento padrão dos dados do sistema.

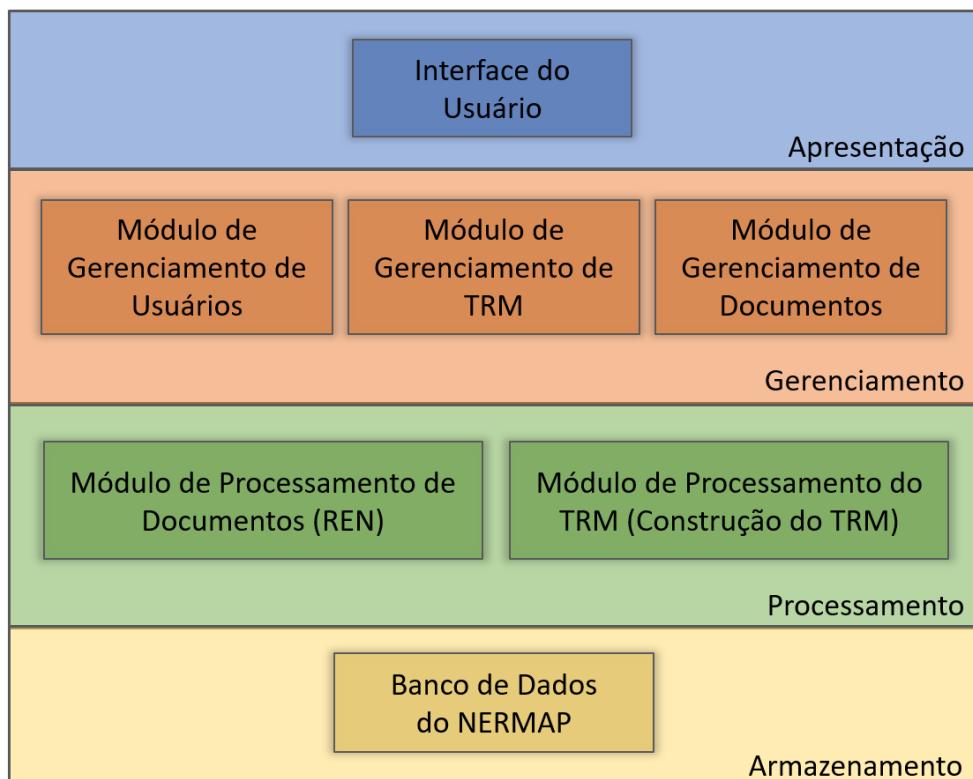


Figura 41. Camadas da arquitetura do NERMAP.

Os artefatos descritos nas camadas de apresentação, gerenciamento, processamento e armazenamento do NERMAP são descritos a seguir e estão organizados na arquitetura apresentada na Figura 42.

- **Interface do Usuário:** Visualização das instanciações de cada módulo;
- **Gerenciamento de Usuários:** Cadastro e edição de cada usuário do sistema;

- **Gerenciamento de TRM:** Criação do TRM, atribuindo parâmetros e papéis (*roles*) para cada usuário participante, podendo orientar a grupos;
- **Gerenciamento de Documentos:** Submissão de documentos para um TRM para dar início ao processo de *Roadmapping* Tecnológico;
- **Processamento de Documentos:** Conversão do documento para texto plano e, em seguida, tagueamento do texto com base no REN;
- **Processamento do TRM:** Obtenção das prospecções datadas para a construção do *roadmap*.
- **Banco de Dados do NERMAP:** Armazenamento de todas as informações inerentes ao sistema.

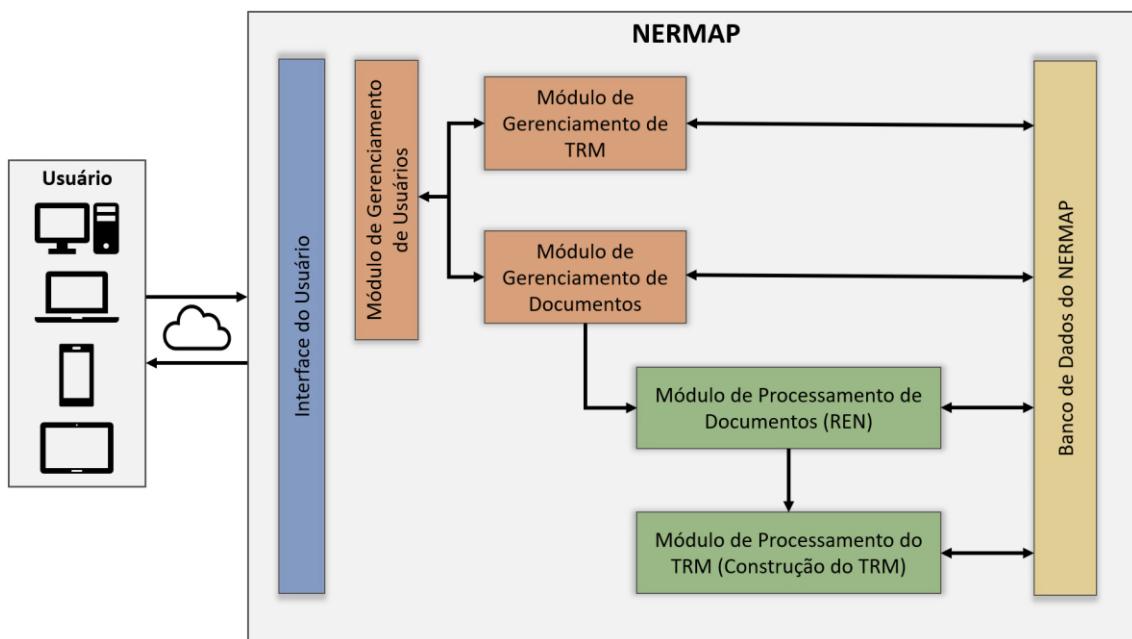


Figura 42. A Arquitetura do NERMAP.

Observando a arquitetura, podemos ver que no acesso do usuário ao sistema, as informações são trocadas através da interface do usuário, que está estreitamente ligada ao controle de acesso do sistema, parte do módulo de Gerenciamento de Usuários, através do *login*. No Gerenciamento de Usuários, um usuário pode se cadastrar no sistema e editar seu perfil, conforme apresentado na Figura 43 e na Figura 44, respectivamente.

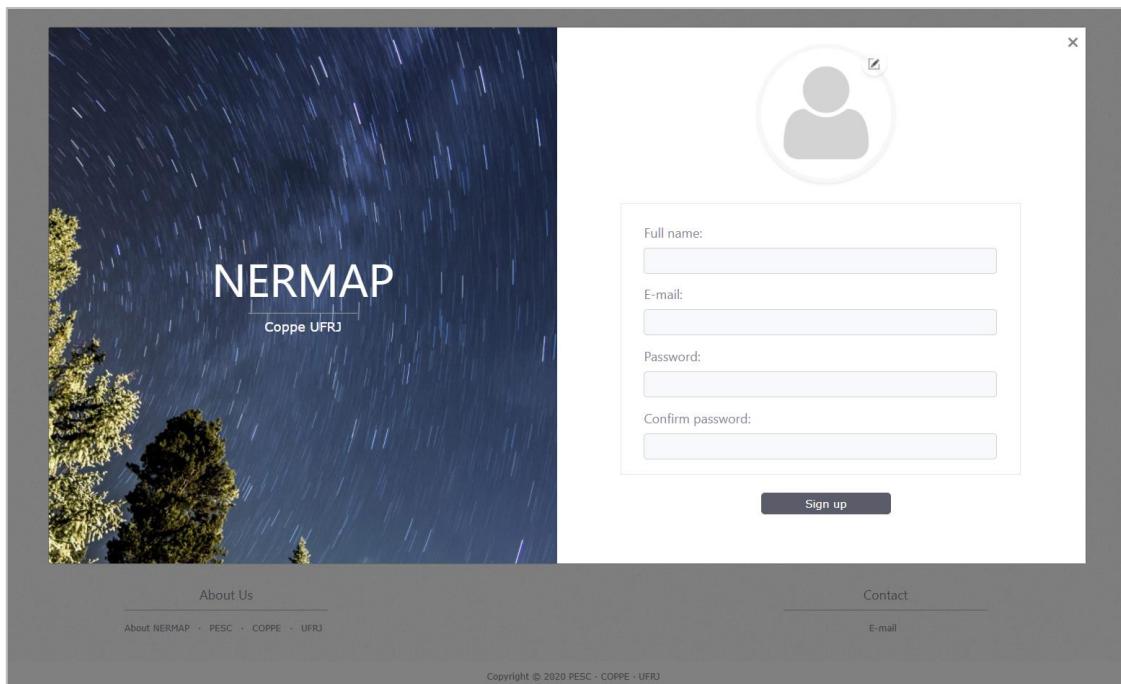


Figura 43. Cadastro de usuário do NERMAP.

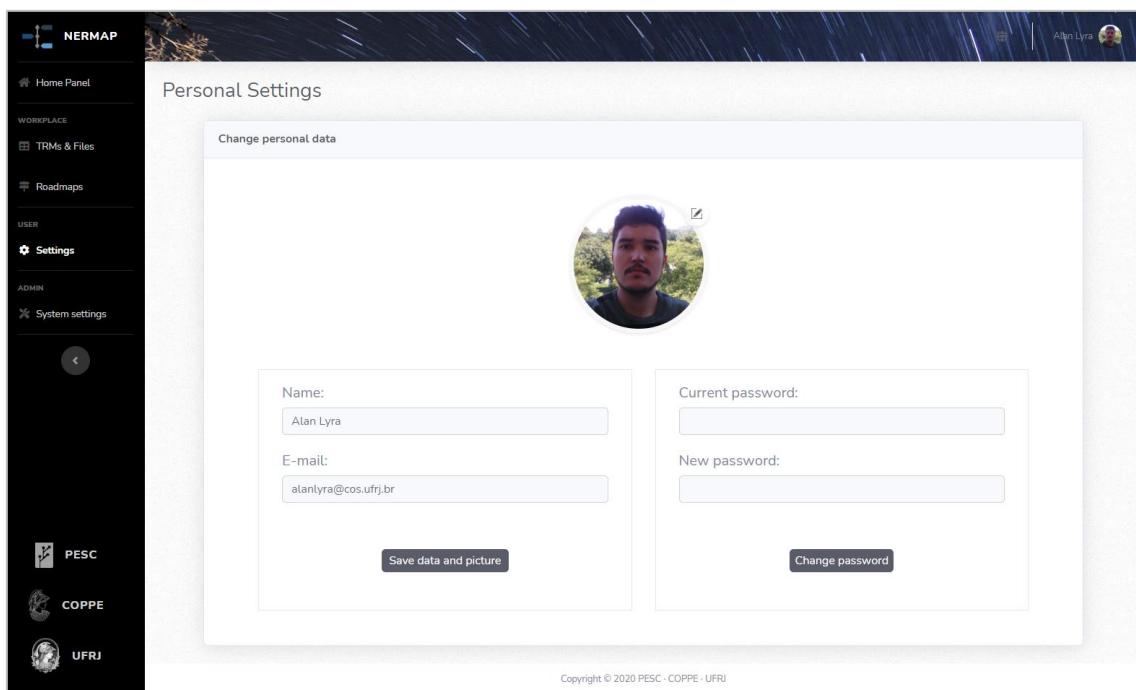


Figura 44. Edição de perfil do usuário do NERMAP.

Os TRMs criados, por sua vez, podem ser gerenciados pelo módulo de Gerenciamento de TRM, como ilustrado na Figura 45. É possível visualizar todas as informações como: nome, área, ano limite, status de processamento, número de documentos anexados, bem como todos os documentos listados, *roadmaps* processados e realizar ações de, quando for um administrador: compartilhamento, edição e exclusão do TRM, ou, quando for um participante: sair do TRM.

Figura 45. Gerenciamento de TRMs do NERMAP.

O compartilhamento configura a orientação a grupos implementada no sistema, fazendo com que um *roadmap* possa ser acessado e ter a colaboração de todos os usuários do NERMAP. O administrador do TRM gerencia o compartilhamento, podendo adicionar ou remover participantes, como ilustrado na Figura 46.

Figura 46. Compartilhamento de TRMs do NERMAP.

Além disso, os usuários podem submeter documentos, tanto em TXT quanto em PDF, para alimentar o TRM por meio do módulo de Gerenciamento de Documentos,

como representado na Figura 47. O usuário pode inserir informações como Título do arquivo, Data da publicação, Autores e a Confiabilidade da fonte.

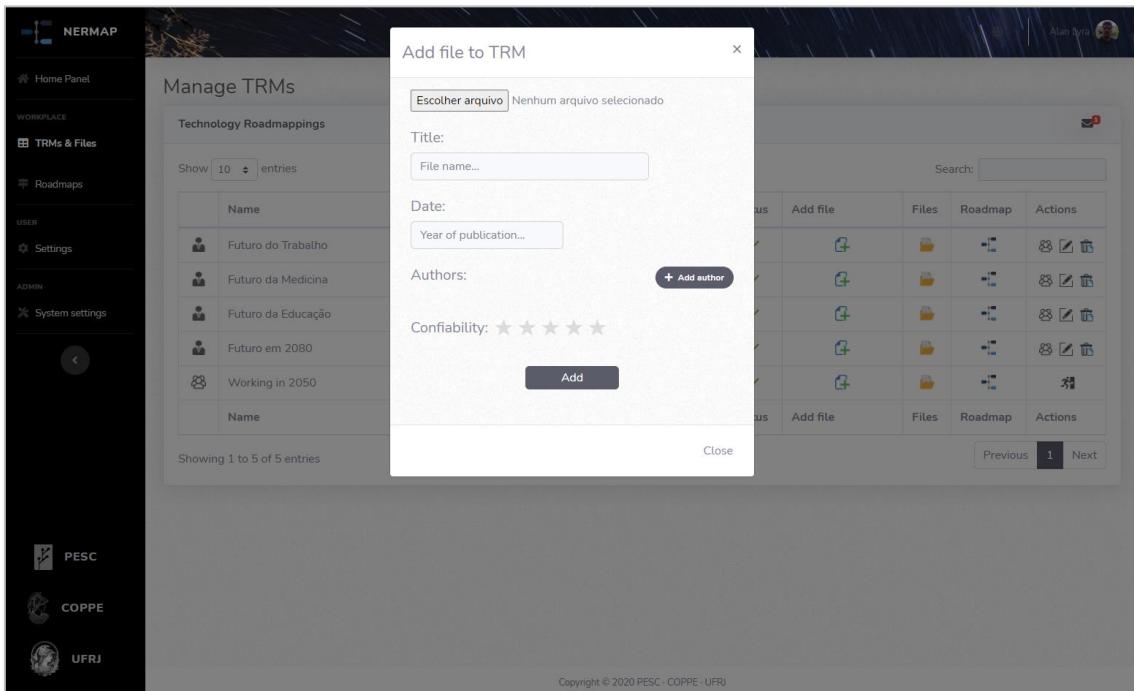


Figura 47. Gerenciamento de Documentos do NERMAP.

Os documentos são processados para gerar *roadmaps* e podem ser acessados individualmente para ações como edição e remoção. Os arquivos são listados na tela junto com as informações relevantes, como mostrado na Figura 48.

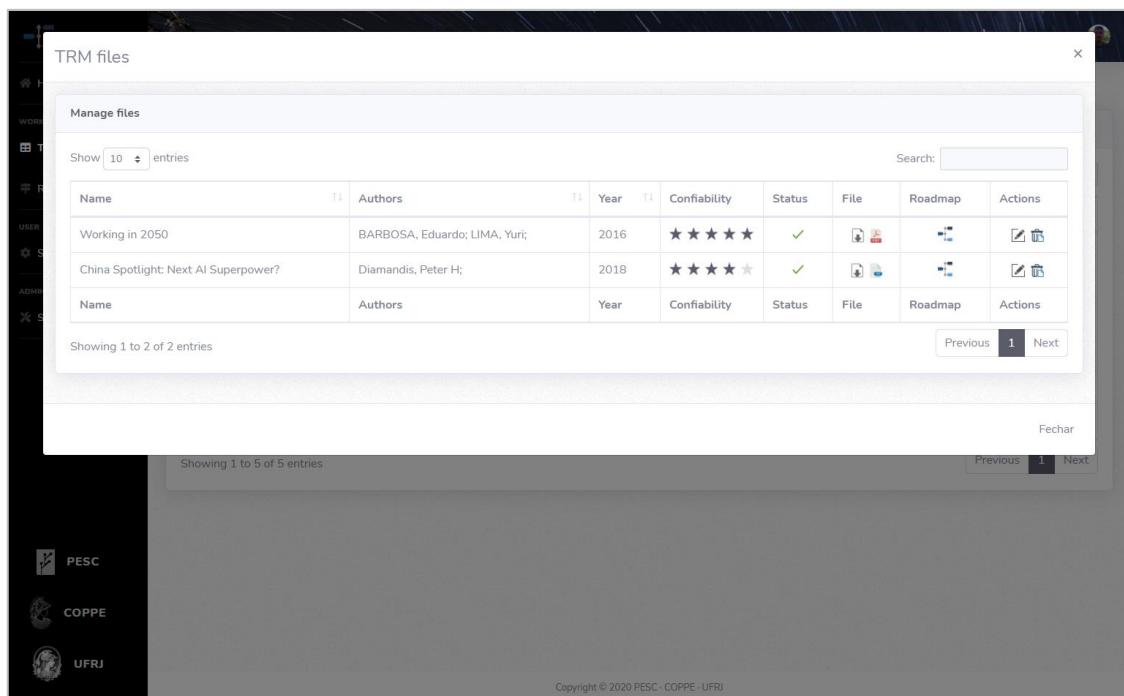


Figura 48. Arquivos do TRM.

Os *roadmaps* por sua vez podem ser acessados de forma individual – para cada arquivo – ou de forma completa (composto de todos os *roadmaps* do TRM) como explicado na figura 36. Os *roadmaps* gerados são exibidos como forma de *timeline*, como mostrado na Figura 49.

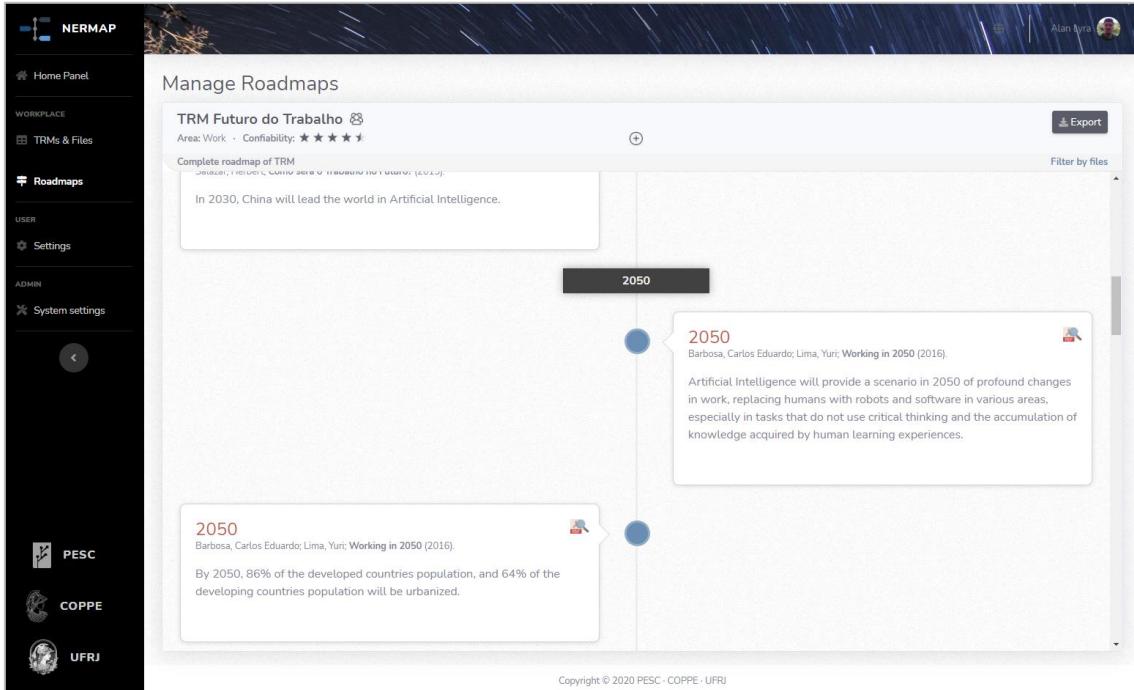


Figura 49. *Roadmap* sendo exibido no NERMAP.

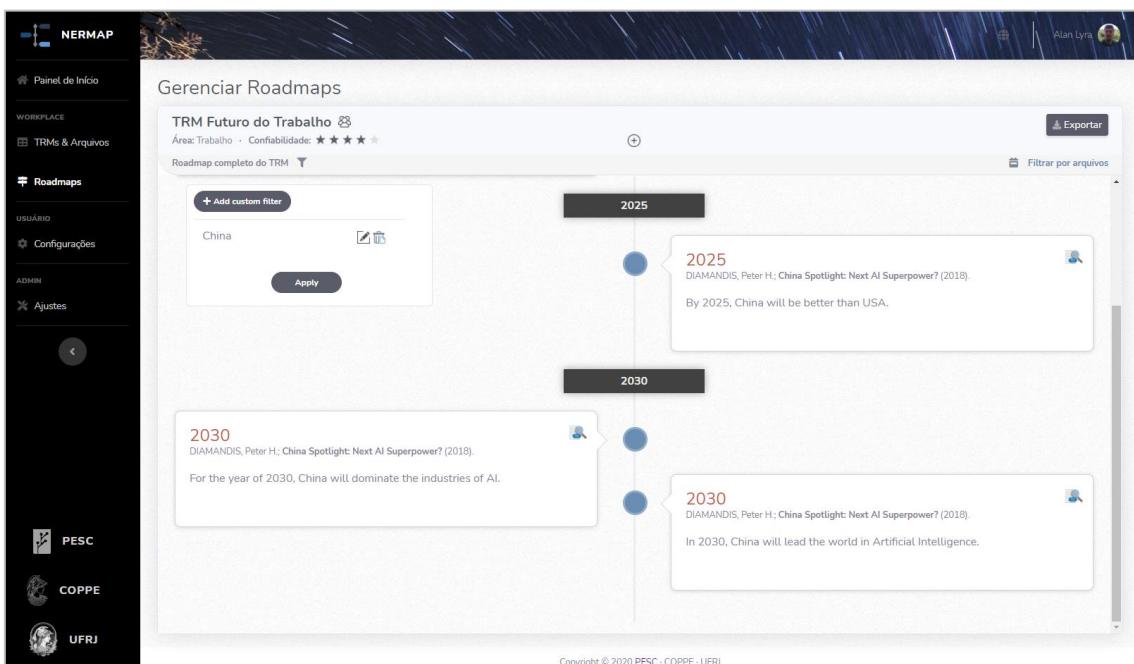


Figura 50. Filtro sendo aplicado ao *roadmap*.

Por se tratar de uma análise de documentos para uma visão “macro” da área escolhida para o TRM, o NERMAP dispõe de um filtro para direcionar o *roadmap* gerado

para subtemas da área selecionada. Na Figura 50, temos como exemplo um TRM para a área do Trabalho, onde documentos que tratam de prospecção para essa área foram submetidos ao NERMAP. Neste TRM, foi aplicado um filtro, que pode conter uma ou mais palavras/expressões que farão com que o NERMAP retorne apenas prospecções que contenham esses termos chaves.

Quando o usuário passa o mouse por cima de uma prospecção, é exibida a confiabilidade da fonte e, também, o botão de edição, conforme apresentado na Figura 51. A confiabilidade média do TRM – com base nos arquivos que o compõem – também é calculada e é exibida no topo da página.

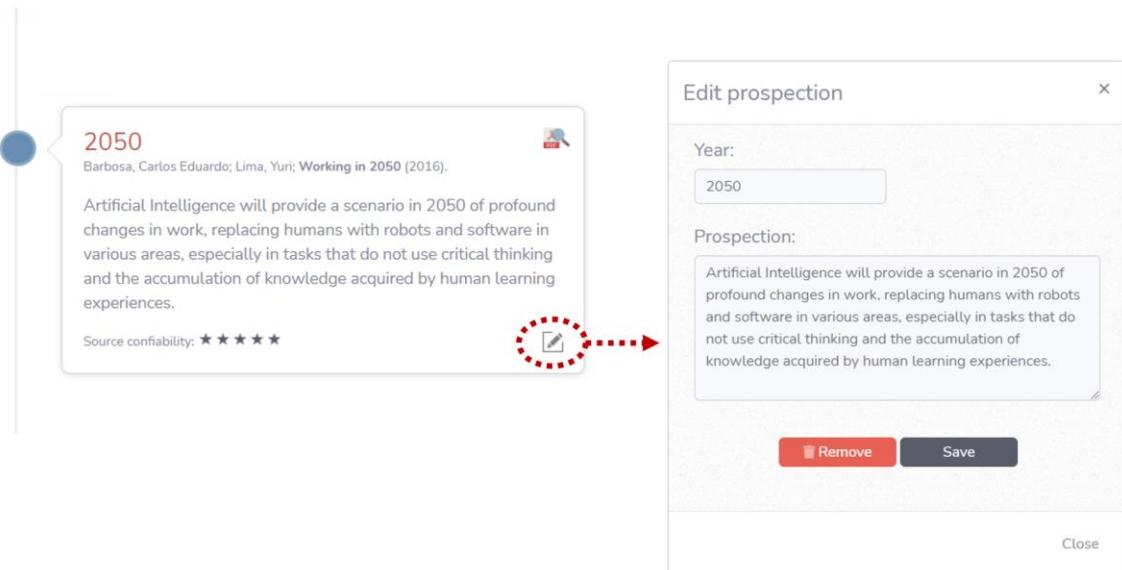


Figura 51. Edição da prospecção.

Na tela de exibição do *roadmap*, o usuário, além de editar cada prospecção, pode realizar ações como visualizar os participantes do TRM, alternar da visualização completa do *roadmap* do TRM para a visualização do *roadmap* de cada arquivo e visualizar a prospecção diretamente no arquivo – na forma de *highlight* no texto, tanto para um arquivo TXT quanto um arquivo PDF, como ilustrado na Figura 52.

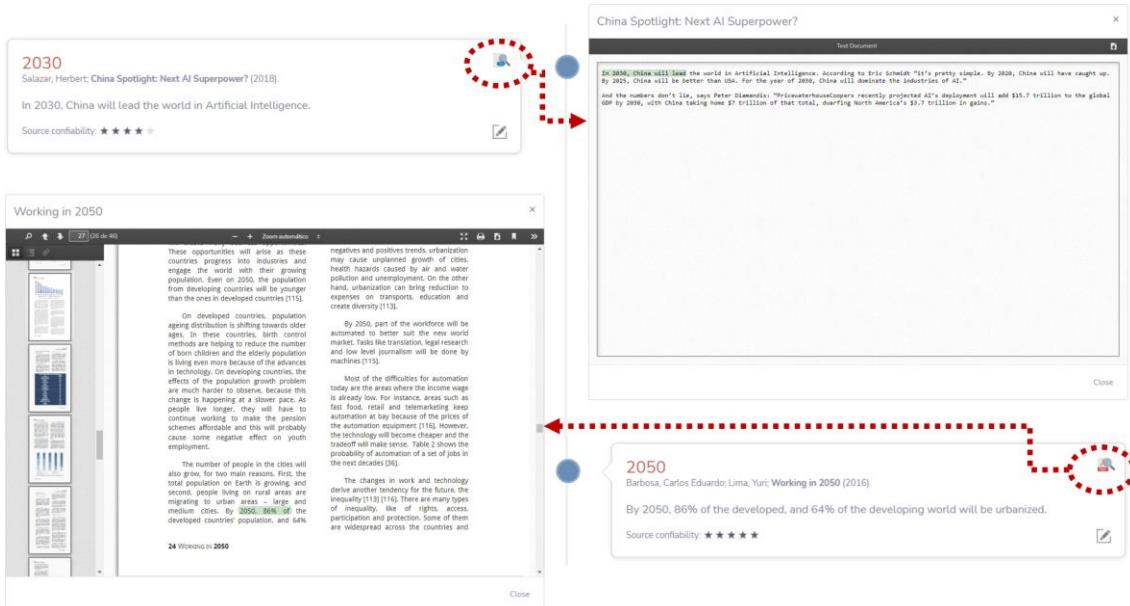


Figura 52. Visualização das prospecções nos documentos no NERMAP.

Prospecções podem ser adicionadas de forma manual à um *roadmap*, onde o usuário também especificará a fonte, autores e a confiabilidade para manter as métricas e a qualidade do TRM, como exposto na Figura 53. Esse processo é importante pois, durante um estudo de *Roadmapping* Tecnológico, prospecções podem ser encontradas em leituras em que não é necessário submeter todo o texto para o NERMAP, fazendo uso apenas do trecho relevante ao estudo.

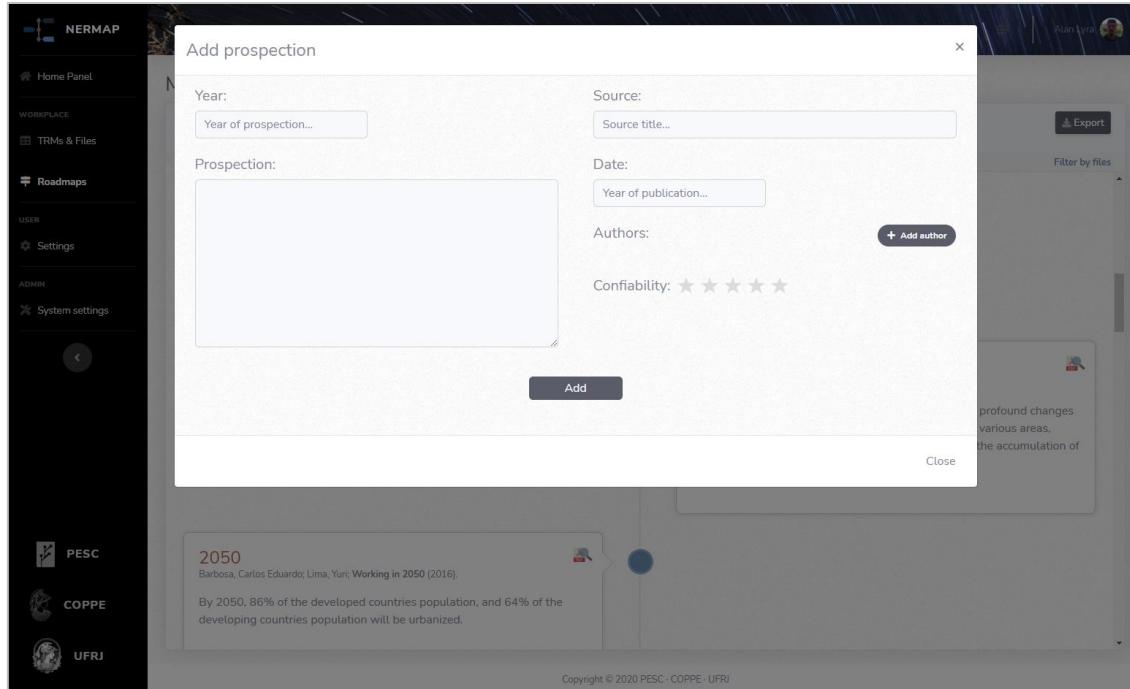


Figura 53. Adicionando prospecção manualmente ao *roadmap*.

A exportação do produto do NERMAP, o *roadmap*, pode ser feita em 5 formatos: CSV, PDF, TXT, DOC e JSON, ou também pode ser disponibilizado na forma de API, para ser consultado/utilizado em qualquer outro projeto de Prospecção Tecnológica, como pode ser observado na Figura 54.

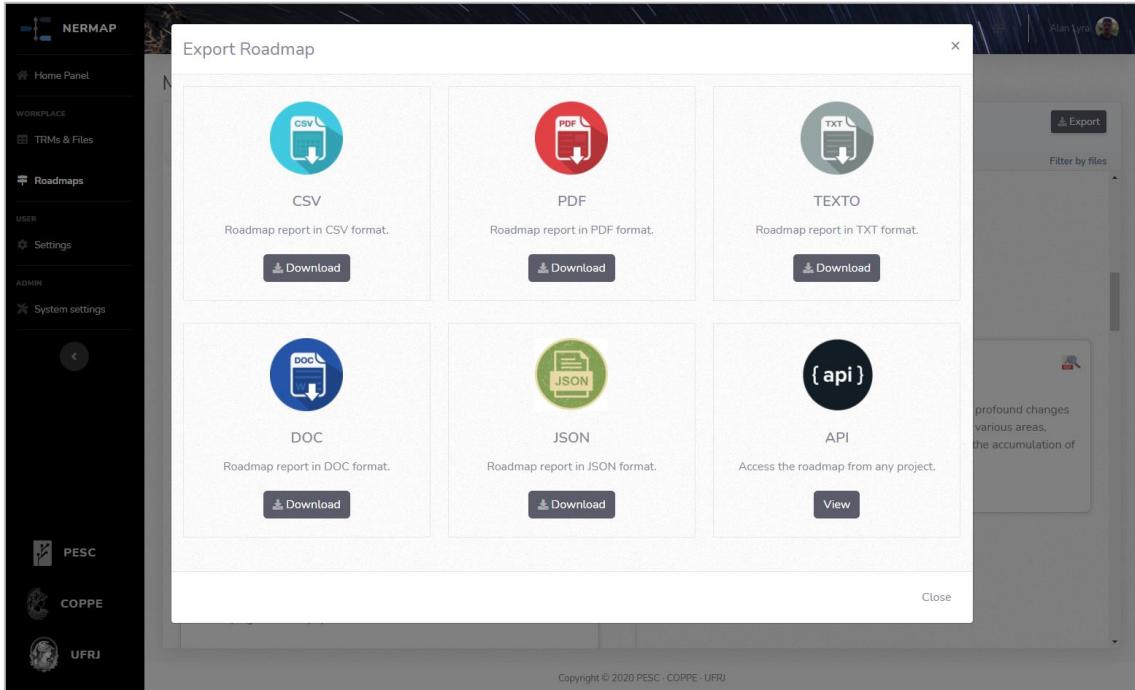


Figura 54. Modos de exportação do *roadmap*.

5.3.2. Modelagem do NERMAP

Apoiado nos conceitos apresentados na arquitetura do NERMAP, definimos modelos que representam diretrizes para a sua construção. Os modelos foram elaborados utilizando diagramas Linguagem de Modelagem Unificada (*Unified Modeling Language* – UML) (BOOCH; RUMBAUGH; JACOBSON, 1999), o modelo de abstração mais comum para o projeto de *software*. A UML se trata de uma notação simples e padronizada para descrever modelos orientados a objetos, tendo um extenso apoio na literatura, além de diversas ferramentas de *software*.

Embora existam mais de uma dúzia de diagramas, nos concentramos em três dos principais diagramas da UML: diagrama de classes, diagrama de caso de uso e diagrama de atividades.

5.3.2.1. Diagrama de Casos de Uso do NERMAP

Conforme BOOCH et al. (1999), os diagramas de Casos de Uso são importantes para modelar o comportamento de um sistema, um subsistema ou uma classe. O diagrama apresenta um conjunto de casos de uso, além de seus atores e relacionamentos. Um caso

de uso envolve modelar o contexto de um sistema. Diagramas de caso de uso são essenciais para visualizar, especificar e documentar o comportamento de um elemento. O diagrama de caso de uso do NERMAP é apresentado na Figura 55.

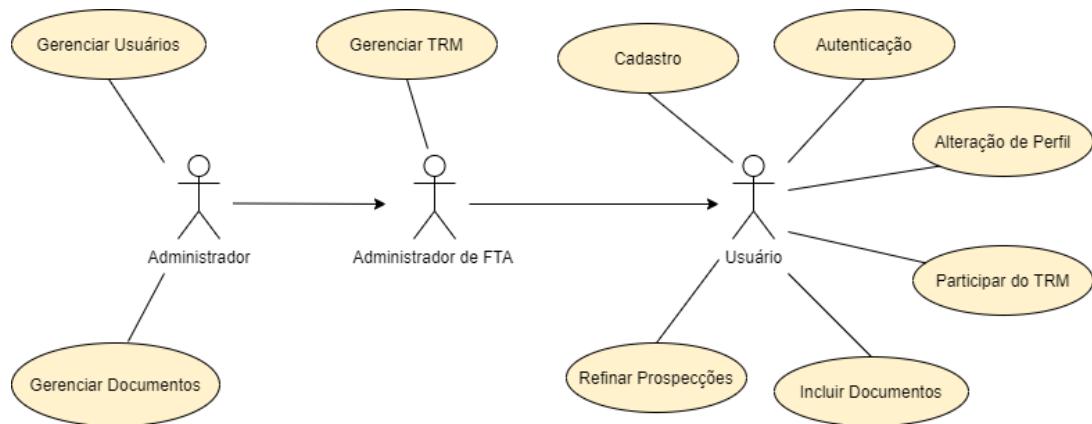


Figura 55. Diagrama de Casos de Uso do NERMAP.

A lista dos Casos de Uso – e suas respectivas descrições – é apresentada na Tabela 12.

Tabela 12. Descrição dos Casos de Uso do NERMAP.

Ator	Caso de Uso	Descrição
Usuário	Cadastro	Cadastro do próprio usuário no sistema.
	Autenticação	Verificação do usuário e senha.
	Alteração de Perfil	Modificação do cadastro do usuário autenticado.
	Participar do TRM	Realização as atividades dentro do TRM ao qual o usuário foi associado.
	Incluir Documentos	Inclusão dos documentos de subsídio no sistema, para geração do <i>roadmap</i> .
	Refinar Prospecções	Edição/remoção de prospecções geradas automaticamente e inclusão de novas prospecções ao <i>roadmap</i> .
Administrador de FTA	Gerenciar TRM	Permite a administração de um TRM e de seus participantes, além de gerenciar a sua execução.
Administrador	Gerenciar Usuários	Permite a edição administrativa do cadastro de usuários no sistema.
	Gerenciar Documentos	Permite a edição administrativa do cadastro de Documentos do TRM no sistema.

5.3.2.2. Diagrama de Classes do NERMAP

Um diagrama é o núcleo da arquitetura e é uma ferramenta importante para documentar um sistema de classes, apresentando os conjuntos de classes, interfaces e colaborações, elementos básicos do sistema, assim como seus relacionamentos e sendo utilizado para fornecer uma descrição estática do sistema que está sendo projetado. É um dos elementos mais importantes da modelagem de sistemas, servindo de base para a concepção de outros diagramas. O diagrama de classes do NERMAP é apresentado na Figura 56.

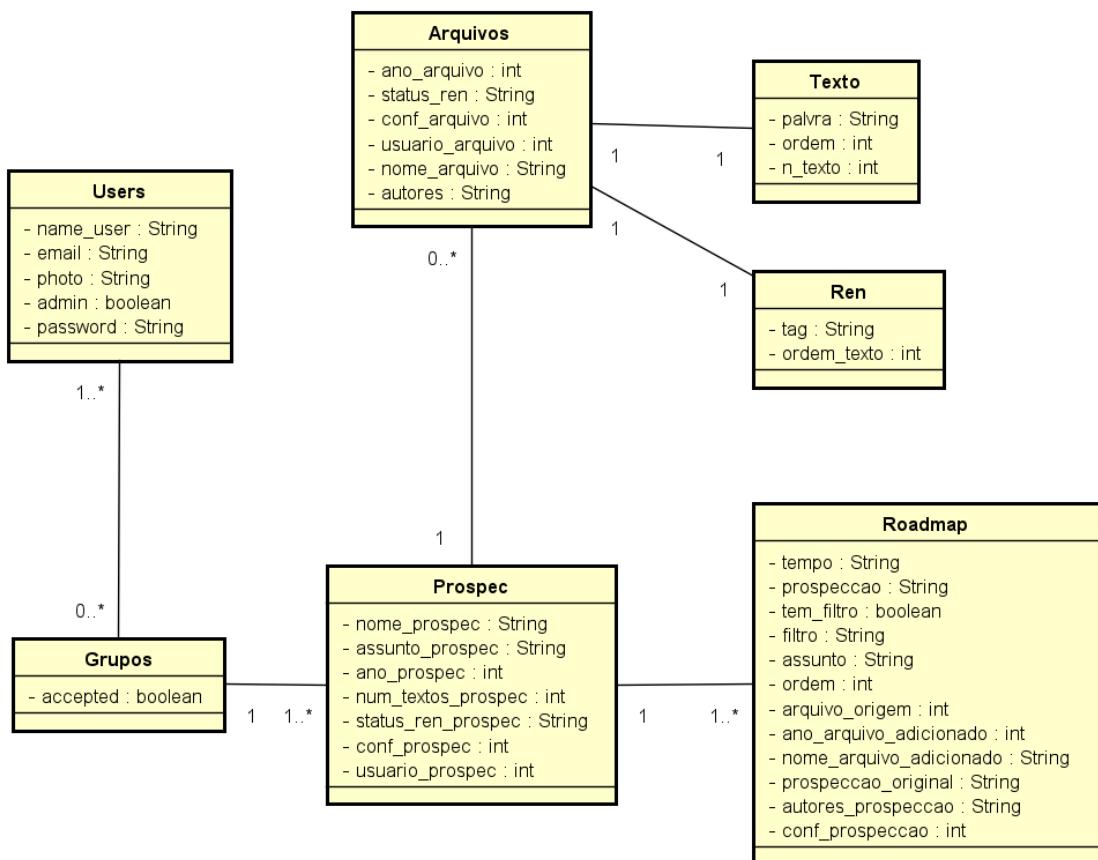


Figura 56. Diagrama de Classes do NERMAP.

As descrições das classes e suas responsabilidades estão descritas na Tabela 13.

Tabela 13. Descrição das Classes do NERMAP.

Classe	Responsabilidade
Users	Registrar as informações do usuário.
Grupos	Registrar os grupos do sistema, relacionando os usuários e os convites.
Prospec	Registrar os TRMs criados no sistema.
Arquivos	Subsídios para os TRMs.
Texto	Prospecções em texto plano extraídas dos documentos.
Ren	Relacionamento com as <i>tags</i> do REN para cada prospecção.
Roadmap	Registrar as prospecções datadas para formação do <i>roadmap</i> .

5.3.2.3. Diagrama de Atividades do NERMAP

Os diagramas de atividades são um dos diagramas na UML utilizados para modelar os aspectos dinâmicos dos sistemas. Um diagrama de atividades é, essencialmente, um fluxograma que evidencia a passagem de controle de uma atividade para outra (BOOCH; RUMBAUGH; JACOBSON, 1999). Um diagrama de atividades envolve modelar a sequência e o paralelismo dos passos de um processo computacional. O diagrama de atividades do NERMAP é apresentado na Figura 57.

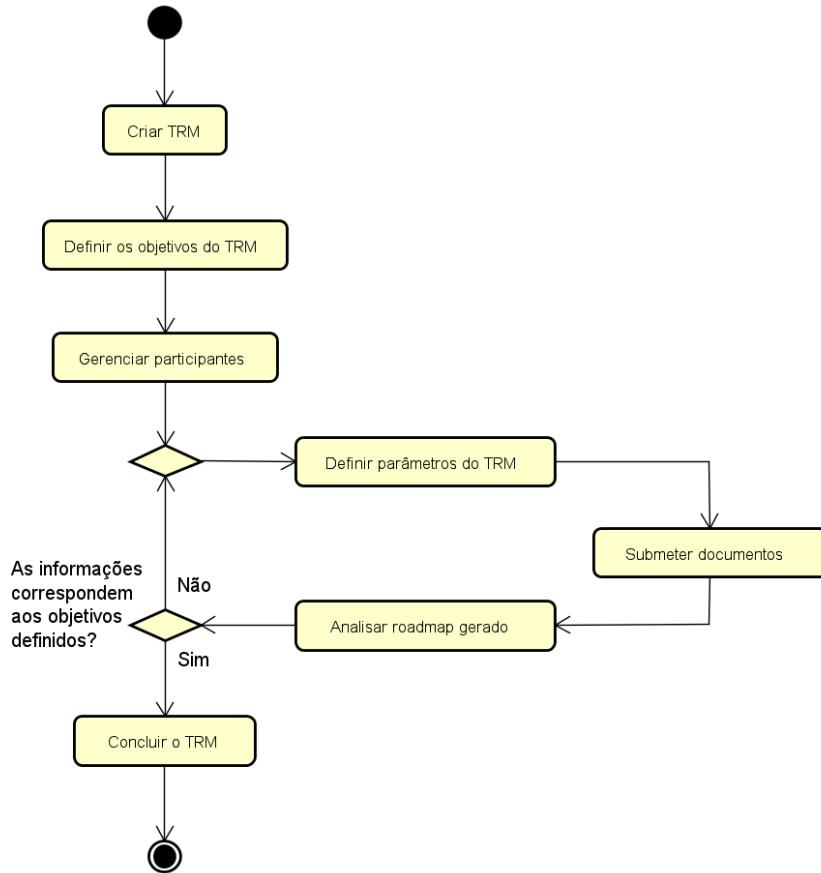


Figura 57. Diagrama de Atividades do NERMAP.

As descrições das atividades e suas descrições estão listadas na Tabela 14.

Tabela 14. Descrição das Atividades do NERMAP.

Atividade	Descrição
Criar TRM	Cadastrar um TRM no NERMAP.
Definir os objetivos do TRM	Especificiar os objetivos para qual a finalidade o TRM se propõe e para qual área/segmento será aplicado.
Gerenciar participantes	Gerenciar convites para a compor os participantes do TRM.
Definir parâmetros do TRM	Especificiar o ano limite e os arquivos que irão compor o TRM, analisando confiabilidade, data de publicações, área e afins.
Submeter documentos	Incluir subsídios para geração do <i>roadmap</i> .
Analizar roadmap gerado	Refinamento do <i>roadmap</i> , editando, removendo ou adicionando novas prospecções.
Concluir o TRM	Finalizar o estudo, podendo exportar o TRM ou acessá-lo como API de outro projeto.

5.3.2.4. Diagrama de Banco de Dados do NERMAP

A modelagem de dados é usada para a especificação das regras de negócios e as estruturas de dados de um banco de dados. O NERMAP utiliza um banco de dados relacional (CODD, 1970), que armazena as informações no formato de tabelas

relacionadas entre si. O modelo relacional de banco de dados é o mais utilizado tanto no ambiente comercial quanto no acadêmico. O Diagrama Entidade Relacionamento (DER) – proposto por CHEN (1976) do banco de dados do NERMAP é apresentado na Figura 58.

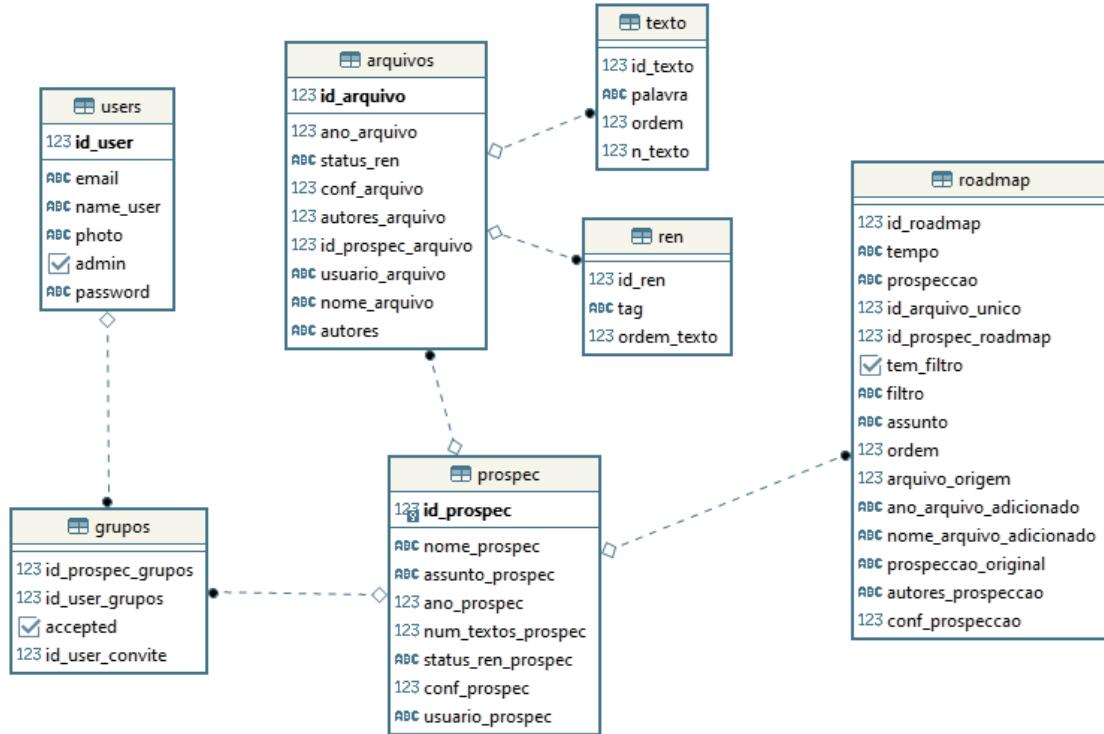


Figura 58. Diagrama Entidade Relacionamento do Banco de Dados do NERMAP.

Capítulo 6 – Avaliação do NERMAP

Neste capítulo é apresentada a avaliação do NERMAP conforme os preceitos do *Design Science*. A metodologia de avaliação é descrita detalhadamente, dividindo a avaliação em conceitos, modelos, métodos e implementação. Após a apresentação desses itens, é realizada a avaliação de cada um dos artefatos aplicando a metodologia de avaliação correspondente.

6.1. Metodologia

Este trabalho se utiliza do *framework Design Science* (MARCH; SMITH, 1995), que divide o processo de *Design Science* na construção e na avaliação de conceitos, modelos, métodos, e instâncias. O processo de avaliação desses artefatos deve ser condizente com os seus objetivos de medir a fidelidade com os processos do mundo real, assim como avaliar sua completeza e consistência. O *Design Science* concilia as atividades de pesquisa com seus resultados, ao propor uma metodologia unificada para ambos, que consiste dos seguintes processos: construir, avaliar, teorizar e justificar.

O *Design Science* é utilizado para produzir artefatos tecnológicos, que podem ser separados em quatro tipos: conceitos, modelos, métodos e implementações. Os **conceitos** são a linguagem básica para caracterizar o fenômeno estudado. Um **modelo** é a combinação dos conceitos, arranjados de forma a descrever artefatos ou atividades. **Métodos** são maneiras de organizar atividades de forma a atingir o objetivo esperado. **Implementação** é a materialização dos artefatos anteriores em produtos específicos, acessíveis e potencialmente utilizáveis na prática – uma instância. O *Design Science* possui duas atividades: construir e avaliar. A construção é o processo de criar um artefato para um propósito específico e a avaliação é o processo de verificação do desempenho do artefato criado (MARCH; SMITH, 1995). A avaliação do artefato é realizada pela análise de sua completeza, simplicidade, elegância, inteligibilidade e facilidade de uso. A relevância da contribuição de uma pesquisa de *Design Science* é relacionada com a novidade do artefato e da maneira com o qual atinge seus objetivos. O avanço ocorre no *Design Science* quando uma tecnologia é substituída por uma mais eficiente.

A atividade de pesquisa que se utiliza do *Design Science* tem como finalidade a melhoria dos processos existentes no mundo real e é composta de etapas que buscam construir e avaliar os artefatos produzidos para o fim a que se destinam. Construir se

baseia em demonstrar que um artefato pode ser construído; avaliar refere-se ao desenvolvimento de critérios de avaliação dos artefatos e executar a sua avaliação conforme esses próprios critérios.

De modo facultativo, esses processos podem ser realizados em conjunto com atividades de pesquisa no domínio das ciências naturais (*natural science*), com o intuito de extrair conhecimento através da proposição e validação de teorias; por esse motivo, são denominadas teorizar e justificar. Teorizar visa à construção de teorias que expliquem como ou por que algo acontece; portanto, consiste basicamente em explicar como ou por qual razão um artefato funciona de forma adequada no contexto ao qual faz parte. Justificar refere-se a provar a teoria anterior, requerendo evidência científica que a suporta ou refuta (MARCH; SMITH, 1995).

Tendo como base esses conceitos, a metodologia de avaliação do NERMAP é apresentada detalhadamente na Tabela 15. Com o objetivo de facilitar o entendimento da metodologia de avaliação utilizada nesta pesquisa, dividimos o processo de construção em objetivos iniciais e seus resultados – os artefatos construídos. De maneira similar, o processo de avaliação foi dividido em métricas e metodologias de avaliação utilizadas.

Tabela 15. Metodologia de Avaliação. Adaptado de MARCH e SMITH (1995).

	Construção		Avaliação	
	Objetivo	Resultados	Métricas	Metodologias
Conceitos	Identificar os conceitos básicos que compõem o processo de <i>Roadmapping</i> Tecnológico no contexto de um estudo de prospecção tecnológica	Definição dos elementos que compõem o processo de <i>Roadmapping</i> Tecnológico	Completeza, inteligibilidade	Argumento Lógico
Modelo	Descrever o modelo de semiautomatizar o processo de <i>Roadmapping</i> Tecnológico	O modelo NERMAP	Fidelidade com o mundo real, inteligibilidade, completeza, precisão, consistência interna	Cenário Ilustrativo, Avaliação de Desempenho
Método	Descrever um Processo de <i>Roadmapping</i> Tecnológico	O processo NERMAP	Fidelidade com o mundo real, completeza, consistência interna	Cenário Ilustrativo
Implementação	Aplicar o <i>framework</i> em aplicações reais	O sistema NERMAP	Aplicabilidade	Protótipo e Cenário Ilustrativo

Com relação à seleção de metodologias de avaliação, utilizamos o estudo de PEFFERS et al. (2012), que analisou os tipos de métodos de avaliação, os tipos de artefatos produzidos e o histórico da aplicação de cada método de avaliação para cada artefato produzido. Com base na análise da aplicabilidade de cada um dos métodos para os artefatos produzidos nesta pesquisa (conceitos, modelos, métodos e implementação), foram selecionados os métodos listados na Tabela 15, definidos como:

- Argumento Lógico: Um argumento com validade aparente (*face validity*);
- Cenário Ilustrativo: Aplicação de um artefato em uma situação real ou sintética, ilustrando sua adequabilidade ou utilidade;
- Estudo de Caso: Aplicação de um artefato em uma situação real, avaliando seu efeito nesta situação real;
- Protótipo: Implementação de um artefato tecnológico, ilustrando sua adequabilidade ou utilidade.

A seguir, são analisados e avaliados os artefatos produzidos neste estudo, sob dois agrupamentos: elementos teóricos (conceitos e modelo) e aplicações práticas/reais (método e instanciação).

6.2. Avaliação dos Conceitos

O processo de avaliação dos conceitos do NERMAP foi realizado utilizando a metodologia de Argumento Lógico, com o objetivo de demonstrar a completeza e inteligibilidade dos conceitos apresentados.

Os Usuários, que, neste estudo, são praticantes do processo de *Roadmapping* Tecnológico no escopo de um estudo de prospecção tecnológica, constituem os atores necessários deste processo, interagindo com o NERMAP para semiautomatizar a geração do *roadmap*.

A Interface do Usuário é responsável pela interação do pesquisador e o sistema, incluindo a inserção dos subsídios (documentos de entrada) que serão processados durante a geração do *roadmap*.

O processo de geração do *roadmap* inclui quatro etapas principais: Conversão para texto plano, Tagueamento, Fatiamento do texto e Atribuição de datas. A Conversão para texto plano é responsável por converter os documentos no formato PDF para TXT, gerando um arquivo limpo e sem formatação. O Tagueamento é responsável por aplicar o método de REN com base no modelo treinado com as EN. O Fatiamento do texto é

encarregado de selecionar as prospecções candidatas a partir do tagueamento da etapa anterior. Finalmente, a Atribuição de data é responsável por analisar e datar cada prospecção candidata, organizando as mesmas em uma série temporal – o *roadmap*.

6.3. Avaliação do Modelo

O processo de avaliação do Modelo NERMAP foi realizado utilizando a metodologias de Cenário Ilustrativo, com o objetivo de demonstrar a fidelidade com o mundo real, a inteligibilidade, a completeza, e a consistência interna do modelo apresentado. Também foram geradas métricas de Avaliação de Desempenho para o Tagueamento.

No cenário ilustrativo, vamos considerar que um pesquisador pretende prospectar a partir de três documentos: DOC1, DOC2, DOC3. Esses documentos contêm uma série de afirmações que podem, ou não, serem prospecções tecnológicas. De início, o pesquisador fornece os três documentos ao NERMAP. Esses documentos serão então processados, seguindo as quatro etapas do modelo: Conversão para texto plano, Tagueamento, Fatiamento do texto e Atribuição de datas, como mostrado na Figura 59.

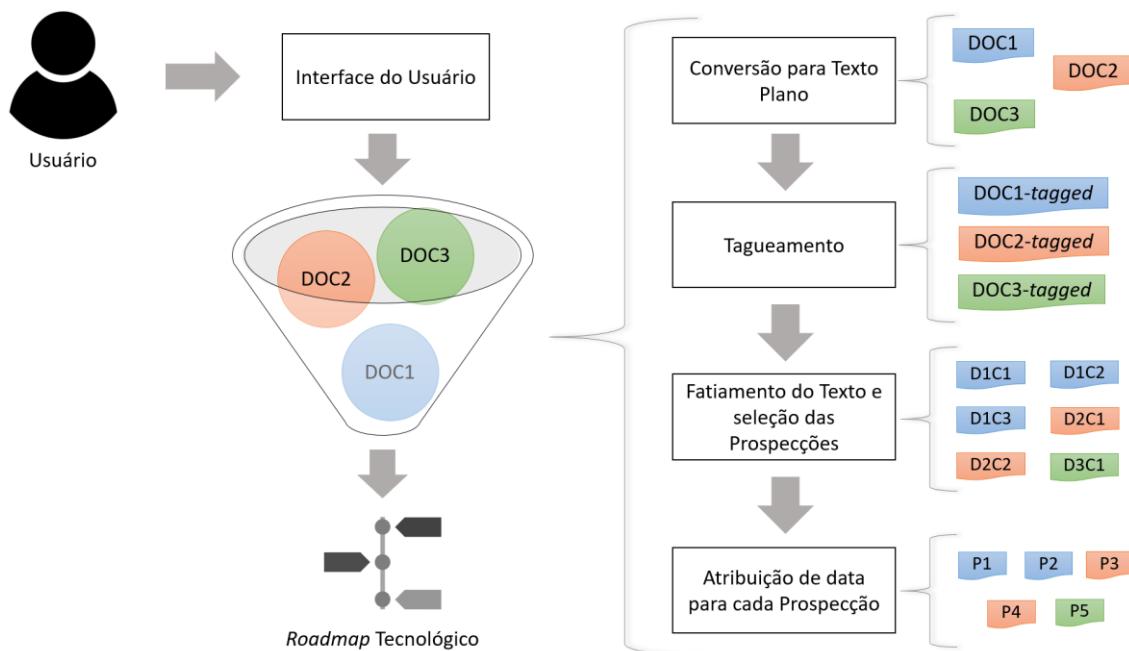


Figura 59. Aplicação do Modelo de geração do *roadmap* para o cenário proposto.

Na etapa de Conversão para texto plano, os documentos serão pré-processados e toda a formatação removida. Esse passo é necessário para o correto tagueamento do texto.

Na etapa de Tagueamento, o modelo REN desenvolvido neste trabalho é aplicado aos documentos, de forma a gerar uma cópia do documento original, contudo

devidamente marcado com as entidades relativas à prospecção tecnológica. Esse passo é fundamental e é onde as medidas de precisão, abrangência e medida-F são calculadas. Podemos dizer que no final deste passo temos os documentos *DOC1-tagged*, *DOC2-tagged* e *DOC3-tagged*.

Na etapa de Fatiamento do texto, os documentos tagueados (*DOC1-tagged*, *DOC2-tagged* e *DOC3-tagged*) são segmentados em frases simples (determinadas do início da frase ao ponto final) as prospecções candidatas detectadas no passo anterior. Ao final deste passo, teremos as prospecções candidatas D1C1, D1C2, D1C3 (prospecções candidatas relativas ao *DOC1*), D2C1, D2C2 (prospecções candidatas relativas ao *DOC2*) e D3C1 (prospecção candidata relativa ao *DOC3*).

Na etapa de Atribuição de datas, cada prospecção candidata é analisada, considerando a data do documento original, a data final do TRM e as datas tagueadas nas prospecções candidatas, dando origem às prospecções P1, P2, P3, P4 e P5. Desta forma, é possível excluir prospecções sobre o passado ou sobre o futuro pós-TRM.

No final essas prospecções datadas são ordenadas de forma crescente, formando um *roadmap*. Os *roadmaps* gerados podem ser visualizados de forma única para cada documento ou gerar um composto de *roadmaps* dentro de um mesmo processo de TRM, como representado na Figura 60.

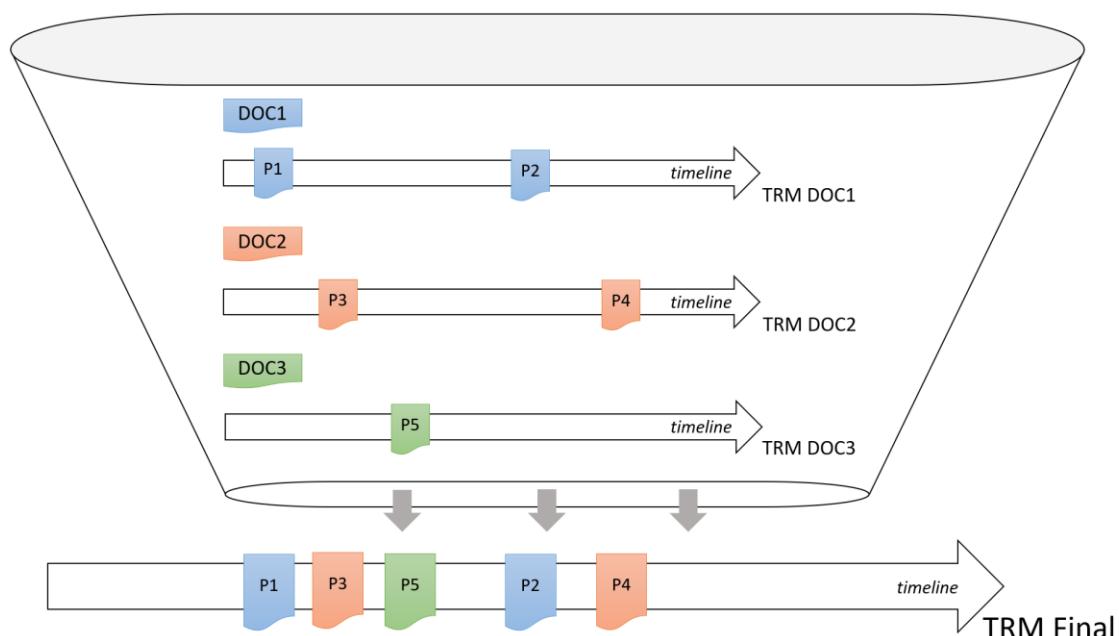


Figura 60. Composto de *roadmaps* para o cenário proposto.

6.3.1. Métricas do Tagueamento

A etapa de Tagueamento do texto inserida no Modelo NERMAP utiliza REN para realizar as anotações no texto conforme o modelo treinado. Esse método é avaliado seguindo medidas de Avaliação e Desempenho detalhadas na seção 4.3.4. Nesta seção, são apresentadas as medidas obtidas para o modelo desenvolvido.

A formação dos corpora e do modelo de REN envolveram publicações da *Futures* (Elsevier) entre os anos de 2010 e 2020. Ao todo foram 445 publicações coletadas e tagueadas manualmente com as EN definidas.

Fizeram parte do conjunto de testes as publicações inerentes aos anos de 2019 e 2020, totalizando 175 arquivos. E fizeram parte do conjunto de treino do modelo o restante das publicações – entre os anos de 2010 e 2011 – totalizando 270 documentos.

Como detalhado na seção 5.1.1.2, foi utilizada a notação BILOU para o tagueamento das EN e o resultado do experimento que reuniu os melhores resultados para os parâmetros testados – descrito na seção 5.1.1.3.1 – é apresentado na Tabela 16.

Tabela 16. Resultado do experimento para as EN.

Entidade	Corpus	Precisão	Abrangência	Medida-F
B_TEMPPRED	<i>Futures</i>	0,4815	0,3611	0,4127
I_TEMPPRED	<i>Futures</i>	0,7778	0,5833	0,6667
L_TEMPPRED	<i>Futures</i>	0,3913	0,5625	0,4615
U_TEMPPRED	<i>Futures</i>	0,8043	0,7551	0,7789

Essas entidades foram as utilizadas para o desenvolvimento do corpus, que ao final foi unificado ao modelo do Stanford NER para obter as entidades DATE e DURATION, já treinados e avaliados pela ferramenta, com um alto valor de eficácia. O corpus do modelo NERMAP é composto das 3 EN: TEMPPRED, DATE e DURATION, seguindo a notação BILOU para o tagueamento.

O resultado do experimento para o modelo desenvolvido é mostrado na Tabela 17.

Tabela 17. Resultado do experimento para o Modelo NERMAP.

Corpus	Precisão	Abrangência	Medida-F
<i>Futures</i>	0,6923	0,6290	0,6592

O processamento do NERMAP identifica as prospecções por meio da ocorrência das tags TEMPPRED + DATE ou DURATION em cada parágrafo do texto. As datações

ocorrem se utilizando como base a data da publicação junto com as informações da prospecção e o ano limite estabelecido para o TRM.

6.4. Avaliação do Método

O processo de avaliação do método NERMAP foi realizado utilizando a metodologia de Cenário Ilustrativo, de maneira semelhante à avaliação do modelo NERMAP.

Nesta seção são apresentados processos que são compatíveis com o modelo discutido na seção 6.3. O Cenário Ilustrativo é relativo ao caso que um pesquisador pretende prospectar a partir de três documentos: DOC1, DOC2, DOC3. O processo é representado na Figura 61.

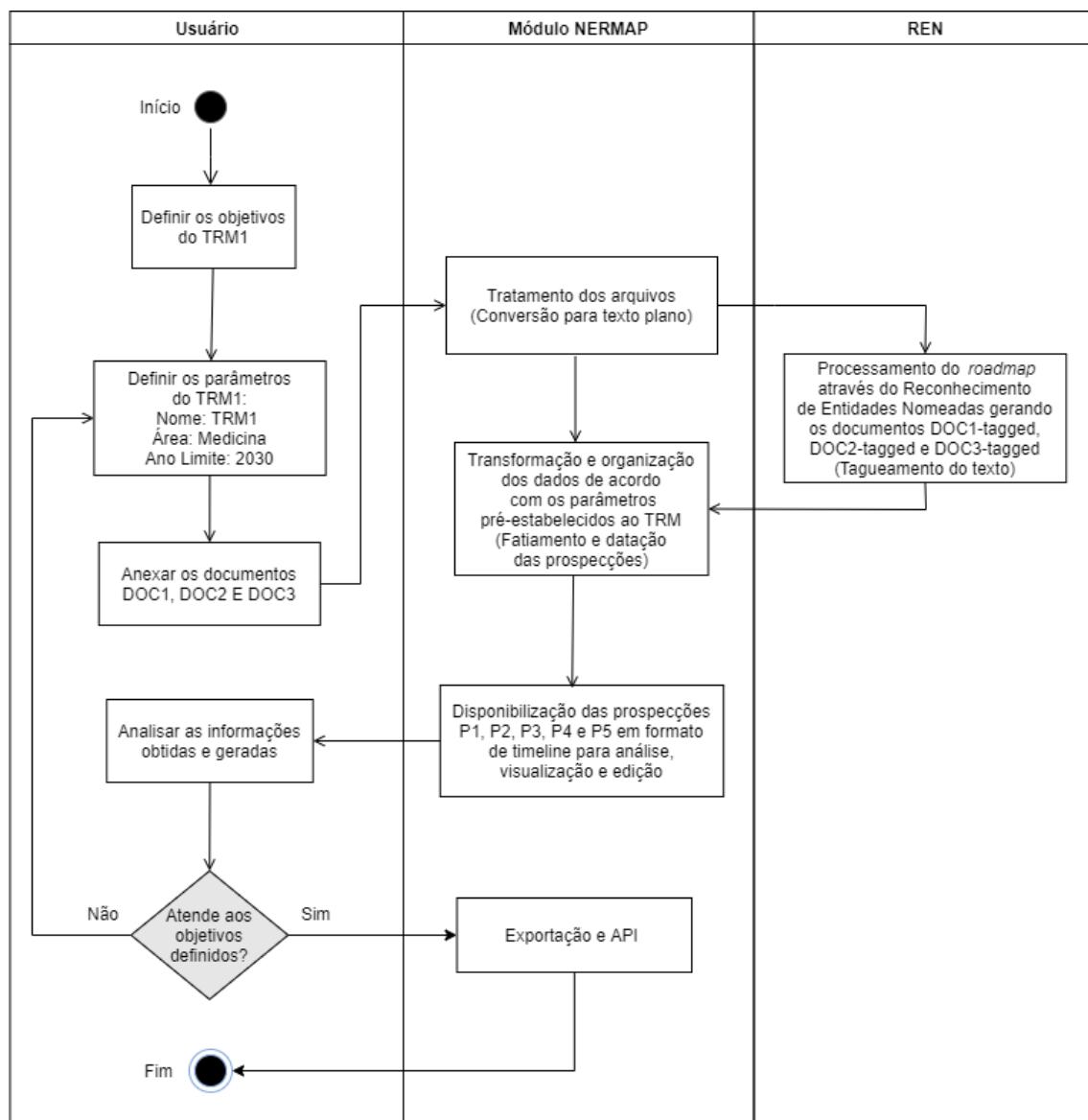


Figura 61. Processo NERMAP aplicado ao cenário proposto.

O processo se inicia com a definição dos objetivos a serem alcançados com TRM dentro do estudo de prospecção tecnológica em que será aplicado. A etapa seguinte configura a definição dos parâmetros do TRM, que neste cenário ilustrativo será: Nome: TRM1; Área: Medicina; Ano Limite: 2030.

O pesquisador, na próxima etapa do processo, submete os arquivos DOC1, DOC2 e DOC3 ao NERMAP, incluindo os metadados como Título da Publicação, Ano da Publicação, Autores e Confiabilidade da Fonte. Essas informações são importantes para a análise das prospecções a serem geradas e do *roadmap* final.

Na etapa de Tratamento dos arquivos, os documentos são processados para serem convertidos em texto plano, sem formatação. Em seguida, é realizado o Tagueamento do texto aplicando o método de REN com o modelo treinado. Nesta etapa serão gerados os arquivos *DOC1-tagged*, *DOC2-tagged* e *DOC3-tagged*.

Na etapa de Transformação e Organização o texto é segmentado em frases simples, gerando as prospecções candidatas D1C1, D1C2, D1C3 (prospecções candidatas relativas ao DOC1), D2C1, D2C2 (prospecções candidatas relativas ao DOC2) e D3C1 (prospecção candidata relativa ao DOC3). Com base em uma análise de acordo com os parâmetros pré-estabelecidos do TRM e os metadados de cada documento, as prospecções candidatas são refinadas de modo a excluir prospecções sobre o passado ou sobre o futuro pós-TRM, gerando as prospecções P1, P2, P3, P4 e P5.

As prospecções são organizadas em forma de *timeline* e disponibilizadas para o pesquisador, permitindo uma análise onde o *roadmap* pode ser refinado de forma manual, editando, excluindo ou adicionando prospecções.

Após a análise e refinamento manual, o pesquisador pode ajustar os parâmetros do TRM como, por exemplo, o Ano Limite e adicionar mais documentos. Tendo alcançado os objetivos definidos, o *roadmap* pode ser exportado nos formatos PDF, DOC, CSV, JSON e TXT ou disponibilizado como API para ser acessado de qualquer outro projeto.

6.5. Avaliação da Instanciação

O processo de avaliação da implementação do NERMAP foi realizado utilizando as metodologias de Protótipo e Estudo de Caso.

6.5.1. Journal Futures (Estudo de Caso)

O Estudo de Caso para a avaliação do sistema NERMAP apresentado nesta seção é relativo à base de publicações do periódico *Futures* (Elsevier). Foram coletadas as publicações referentes ao ano de 2020 que estavam disponíveis até a presente data do experimento.

Ao todo 55 publicações foram analisadas manualmente de modo a identificar as prospecções tecnológicas contidas no texto. A análise dessas publicações está exposta no Anexo 1, junto com a avaliação do experimento detalhada nesta seção.

Para o experimento, os documentos foram submetidos ao NERMAP, servindo de subsídio para o processo de TRM, conforme ilustrado na Figura 62.

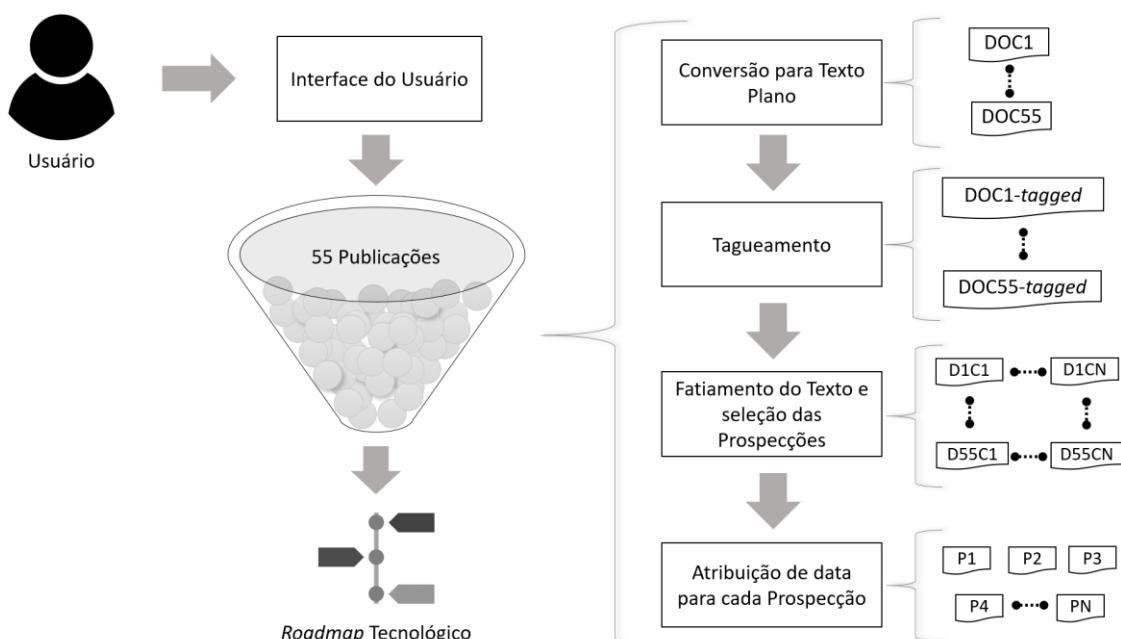


Figura 62. Aplicação do Modelo de geração do *roadmap* para o Estudo de Caso.

O experimento foi dividido em ciclos, todos com base nesse estudo de caso, tendo o primeiro deles servido como base para aprimoramentos no segundo, que por sua vez também serviu de base para aprimoramentos no terceiro. Esses ciclos são descritos nas próximas seções.

Os ciclos se baseiam em submeter ao NERMAP as 55 publicações da *Futures*, seguindo o processamento ilustrado na Figura 62, e analisar as prospecções retornadas. A análise se baseia em categorizar a prospecção retornada como: *Completa*, ou seja, apenas com o resultado da prospecção já podemos definir qual será o acontecimento previsto e a sua data; *Parcial*, ou seja, apresenta uma prospecção incompleta, sendo necessário usar a função de busca do NERMAP para ler mais trechos antes ou depois do parágrafo

selecionado e completar a prospecção; *Inválida*, ou seja, um trecho que não se configura como prospecção. Além disso, é avaliado se a prospecção retornada pelo NERMAP também foi encontrada na análise manual realizada nas publicações, ou seja, se ela está presente no Anexo 1.

6.5.1.1. Ciclo 1

No ciclo 1 foi realizada a análise das prospecções obtidas pelo NERMAP e a validação das prospecções encontradas manualmente.

A validação das prospecções encontradas manualmente – verificação se o NERMAP retornou determinada prospecção – avaliou cada caso em específico para detalhar os problemas e limitações do sistema. Com isso, as prospecções foram categorizadas da seguinte forma:

- **Encontrado:** 27 prospecções (36%) foram retornadas pelo NERMAP;
- **Não encontrado:**
 - **Falha:** O NERMAP falhou ao não retornar 14 prospecções (18%);
 - **Erro na conversão do arquivo:** 13 prospecções (17%) não foram retornadas pelo NERMAP por falha no processamento do arquivo para conversão de PDF para texto plano;
 - **Caso não pertence ao modelo de treino:** 9 prospecções (12%) não foram retornadas pelo NERMAP por conta de o treinamento do modelo não conter o caso específico da prospecção;
 - **Caso não pertence ao modelo de treino e arquivo apresentou erro na conversão:** 6 prospecções (8%), não foram retornadas por conta de o treinamento do modelo não conter o caso específico da prospecção e o arquivo apresentar falha na conversão de PDF para texto plano;
 - **Precisa de tratamento para o caso e arquivo apresentou erro na conversão:** 3 prospecções (4%) seriam tagueadas corretamente pelo modelo, porém precisam de tratamento para casos específicos que apresentaram problemas e o arquivo apresentou falha na conversão de PDF para texto plano;
 - **Precisa de tratamento para o caso:** 2 prospecções (3%) foram tagueadas corretamente pelo modelo, porém precisam de tratamento para casos específicos que apresentaram problemas;

- **Fora do escopo do ano limite do TRM:** 1 prospecção (1%) não foi retornada por estar fora do ano limite estabelecido para o TRM no experimento;
- **Fora do escopo do ano limite do TRM e arquivo apresentou erro na conversão:** 1 prospecção (1%) além de estar fora do ano limite estabelecido para o TRM no experimento, não foi retornada por falha na conversão de PDF para texto plano.

Essa validação pode ser observada na coluna 3 do Anexo 1 e o gráfico que reflete essa análise é ilustrado na Figura 63.

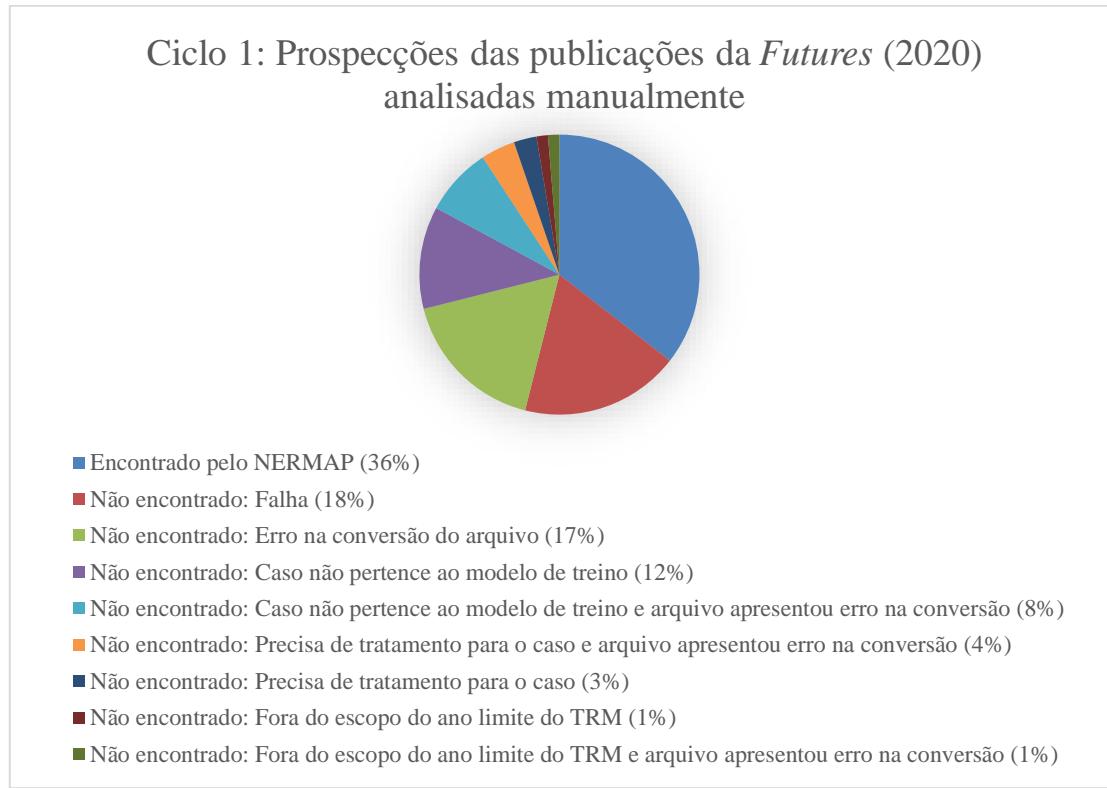


Figura 63. Ciclo 1: Gráfico das prospecções das publicações da *Futures* (2020) analisadas manualmente.

Os resultados obtidos na validação das prospecções encontradas manualmente para o ciclo 1 revelou o problema no processamento dos documentos, que representou, no total, 30% dos erros para encontrar as prospecções. Além disso, os casos que necessitam de tratamento para peculiaridades de cada prospecção e o escopo temporal do TRM podem ser tratados para aprimoramento dos resultados.

A análise dos resultados para as prospecções obtidas pelo NERMAP, é representada no Anexo 2 e o gráfico que reflete essa análise é ilustrado na Figura 64.

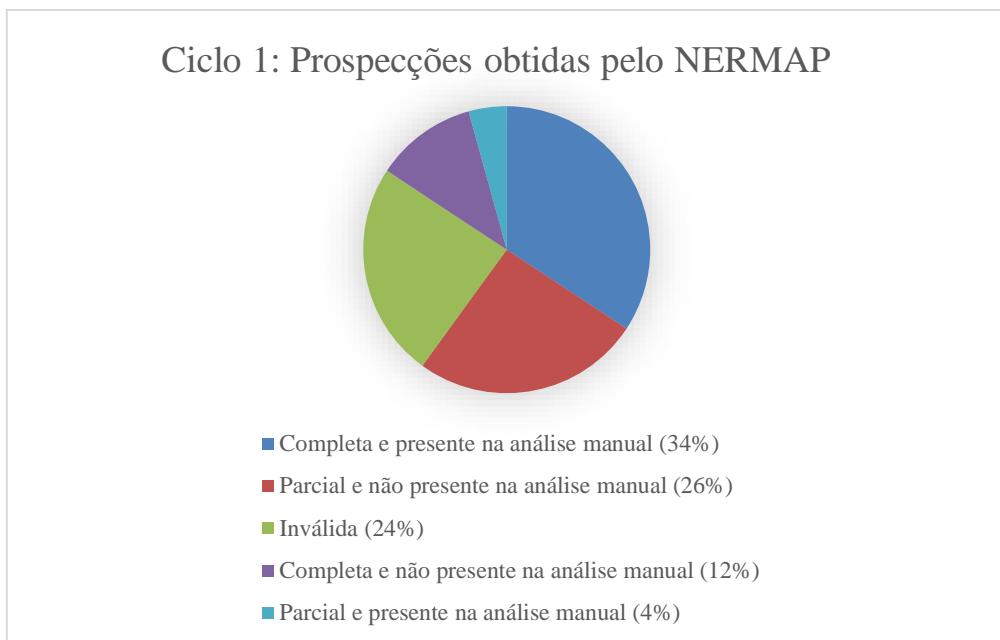


Figura 64. Ciclo 1: Gráfico das prospecções obtidas no NERMAP.

A análise é impactada pelos problemas citados na validação das prospecções encontradas manualmente e mostra que, dentre as 70 prospecções retornadas pelo NERMAP, a maioria – 24 prospecções (34%) – representa prospecções completas e que foram encontradas na análise manual; 18 prospecções (26%) parciais e não encontradas na análise manual; 17 prospecções (24%) inválidas; 8 (12%) prospecções completas e não encontradas na análise manual e 3 prospecções (4%) parciais e encontradas na análise manual.

Prospecções tanto completas quanto parciais retornadas e que não foram encontradas manualmente, demonstram que o NERMAP é capaz de retornar resultados que passam despercebidos por análises manuais, dessa forma auxiliando no processo de montar um *roadmap* mais completo.

6.5.1.2. Ciclo 2

No ciclo 2, os problemas descobertos no ciclo 1 foram mitigados, de forma a aumentar a porcentagem de prospecções que foram encontradas manualmente e que foram retornadas pelo NERMAP.

No problema da conversão de arquivos, as oito publicações que apresentaram falha para serem convertidas de forma automática do formato PDF para o formato TXT (texto plano) no ciclo 1, foram convertidos de forma manual. Já os casos que pertenciam ao modelo de treino e foram tagueados corretamente, porém necessitavam de tratamento para serem identificados pelo NERMAP como prospecções, foram tratados.

O principal erro por necessidade de tratamento ocorreu quando uma prospecção apresentava mais de uma palavra/expressão identificada como data. Desse modo, foi optado por selecionar a data mais distante da publicação e que estivesse dentro do escopo do TRM – as prospecções 43, 45, 59, 61 e 76 do Anexo 1 representam esses casos.

As correções representaram um aumento na validação das prospecções tagueadas manualmente com as que foram retornadas pelo NERMAP. As prospecções encontradas obtiveram um aumento de 39%, saltando de 36% para 75%. O gráfico que reflete o resultado do segundo ciclo é representado na Figura 65.

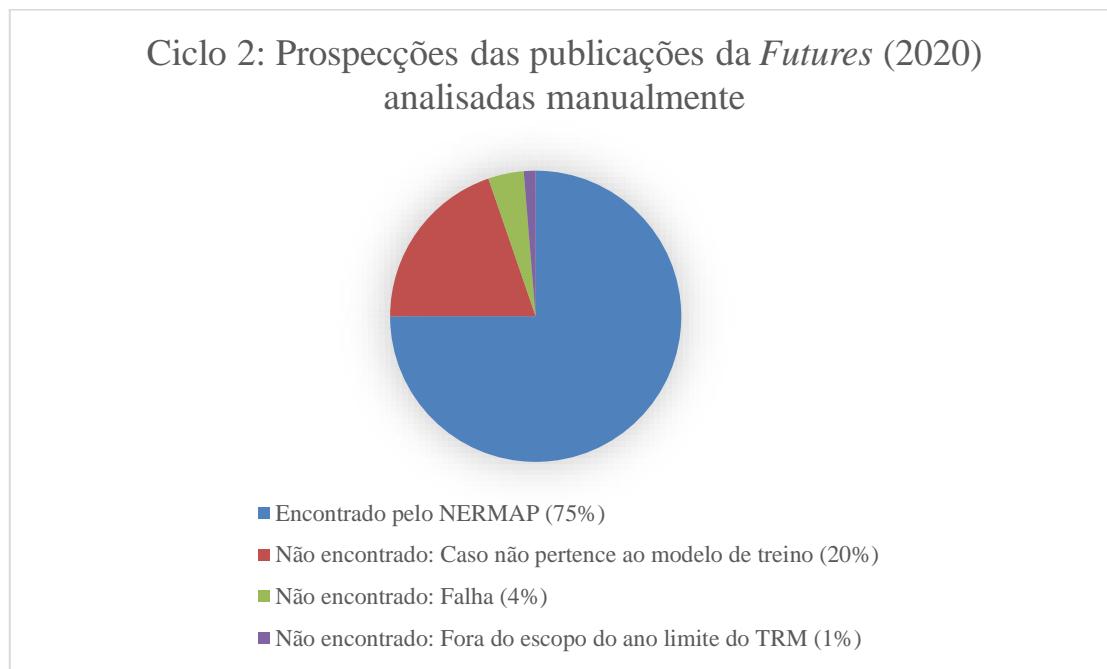


Figura 65. Ciclo 2: Gráfico das prospecções das publicações da *Futures* (2020) analisadas manualmente.

Com o tratamento para os problemas observados com o ciclo 1, o NERMAP retornou um total de 117 prospecções, sendo 47 (67%) a mais do que no ciclo 1.

A análise dos resultados para as prospecções obtidas pelo NERMAP, é representada no Anexo 3 e o gráfico que reflete essa análise é mostrado na Figura 66.

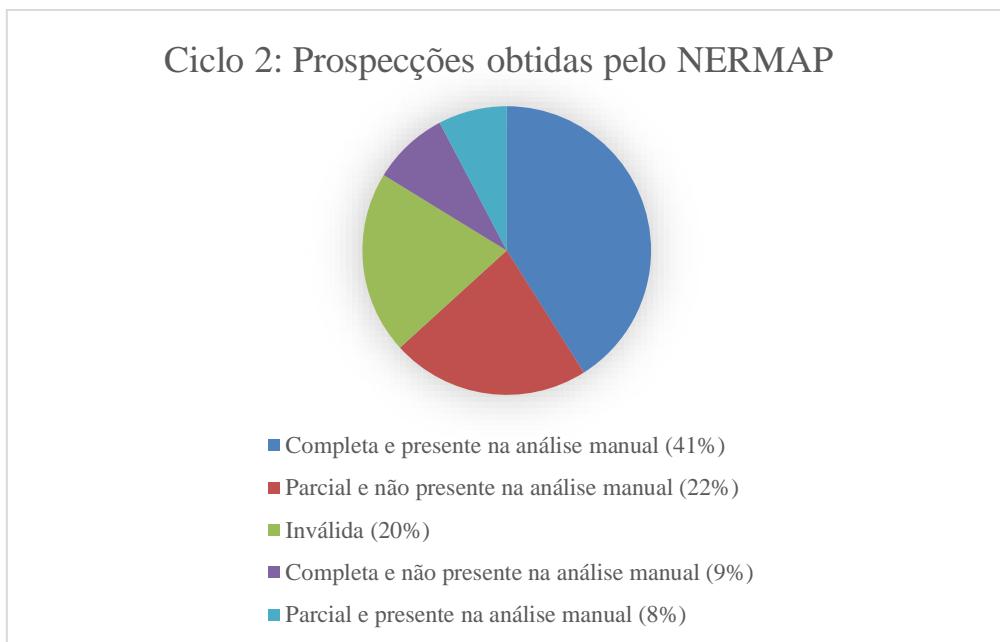


Figura 66. Ciclo 2: Gráfico das prospecções obtidas no NERMAP.

Em comparação com o ciclo 1, o NERMAP encontrou um número maior de prospecções, tanto completas quanto parciais. As prospecções presentes na análise manual obtiveram um ganho de 11% (38% no ciclo 1 e 49% no ciclo 2). Já as prospecções detectadas como inválidas e que não deveriam ser retornadas pelo NERMAP decaíram de 24% para 20%.

As prospecções completas e não presentes na análise manual demonstram que o NERMAP é capaz de encontrar prospecções que o ser humano manualmente pode deixar passar despercebida. Além disso, as prospecções parciais que não foram validadas manualmente representam possíveis prospecções a serem refinadas que colaboram com o resultado do *roadmap*.

6.5.1.3. Ciclo 3

No ciclo 3, o foco foi direcionado para a melhoria do modelo de treino para o REN, de modo a buscar os casos omissos no modelo atual aplicados aos ciclos 1 e 2. As publicações do *Journal Futures* pertencentes aos anos de 2012 a 2018, que totalizam 646 documentos, foram analisadas manualmente para encontrar novos exemplos similares aos casos que não pertenciam ao modelo de treino.

Os casos omissos buscados nessas publicações podem ser observados nos índices 2, 9, 10, 17, 19, 21, 25, 26, 31, 33, 48, 70, 71, 72 e 74 do Anexo 1.

O novo modelo que conglorou prospecções anotadas manualmente das publicações de 2010 a 2018 do *Journal Futures* obteve as métricas apresentadas na Tabela

18, quando aplicado às publicações dos anos de 2019 e 2020. Nota-se um ganho de desempenho em comparação com o modelo utilizado nos ciclos 1 e 2 (Tabela 17).

Tabela 18. Resultado do experimento para o novo Modelo NERMAP.

Corpus	Modelo	Precisão	Abrangência	Medida-F
<i>Futures</i>	Ciclos 1 e 2	0,6923	0,6290	0,6592
<i>Futures</i>	Ciclo 3	0,7079 (+ 2,25%)	0,6774 (+ 7,7%)	0,6923 (+ 5%)

A análise das publicações coletadas manualmente para o ano de 2020 tendo em vista o que foi retornado pelo NERMAP para o mesmo grupo de documentos é exposta, graficamente, na Figura 67.

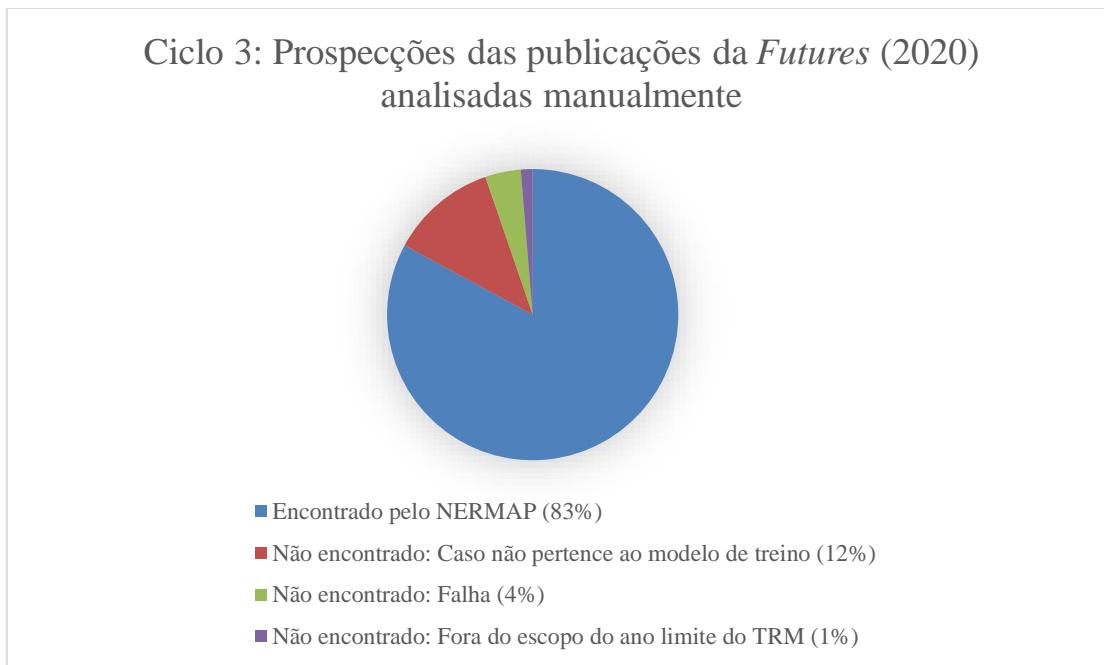


Figura 67. Ciclo 3: Gráfico das prospecções das publicações da *Futures* (2020) analisadas manualmente.

Nem todos os casos omissos foram encontrados entre 2012 e 2018. Dessa forma, 12% das prospecções – 8% menos que no ciclo 2 – ainda não puderam ser retornadas pelo NERMAP pois seus casos específicos de tagueamento não estão presentes no modelo de treino. Em contrapartida, houve um ganho de publicações encontradas pelo NERMAP, totalizando 83%. Os casos não encontrados ou por falha ou por estarem fora do escopo do TRM, se mantiveram em 4% e 1%, respectivamente.

No ciclo 3 o NERMAP retornou 125 prospecções, sendo 8 a mais do que no ciclo 2. Desses publicações, 115 estão presentes no ciclo 2 e 10 se configuraram como novas prospecções retornadas pelo NERMAP. Além disso, 2 prospecções inválidas retornadas no ciclo 2 – índices 54 e 87 do Anexo 3 – não foram retornadas no ciclo 3.

Dessas 10 novas prospecções retornadas: 5 se configuram como completas e presentes na análise manual exposta no Anexo 1, 4 como parciais (sendo uma dessas prospecções presente no Anexo 1) e 1 inválida. O resultado completo para as 125 prospecções pode ser observado no Anexo 4 e, graficamente, na Figura 68.

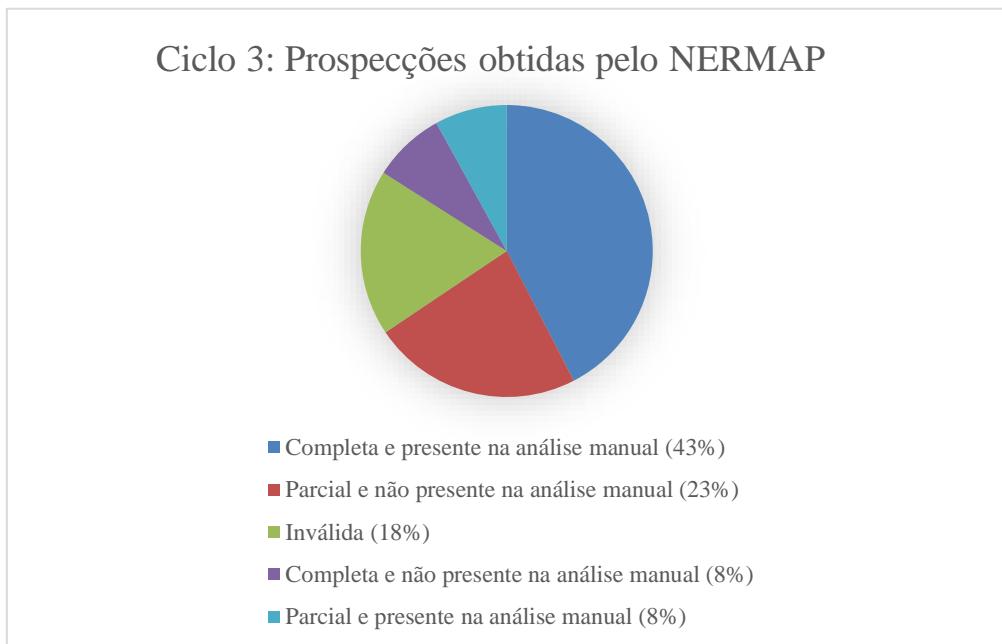


Figura 68. Ciclo 3: Gráfico das prospecções obtidas no NERMAP.

6.5.2. Discussão

Os ciclos do experimento mostraram um crescimento gradativo do *framework* NERMAP, sendo possível perceber problemas e novos casos de prospecções para somar ao modelo. Com base na análise das prospecções encontradas manualmente, o NERMAP evoluiu de 36% de publicações retornadas no ciclo 1 para 83% no ciclo 3.

Impulsionado por correções no processamento dos documentos e melhoria no modelo de treino do REN, o NERMAP aumentou o número de prospecções retornadas de 70 para 125 prospecções, do ciclo 1 ao ciclo 3, representando um ganho de 78.5%.

Dessas prospecções, foi observado durante os ciclos um aumento das prospecções consideradas completas e parciais e uma redução percentual das prospecções consideradas como inválidas e que não deveriam ser retornadas pelo NERMAP.

No resultado do ciclo 3, 51% das publicações retornadas foram também encontradas na análise manual nos artigos da Futures do ano de 2020. Entretanto, dentre o restante das prospecções, 31% consideradas completas ou parciais foram retornadas pelo NERMAP e não haviam sido encontradas manualmente. Isso representa o poder do

NERMAP em realizar um trabalho que, além de ser realizado em massa e em pouco tempo, pode encontrar prospecções que possam passar de maneira despercebida em uma análise manual.

O refinamento manual a ser realizado pelo pesquisador que utiliza o NERMAP se configura por remover as prospecções consideradas como inválidas e completar/remover as prospecções consideradas como parciais. As prospecções consideradas como completas representam um resultado perfeito do NERMAP de uma prospecção para fazer parte do *roadmap*, sem a necessidade de qualquer melhoria.

Em um contexto de um grande estudo de FTA, onde centenas de documentos são analisados, o NERMAP se mostrou uma ferramenta bastante útil para auxiliar o pesquisador de Prospecção Tecnológica, permitindo realizar o processo de *Roadmapping* Tecnológico de maneira rápida e para um grande conjunto de publicações.

Capítulo 7 – Considerações Finais

O *Roadmapping* Tecnológico configura um dos métodos pertencentes a um estudo de *Future-oriented Technology Analysis*, tendo como um produto esperado, o *roadmap*. Apresentando um cronograma temporal relacionando tecnologias e produtos, de forma a apresentar as prospecções tecnológicas previstas em um espaço de tempo.

O *framework* NERMAP, desenvolvido neste trabalho, pode contribuir de forma essencial no processo de TRM, de modo a semiautomatizá-lo fazendo uso da técnica de Reconhecimento de Entidades Nomeadas e permitindo que uma grande massa de documentos inseridos no âmbito de Prospecção Tecnológica possa ser analisada para gerar um *roadmap*.

Essa contribuição foi demonstrada de maneira teórica e prática ao longo deste trabalho, através dos artefatos que deram origem ao sistema NERMAP, sob orientação do *framework Design Science*. Esse *framework* consiste na construção e na avaliação de conceitos, modelos, métodos e instanciações, avaliados de modo a medir a sua fidelidade com os fenômenos do mundo real, sua completeza e sua consistência interna.

Como produtos do trabalho realizado durante essa pesquisa, podemos destacar o seguinte conjunto de contribuições:

- Uma revisão da literatura sobre *Future-oriented Technology Analysis* realizada no Capítulo 2 e uma revisão com ênfase para o método *Roadmapping* Tecnológico e seu produto, o *roadmap*, realizada no Capítulo 3;
- Uma revisão da literatura sobre Reconhecimento de Entidades Nomeadas e suas áreas relacionadas, dentro do escopo de Mineração de Textos, realizada no Capítulo 4;
- O *framework* NERMAP para semiautomatizar o processo de *Roadmapping* Tecnológico a ser executado num contexto de estudo de Prospecção Tecnológica, que consiste em conceitos, modelo e processo, à luz do *framework Design Science*, descritos detalhadamente no Capítulo 5 e avaliados no Capítulo 6;
- O sistema que representa a instanciação desse *framework*, denominado NERMAP, também descrito no Capítulo 5, cuja avaliação é discutida detalhadamente no Capítulo 6.

O sistema NERMAP obteve 83% de eficácia ao encontrar prospecções em textos, sendo possível também encontrar prospecções em um grande volume de documentos que, dificilmente poderiam ser encontradas manualmente. Dessa forma, o sistema proposto se configura uma ferramenta extremamente útil para o estudo de FTA, contribuindo para todo o processo de um TRM, além de prover uma estrutura para organizar, manter e compartilhar os *roadmaps* gerados.

Analizando este trabalho de maneira crítica, algumas limitações podem ser apontadas, como, por exemplo: O modelo treinado para uso no REN não contempla todos os casos possíveis de prospecções a serem encontradas pelo NERMAP para gerar *roadmaps*; O processamento de documentos apresentou falhas durante o experimento (ciclo 1) por conta da biblioteca utilizada no trabalho, sendo necessário uma melhoria na mesma ou busca por outras bibliotecas similares com melhor desempenho; Prospecções inválidas sendo retornadas pelo NERMAP (18% no ciclo 3) representam uma necessidade de melhoria no REN e no processamento das prospecções candidatas em prospecções datadas.

Portanto, à luz das contribuições e limitações mencionadas anteriormente, apresentam-se, a seguir, oportunidades de trabalhos futuros, como, por exemplo: O modelo usado pelo REN pode ser constantemente melhorado, inicialmente, utilizando como insumo as publicações dos anos do *Journal Futures* que foram utilizados para testes e avaliação do NERMAP (2019 e 2020), que não entraram para o modelo por conta de não o enviesar. A inclusão dos casos de prospecção observados nesses documentos tende a cobrir os 12% de prospecções não encontradas pelo NERMAP no ciclo 3 do experimento, como pode ser observado na Figura 67, de modo a atingir 95% de eficácia e aumentar as medidas de Precisão, Abrangência e Medida-F. Novos artigos da *Futures* além de outros documentos no âmbito de Prospecção Tecnológica também podem fazer parte do aprimoramento do modelo utilizado no NERMAP.

Quanto ao processamento de documentos, existe uma constante melhora na biblioteca de conversão de arquivos que, naturalmente, implicará em melhores resultados. Além de bibliotecas mais eficientes para outras linguagens, que podem ser implementadas de modo complementar ao NERMAP para converter uma gama maior de arquivos.

As prospecções inválidas sendo retornadas pelo NERMAP tendem a diminuir gradativamente em consonância com o aperfeiçoamento do modelo. Outrossim, com um uso maior do sistema NERMAP e novos resultados observados, a etapa de escolher as

prospecções candidatas em prospecções datadas pode avaliar casos mais propícios a serem inválidos e descartá-los previamente.

Assim, através deste trabalho, composto por todas as suas etapas de revisão literária, contextualização e proposta de solução, e construção dos artefatos segundo o *framework Design Science*, é oferecido como seu produto o sistema NERMAP, uma ferramenta para apoiar o processo de *Roadmapping* Tecnológico em um estudo de FTA.

Referências Bibliográficas

- ALBRIGHT, R. E. Roadmapping Convergence. In: BAINBRIDGE, W. S.; ROCO, M. C. (Eds.). . **Managing nano-bio-info-cognito innovations**. Dordrecht: Kluwer Academic Publishers, 2005. p. 23–31.
- ALLEN, J. **Natural Language Understanding**. 2. ed. United States: The Benjamin/Cummings Publishing Company, 1995.
- ALPAYDIN, E. **Introduction to machine learning**. Third edition ed. Cambridge, Massachusetts: The MIT Press, 2014.
- ALVES, A. **Maximização de entropia em linguística computacional para língua portuguesa**. Dissertação de Mestrado—São Paulo: Universidade de São Paulo, 2003.
- AMANATIDOU, E. Beyond the veil — The real value of Foresight. **Technological Forecasting and Social Change**, v. 87, p. 274–291, Setembro 2014.
- AMARAL, D. **O Reconhecimento de Entidades Nomeadas por meio de Conditional Random Fields para a Língua Portuguesa**. Porto Alegre: Pontifícia Universidade Católica do Rio Grande do Sul, 2013.
- AMARAL, D.; VIEIRA, R. NERP-CRF: uma ferramenta para o reconhecimento de entidades nomeadas por meio de Conditional Random Fields. **Linguamática**, v. 6, n. 1, p. 41–49, 2014.
- APPELT, D. E. et al. **FASTUS: A finite-state processor for information extraction from real-world text**. IJCAI. **Anais...** 1993
- ARANHA, C. N. **Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional**. Tese de Doutorado—Rio de Janeiro, RJ: Pontifícia Universidade Católica do Rio de Janeiro, 2007.
- BARBOSA, C. E. **TIAMAT: Um framework para apoiar a integração de métodos de Prospecção Tecnológica**. Tese de Doutorado—Rio de Janeiro: UFRJ, mar. 2018.
- BATISTA, D. S. et al. **Geographic signatures for semantic retrieval**. Proceedings of the 6th Workshop on Geographic Information Retrieval - GIR '10. **Anais...** In: THE 6TH WORKSHOP. Zurich, Switzerland: ACM Press, 2010Disponível em: <<http://portal.acm.org/citation.cfm?doid=1722080.1722104>>. Acesso em: 30 nov. 2020
- BENDER, O.; OCH, F. J.; NEY, H. **Maximum Entropy Models for Named Entity Recognition**. Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. **Anais...** 2003Disponível em: <<https://www.aclweb.org/anthology/W03-0420>>. Acesso em: 7 jun. 2020
- BEPPLER, M. D.; FERNANDES, A. M. DA R. Aplicação de Text mining para a Extração de Conhecimento Jurisprudencial. **Primeiro Congresso Sul Catarinense de Educação**, p. 8, 2005.
- BIKEL, D. M. et al. **Nymble: a high-performance learning name-finder**. Proceedings of the fifth conference on Applied natural language processing -. **Anais...** In: THE FIFTH CONFERENCE. Washington, DC: Association for Computational Linguistics, 1997Disponível em: <<http://portal.acm.org/citation.cfm?doid=974557.974586>>. Acesso em: 7 jun. 2020
- BLAIR, P. Technology assessment; current trends and the myth of a formula. **International Association of Technology Assessment and Forecasting Institutions**, 2 maio 1994.
- BOOCH, G.; RUMBAUGH, J.; JACOBSON, I. **The unified modeling language user guide**. Reading Mass: Addison-Wesley, 1999.
- BORTHWICK, A. et al. **NYU: Description of the MENE Named Entity System as Used in MUC-7**. Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax,

Virginia, April 29 - May 1, 1998. **Anais...** In: MUC 1998. 1998Disponível em: <<https://www.aclweb.org/anthology/M98-1018>>. Acesso em: 12 jun. 2020

BRIN, S. Extracting Patterns and Relations from the World Wide Web. In: ATZENI, P.; MENDELZON, A.; MECCA, G. (Eds.). **The World Wide Web and Databases**. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999. v. 1590p. 172–183.

BROWN, R.; PHAAL, R. **The use of technology roadmaps as a tool to manage technology developments and maximise the value of research activity**. In: IMECHE MAIL TECHNOLOGY CONFERENCE (MTC). Brighton: 24 abr. 2001

BUNESCU, R. C. **Learning for Information Extraction: From Named Entity Recognition and Disambiguation to Relation Extraction**. Ph.D. Thesis—Austin: University of Texas of Austin, 2007.

CAGNIN, C. et al. (EDS.). **Future-Oriented Technology Analysis**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

CARDOSO, N. F. P. F. **Avaliação de sistemas de reconhecimento de entidades mencionadas**. Dissertação de Mestrado,—Porto: Faculdade de Engenharia da Universidade do Porto., out. 2006.

CARVALHO, R. C. DE. **Aplicação de técnicas de mineração de texto na recuperação de informação clínica em prontuário eletrônico do paciente**. Dissertação (Mestrado em Ciência da Informação)—Marília: Unesp, 2017.

CHATZIS, S. P.; DEMIRIS, Y. The echo state conditional random field model for sequential data modeling. **Expert Systems with Applications**, v. 39, n. 11, p. 10303–10309, set. 2012.

CHEN, P. P.-S. The entity-relationship model—toward a unified view of data. **ACM Transactions on Database Systems**, v. 1, n. 1, p. 9–36, mar. 1976.

CHEUNG, D. W. et al. **A fast distributed algorithm for mining association rules**. Fourth International Conference on Parallel and Distributed Information Systems. **Anais...** In: 4TH INTERNATIONAL CONFERENCE ON PARALLEL AND DISTRIBUTED INFORMATION SYSTEMS. Miami Beach, FL, USA: IEEE Comput. Soc. Press, 1996Disponível em: <<http://ieeexplore.ieee.org/document/568665/>>. Acesso em: 12 jun. 2020

CHIEU, H. L.; NG, H. T. **Named Entity Recognition with a Maximum Entropy Approach**. Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. **Anais...** 2003Disponível em: <<https://www.aclweb.org/anthology/W03-0423>>. Acesso em: 7 jun. 2020

CHINCHOR, N.; HIRSCHMAN, L.; LEWIS, D. D. Evaluating Message Understanding Systems: An Analysis of the Third Message Understanding Conference (MUC-3). **Computational Linguistics**, v. 19, n. 3, p. 409–450, 1993.

CIRAVEGNA, F. Adaptive Information Extraction from Text by Rule Induction and Generalisation. **17th International Joint Conference on Artificial Intelligence**, v. 2, p. 6, 2001.

CODD, E. F. A Relational Model of Data for Large Shared Data Banks. v. 13, n. 6, p. 11, 1970.

COELHO, G. M. **Prospecção tecnológica: metodologias e experiências nacionais e internacionais**. Rio de Janeiro: INT/Finep/ANP Projeto CT-Petro, , 2003.

COELHO, G. M. et al. Prospecção de tecnologias de futuro: métodos, técnicas e abordagens. p. 41, dez. 2004.

COELHO, G. M. et al. Caminhos para o desenvolvimento em prospecção tecnológica: Technology Roadmapping – um olhar sobre formatos e processos. p. 36, dez. 2005.

COLLINS, M. **Ranking Algorithms for Named Entity Extraction: Boosting and the VotedPerceptron**. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. **Anais...** In:

ACL 2002. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, jul. 2002Disponível em: <<https://www.aclweb.org/anthology/P02-1062>>. Acesso em: 12 jun. 2020

COLLINS, M.; SINGER, Y. **Unsupervised Models for Named Entity Classification**. 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora. *Anais...*1999Disponível em: <<https://www.aclweb.org/anthology/W99-0613>>. Acesso em: 12 jun. 2020

COSTANTINO, M.; MORGAN, R. G.; COLLINGHAM, R. J. Financial information extraction using pre-defined and user-definable templates in the LOLITA System. **Journal of Computing and Information Technology**, v. 4, n. 4, p. 15, 1996.

COWIE, J.; LEHNERT, W. Information extraction. **Communications of the ACM**, v. 39, n. 1, p. 80–91, 1 jan. 1996.

CURRAN, J.; CLARK, S. **Language Independent NER using a Maximum Entropy Tagger**. Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. *Anais...*2003Disponível em: <<https://www.aclweb.org/anthology/W03-0424>>. Acesso em: 7 jun. 2020

DA SILVA, L. H.; CASELI, H. DE M. Reconhecimento de entidades nomeadas em textos em português do Brasil no domínio do e-commerce. p. 7, 2015.

DODDINGTON, G. et al. The Automatic Content Extraction (ACE) Program Tasks, Data, and Evaluation. p. 4, 2004.

EIRMA. **Technology roadmapping : delivering business vision**. Paris: EIRMA, 1997. Disponível em: <<https://trove.nla.gov.au/version/39783802>>. Acesso em: 17 set. 2019.

ELLOUMI, S. et al. General learning approach for event extraction: Case of management change event. **Journal of Information Science**, v. 39, n. 2, p. 211–224, abr. 2013.

ETZIONI, O. et al. Unsupervised named-entity extraction from the Web: An experimental study. **Artificial Intelligence**, v. 165, n. 1, p. 91–134, jun. 2005.

FAYYAD, U. et al. (EDS.). **Advances in knowledge discovery and data mining**. Menlo Park, Calif.: AAAI Press [u.a.], 1996.

FELDMAN, R.; DAGAN, I. Knowledge Discovery in Textual Databases (KDT). p. 6, 1995.

FELDMAN, R.; HIRSH, H. Exploiting Background Information in Knowledge Discovery from Text. **Journal of Intelligent Information Systems**, v. 9, n. 1, p. 83–97, 1997.

FINKEL, J. et al. Exploring the boundaries: gene and protein identification in biomedical text. **BMC Bioinformatics**, v. 6, n. Suppl 1, p. S5, 2005.

FINKEL, J. R.; GRENAGER, T.; MANNING, C. **Incorporating non-local information into information extraction systems by Gibbs sampling**. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05. *Anais...* In: THE 43RD ANNUAL MEETING. Ann Arbor, Michigan: Association for Computational Linguistics, 2005Disponível em: <<http://portal.acm.org/citation.cfm?doid=1219840.1219885>>. Acesso em: 14 jun. 2020

GAIZAUSKAS, R.; WILKS, Y. Information extraction: beyond document retrieval. **Journal of Documentation**, v. 54, n. 1, p. 70–105, mar. 1998.

GALVES, C.; BRITTO, H. **The Tycho Brahe Corpus of Historical Portuguese**. Campinas: Department of Linguistics, University of Campinas, 2002.

GALVIN, R. Roadmapping—A practitioner's update. **Technological Forecasting and Social Change**, v. 71, n. 1–2, p. 101–103, jan. 2004.

GARCIA, M. L.; BRAY, O. H. **Fundamentals of technology roadmapping**. [s.l.] Sandia National Laboratories, 1 abr. 1997. Disponível em: <<http://www.osti.gov/servlets/purl/471364-PDo152/webviewable/>>. Acesso em: 14 set. 2019.

GASPERIN, C. V.; LIMA, V. L. S. DE. **Fundamentos do Processamento Estatístico da Linguagem Natural**. Porto Alegre – RS – Brazil: PUCRS, 2001.

GERDSRI, N.; VATANANAN, R. S.; DANSAMASATID, S. Dealing with the dynamics of technology roadmapping implementation: A case study. **Technological Forecasting and Social Change**, v. 76, n. 1, p. 50–60, jan. 2009.

GRISHMAN, R. (ED.). **Information extraction: Techniques and challenges**. Frascati, Italy: Springer Verlag, 1997. v. 1299

GRISHMAN, R.; SUNDHEIM, B. **Message Understanding Conference- 6: A Brief History**. COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics. **Anais...** In: COLING 1996. 1996Disponível em: <<https://www.aclweb.org/anthology/C96-1079>>. Acesso em: 28 jul. 2019

GROENVELD, P. Roadmapping Integrates Business and Technology. **Research-Technology Management**, v. 50, n. 6, p. 49–58, nov. 2007.

GÜLBAHAR, Y. Technology planning: A roadmap to successful technology integration in schools. **Computers & Education**, v. 49, n. 4, p. 943–956, dez. 2007.

HARTIGAN, J. A.; WONG, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. **Applied Statistics**, v. 28, n. 1, p. 100, 1979.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. Unsupervised Learning. In: **The Elements of Statistical Learning**. 2. ed. New York: Springer, 2009. p. 485–585.

HORTON, A. Foresight: how to do simply and successfully. v. 1, n. 1, 1999.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. [s.l.] OTexts, 2014.

IMTR. **Information systems for the manufacturing enterprise**. Integrated manufacturing technology roadmapping project: 1999

ITRI. **Electronic Manufacturing and Packaging in Japan**. [s.l: s.n.].

JIANG, J. Information Extraction from Text. In: AGGARWAL, C. C.; ZHAI, C. (Eds.). . **Mining Text Data**. Boston, MA: Springer US, 2012. p. 11–41.

JOHNSTON, R. Historical Review of the Development of Future-Oriented Technology Analysis. In: CAGNIN, C. et al. (Eds.). . **Future-Oriented Technology Analysis**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 17–23.

JUNG, J. J. Online named entity recognition method for microtexts in social networking services: A case study of twitter. **Expert Systems with Applications**, v. 39, n. 9, p. 8066–8070, jul. 2012.

KAO, A.; POTEET, S. R. (EDS.). **Natural language processing and text mining**. London: Springer, 2007.

KAPPEL, T. A. Perspectives on roadmaps: how organizations talk about the future. **Journal of Product Innovation Management**, v. 18, n. 1, p. 39–50, jan. 2001.

KONKOL, M. **Named Entity Recognition**. Czech Republic: University of West Bohemia in Pilsen, jun. 2012.

KOSTOFF, R. N.; SCHALLER, R. R. Science and technology roadmaps. **IEEE Trans. Engineering Management**, v. 48, p. 132–143, 2001.

KRENN, B.; SAMUELSSON, C. **The Linguist's Guide to Statistics: Don't Panic.** [s.l.] Los autores, 1997.

KROCH, A.; TAYLOR, A. **The Penn-Helsinki Parsed Corpus of Middle English (PPCME2)** Philadelphia: University of Pennsylvania, , 2000.

KUMAR, S.; HEBERT, M. Discriminative Fields for Modeling Spatial Dependencies in Natural Images. In: THRUN, S.; SAUL, L. K.; SCHÖLKOPF, B. (Eds.). . **Advances in Neural Information Processing Systems 16.** [s.l.] MIT Press, 2004. p. 1531–1538.

KUSHMERIK, N. Gleaning the Web. **IEEE Intelligent Systems**, v. 14, n. 2, p. 20–22, mar. 1999.

LAFFERTY, J.; MCCALLUM, A.; PEREIRA, F. C. N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. **18th International Conference on Machine Learning (ICML)**, p. 10, 2001.

LEE, J. H.; PHAAL, R.; LEE, C. An empirical analysis of the determinants of technology roadmap utilization: Determinants of technology roadmap utilization. **R&D Management**, v. 41, n. 5, p. 485–508, nov. 2011.

LIU, Y. et al. Segmentation Conditional Random Fields (SCRFs): A New Approach for Protein Fold Recognition. In: MIYANO, S. et al. (Eds.). . **Research in Computational Molecular Biology**. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. v. 3500p. 408–422.

MACINTOSH, A.; FILBY, I.; TATE, A. **Knowledge Asset Road Maps**. . In: PROCEEDINGS OF THE 2ND INT. CONF. ON PRACTICAL ASPECTS OF KNOWLEDGE MANAGEMENT (PAKM98). Basel, Switzerland: 29 out. 1998

MACLEAN, M.; ANDERSON, J.; MARTIN, B. R. Identifying research priorities in public sector funding agencies: mapping science outputs on to user needs. **Technology Analysis & Strategic Management**, v. 10, n. 2, p. 139–155, jan. 1998.

MAGALHÃES, M. E. DE. **Uma ferramenta de apoio na identificação de novos elementos geográficos de baixa granularidade em notícias para a atualização de dicionários geográficos**. Dissertação (mestrado)—Rio de Janeiro: UFRJ, 2016.

MANNING, C.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. 2. ed. Massachusetts Institute of Technology: The MIT Press Cambridge, 1999.

MARCH, S. T.; SMITH, G. F. Design and natural science research on information technology. **Decision Support Systems**, v. 15, n. 4, p. 251–266, dez. 1995.

MARINELLI, E.; MILLER, R.; WARNKE, P. Future-oriented technology analysis: Practice in search of theory? **Futures**, v. 59, p. 1–4, jun. 2014.

MARTIN, B. R. The origins of the concept of 'foresight' in science and technology: An insider's perspective. **Technological Forecasting and Social Change**, v. 77, n. 9, p. 1438–1447, nov. 2010.

MARTINO, J. P. A review of selected recent advances in technological forecasting. **Technological Forecasting and Social Change**, v. 70, n. 8, p. 719–733, 2003.

MCCALLUM, A.; LI, W. **Early results for Named Entity Recognition with Conditional Random Fields, Feature Induction and Web-Enhanced Lexicons**. Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003. **Anais...**2003Disponível em: <<https://www.aclweb.org/anthology/W03-0430>>. Acesso em: 12 jun. 2020

MIKHEEV, A.; GROVER, C.; MOENS, M. **Description of the LTG System Used for MUC-7**. Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29 - May 1, 1998. **Anais...** In: MUC 1998. 1998Disponível em: <<https://www.aclweb.org/anthology/M98-1021>>. Acesso em: 9 jun. 2020

MINER, G. et al. (EDS.). **Practical text mining and statistical analysis for non-structured text data applications**. 1st ed ed. Waltham, MA: Academic Press, 2012.

MØLLER, M. F. A scaled conjugate gradient algorithm for fast supervised learning. **Neural Networks**, v. 6, n. 4, p. 525–533, jan. 1993.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre Aprendizado de Máquina. In: **Sistemas Inteligentes-Fundamentos e Aplicações**. 1. ed. Barueri-SP: Manole Ltda, 2003. p. 89–114.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. **Mineração de textos**. Goiás: Instituto de Informática da Universidade Federal de Goiás, 2007. Disponível em: <http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf>. Acesso em: 8 dez. 2019.

MOTA, C.; SANTOS, D. **Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM**. [s.l.] Linguateca, 2008.

MOTA, C.; SANTOS, D.; RANCHHOD, E. (EDS.). **Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa**. Lisboa: IST PRESS, 2007.

NADEAU, D.; SEKINE, S. A survey of named entity recognition and classification. **Lingvisticae Investigationes**, v. 30, n. 1, p. 3–26, 15 ago. 2007.

OBJECT MANAGEMENT GROUP. Business Process Model and Notation (BPMN), Version 2.0. p. 532, 2013.

OLIVEIRA, H. G.; MOTA, C.; FREITAS, C. Avaliação à medida no Segundo HAREM. In: **Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM**. [s.l.] Linguateca, 2008. p. 33.

PEFFERS, K. et al. Design Science Research Evaluation. In: PEFFERS, K.; ROTHENBERGER, M.; KUECHLER, B. (Eds.). **Design Science Research in Information Systems. Advances in Theory and Practice**. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. v. 7286p. 398–410.

PENG, F.; MCCALLUM, A. **Accurate Information Extraction from Research Papers using Conditional Random Fields**. Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004. **Anais...** In: HLT-NAACL 2004. Boston, Massachusetts, USA: Association for Computational Linguistics, 2 maio 2004Disponível em: <<https://www.aclweb.org/anthology/N04-1042>>. Acesso em: 12 jun. 2020

PHAAL, R.; FARRUKH, C. J. P.; PROBERT, D. R. Technology roadmapping—A planning framework for evolution and revolution. **Technological Forecasting and Social Change**, v. 71, n. 1–2, p. 5–26, jan. 2004.

PHAAL, R.; FARRUKH, C.; PROBERT, D. Roadmapping for strategy and innovation: Aligning technology and markets in a dynamic world. **UK: Cambridge University - Institute of Manufacturing**, p. 240, 2010.

POPPER, R. How are foresight methods selected? **Foresight**, v. 10, n. 6, p. 62–89, 15 out. 2008.

PORTER, A. L. et al. **Forecasting and Management of Technology**. [s.l.] John Wiley & Sons, 1991.

PORTER, A. L. et al. Technology futures analysis: Toward integration of the field and new methods. **Technological Forecasting and Social Change**, v. 71, n. 3, p. 287–303, mar. 2004.

RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. **Proceedings of the IEEE**, v. 77, n. 2, p. 257–286, fev. 1989.

RAJMAN, M.; BESANÇON, R. Text Mining: Natural Language techniques and Text Mining applications. In: SPACCAPIETRA, S.; MARYANSKI, F. (Eds.). **Data Mining and Reverse Engineering**. Boston, MA: Springer US, 1998. p. 50–64.

RATINOV, L.; ROTH, D. **Design challenges and misconceptions in named entity recognition**. Proceedings of the Thirteenth Conference on Computational Natural Language Learning. **Anais...** CoNLL '09. USA: Association for Computational Linguistics, 4 jun. 2009. Acesso em: 30 nov. 2020

RILEY, M. D. **Some applications of tree-based modelling to speech and language**. Proceedings of the workshop on Speech and Natural Language - HLT '89. **Anais...** In: THE WORKSHOP. Cape Cod, Massachusetts: Association for Computational Linguistics, 1989Disponível em: <<http://portal.acm.org/citation.cfm?doid=1075434.1075492>>. Acesso em: 28 maio. 2020

RILOFF, E.; JONES, R. **Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping**. In: PROCEEDINGS OF THE SIXTEENTH NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE. 1999

RINNE, M. Technology roadmaps: Infrastructure for innovation. **Technological Forecasting and Social Change**, v. 71, n. 1–2, p. 67–80, jan. 2004.

RIZZO, G.; TRONCY, R. NERD: Evaluating Named Entity Recognition Tools in the Web of Data. p. 16, 2011.

SANTOS, D. O modelo semântico usado no Primeiro HAREM. In: SANTOS, D.; CARDOSO, N. (Eds.). **Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área**. [s.l.] Linguateca, 2007. p. 43–57.

SARAWAGI, S. Information Extraction. **Foundation and Trends in Databases**, v. 1, n. 3, p. 261–377, 2007.

SATO, K.; SAKAKIBARA, Y. RNA secondary structural alignment with conditional random fields. **Bioinformatics**, v. 21, n. Suppl 2, p. ii237–ii242, 1 set. 2005.

SCAPOLO, F. New horizons and challenges for future-oriented technology analysis—The 2004 EU-US seminar. **Technological Forecasting and Social Change**, v. 72, n. 9, p. 1059–1063, nov. 2005.

SEIDEL, W. **Reconhecimento de entidades mencionadas em português utilizando aprendizado de máquina**. Dissertação de mestrado—São Paulo: Universidade de São Paulo, 2012.

SEKINE, S. **Description of the Japanese NE System Used for MET-2**. Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29 - May 1, 1998. **Anais...** In: MUC 1998. 1998Disponível em: <<https://www.aclweb.org/anthology/M98-1019>>. Acesso em: 12 jun. 2020

SETTLES, B. **Biomedical Named Entity Recognition using Conditional Random Fields and Rich Feature Sets**. Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications (NLPBA/BioNLP). **Anais...** Geneva, Switzerland: COLING, ago. 2004Disponível em: <<https://www.aclweb.org/anthology/W04-1221>>. Acesso em: 7 jun. 2020

SILVA, T. S. DA. **Reconhecimento de Entidades Nomeadas em Notícias de Governo**. Dissertação de Mestrado—Rio de Janeiro: UFRJ, 2012.

SILVA, E. F. A.; BARROS, F. A.; PRUDÊNCIO, R. B. C. Uma Abordagem de Aprendizagem Híbrida para Extração de Informação em Textos Semi-Estruturados. **Anais do XXV Congresso da Sociedade Brasileira de Computação**, p. 10, 2005.

SKUMANICH, M.; SILBERNAGEL, M. Foresighting around the world: a review of seven best-in-kind programs. **Seattle: Battelle**, 1997.

SUREKA, A. et al. Polarity Classification of Subjective Words Using Common-Sense Knowledge-Base. In: SAKAI, H. et al. (Eds.). . **Rough Sets, Fuzzy Sets, Data Mining and Granular Computing**. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. v. 5908p. 486–493.

SUTTON, C.; MCCALLUM, A. An Introduction to Conditional Random Fields for Relational Learning. In: GETOOR, L.; TASKAR, B. (Eds.). . **Introduction to Statistical Relational Learning**. [s.l.] The MIT Press, 2007.

TASKAR, B.; ABBEEL, P.; KOLLER, D. Discriminative Probabilistic Models for Relational Data. **arXiv:1301.0604 [cs, stat]**, 12 dez. 2012.

WALSH, S. T.; LINTON, J. D. Infrastructure for Emergent Industries Based on Discontinuous Innovations. **Engineering Management Journal**, v. 12, n. 2, p. 23–32, jun. 2000.

WEIKUM, G. et al. Database and information-retrieval methods for knowledge discovery. **Communications of the ACM**, v. 52, n. 4, p. 56, 1 abr. 2009.

WEISS, S. M. et al. **Text Mining: Predictive Methods for Analyzing Unstructured Information**. [s.l.] Springer Science & Business Media, 2010.

WILLYARD, C. H.; MCCLEES, C. W. Motorola's Technology Roadmap Process. **Research Management**, v. 30, n. 5, p. 13–19, set. 1987.

WIVES, L. K. **Tecnologias de descoberta de conhecimento em textos aplicadas à inteligência competitiva**PPGC-UFRGS, , 2002. . Acesso em: 8 dez. 2019

XUMING HE; ZEMEL, R. S.; CARREIRA-PERPINAN, M. A. **Multiscale conditional random fields for image labeling**. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. **Anais...** In: PROCEEDINGS OF THE 2004 IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2004. CVPR 2004. Washington, DC, USA: IEEE, 2004Disponível em: <<http://ieeexplore.ieee.org/document/1315232/>>. Acesso em: 12 jun. 2020

ZAMBENEDETTI, C. **Extração de Informação sobre Bases de Dados Textuais**. Porto Alegre: UFRGS, 2002.

ZHANG, Y.; TSAI, F. S. **Combining named entities and tags for novel sentence detection**. Proceedings of the WSDM '09 Workshop on Exploiting Semantic Annotations in Information Retrieval - ESAIR '09. **Anais...** In: THE WSDM '09 WORKSHOP. Barcelona, Spain: ACM Press, 2009Disponível em: <<http://portal.acm.org/citation.cfm?doid=1506250.1506256>>. Acesso em: 30 nov. 2020

Anexos

1. Experimento: Prospecções do *Journal Futures* (2020)

ID	Prospecção	NERMAP Ciclo 1	NERMAP Ciclo 2	NERMAP Ciclo 3
1	In 2025 this track of land housed an experimental atom modifier, which during its explosion caused a chemical spill in a nearby plant. these two haphazard, hazardous events have chemically altered my body forever.	Encontrado	Encontrado	Encontrado
2	The 2030 agenda emphasises the SDGs are interlinked, and recognises the importance of ensuring integration across all the sectors they represent, without indicating ways in which this might be achieved.	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino	Encontrado
3	However because of the inherent complexity of the inter-related dilemmas involved, attaining improvement toward one SDG by 2030 may come at the expense of another. Not knowing where to prioritise action is a consequence of the overwhelming scale of issues that need to be addressed. Furthermore, there are dangers if national planning strategies focus on a goal-by-goal and target by target approach, neglecting the mutual influences between them	Não encontrado: Falha	Encontrado	Encontrado
4	Whilst pragmatically focusing on the short term through to 2030, the lack of long-term planning has been highlighted by the absence of global population goals and boundaries in those targets considering environmental constraints, despite well-known connections between population growth and environmental degradation	Não encontrado: Falha	Encontrado	Encontrado
5	To deliver all the SDGs by 2030 will require a concerted global effort, in which both time and monetary constraints must be overcome. Recent reports have highlighted the importance of the inter-relationships and dependancies between the various goals with Stafford-Smith et al., 2016	Encontrado	Encontrado	Encontrado

	noting: “there must be greater attention on inter-linkages in three areas: across sectors, across societal actors, and between and among low, medium and high-income countries”.			
6	Using a quantitative global simulation model called Earth have considered socio-economic and biophysical development through to 2050, and using one modelable indicator for each goal have assessed the number of SDGs likely to be achieved by 2030 and 2050.	Encontrado	Encontrado	Encontrado
7	In a business as usual scenario their results show human societies becoming richer (as measured by GDP per person) but people live in more unequal societies and in an environment increasingly damaged by human activity. Accelerating economic growth leads to higher GDP, more energy use, more CO2 emissions and more use of resources with 11 SDGs achieved by 2030 but with greater global pressures on planetary boundaries.	Encontrado	Encontrado	Encontrado
8	In a third scenario which sees a stronger policy focus on achieving the SDGs leads to more (but not all) goals being attained by 2030 and a halt in the downward trend of environmental goals. This leads to their conclusion that non-conventional - even transformational - action is required to create a sustainable world and satisfactory levels of well being for all.	Encontrado	Encontrado	Encontrado
9	The 2030 target is largely political, set for consistency to match the 15-year timeframe of the Millennium Development Goals. However, in light of the incomplete completion of the Millennium Development Goals, which only consisted of 17 targets across eight goals, it is unreasonable to think that in the same timeframe the Sustainable Development Goals will be achieved, which consist of 169 targets across 17 goals. Progress towards the completion of the Sustainable Development Goals is not linear, with many factors and causal connections within the system that will stagger the progress and thus the final achievement of the goals.	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino	Encontrado
10	However, even by prioritising the foundational goals and feedback loops, it is unlikely that the 2030 target will be met.	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino
11	The vaccines market is estimated to rise from the present value of \$38 billion to over \$70 billion by 2025.	Encontrado	Encontrado	Encontrado

12	A 2008 study found that southern Africa could lose 30 % of its maize crop by 2030 due to the negative effects of climate change. Losses of maize and rice crops in South Asia could also be significant	Encontrado	Encontrado	Encontrado
13	Without further commitments and action to reduce greenhouse gas emissions, the world is likely to warm by more than 3 °C above the preindustrial climate. Even with the current mitigation commitments and pledges fully implemented, there is roughly a 20 percent likelihood of exceeding 4 °C by 2100. If they are not met, a warming of 4 °C could occur as early as the 2060s.	Encontrado	Encontrado	Encontrado
14	Such a warming level and associated sea-level rise of 0.5–1 meter, or more, by 2100 would not be the end point: a further warming to levels over 6 °C, with several meters of sea-level rise, would likely occur over the following centuries.	Não encontrado: Falha	Encontrado	Encontrado
15	By 2050 , under a typical middle-of-the-road emissions scenario, you're looking at a doubling of the volatility for grains in the midlatitudes. In places like China, the U.S., Europe, Ukraine—the breadbasket countries of the world—the volatility from year-to-year just from natural climate variability at a higher temperature is going to be much higher. The impact on crops is going to be greater and greater.	Encontrado	Encontrado	Encontrado
16	The most widely accepted view of population growth is the “demographic transition.” If incomes continue to rise in most countries and richer people have fewer children, then world population should peak at 9–11 billion around the year 2100.	Não encontrado: Falha	Encontrado	Encontrado
17	Over the next several thousand years, this dependence led to agriculture and large-scale state Societies.	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino
18	Climate models indicate that the Earth could warm by 3°C-4 °C by the year 2100 and eventually by as much as 8 °C or more.	Encontrado	Encontrado	Encontrado
19	It took only a few thousand years after sedentary agriculture began for it to spread and become dominant in the Middle East, South Asia, China, and Mesoamerica.	Não encontrado: Caso não	Não encontrado: Caso não	Não encontrado: Caso não

		pertence ao modelo de treino	pertence ao modelo de treino	pertence ao modelo de treino
20	In just the past 70 years human activity has increased atmospheric CO ₂ levels by 100 ppm to over 400 ppm, and the Earth's average temperature has warmed by 1 °C. Unless draconian measures are taken to halt the increase in atmospheric CO ₂ , global temperature will likely increase by at least 3 °C c today's by the year 2100 and could eventually increase by 8 °C or more (the so-called mega-greenhouse).	Encontrado	Encontrado	Encontrado
21	After the initial establishment of agriculture there was a period of several thousand years of small, settled communities "stateless" societies that practiced a combination of agriculture and foraging.	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino
22	Most projections of global warming focus on either the year 2100 or the effects of a doubling of CO ₂ (from the pre-industrial level of 275 ppm–550 ppm).	Encontrado	Encontrado	Encontrado
23	The lack of attention to the very long run is a serious shortcoming, since integrated carbon-climate models project that if CO ₂ from current in situ fossil fuel resources continues to be released into the atmosphere, the peak concentration of atmospheric CO ₂ could exceed 1400 ppm by the year 2300 and the average global temperature could warm by 8 °C or more.	Não encontrado: Fora do escopo do ano limite do TRM	Encontrado	Encontrado
24	In Europe, North America and Australia, 2.4 million people are predicted to die from infections with resistant microorganisms in the next 30 years and could cost up to US\$3.5 billion per year.	Não encontrado: Falha	Não encontrado: Falha	Não encontrado: Falha
25	A second example, the Association of Research Libraries (ARL US) 2030 scenarios , is represented by the matrix: (a) constraints on individual researchers, and (b) research enterprise aggregation. (Fig. D2). Constraints on researchers and level of research enterprise aggregation are, similarly, not variables ARL or its members may expect to significantly influence. The interviewee (see Table B1) stressed that the purpose was to create scenarios that were different	Não encontrado: Caso não pertence ao modelo de	Não encontrado: Caso não pertence ao	Encontrado

	to each other, without consideration to which of them might be preferred, in order to “think about different ways things could turn out,” that is to inform adaptation to future eventualities beyond the ARL’s control.	treino e arquivo apresentou erro na conversão	modelo de treino	
26	Considering examples of successes of adaptive type, the Association of Research Libraries 2030 Scenarios , structured by exogenous axes, were used by librarians in perceiving and preparing for potential new industry externalities beyond ARL’s control, suggesting ways its systems and offerings may need adapting to remain current and competitive. Similarly, Bord Bia’s scenarios were used in follow-on workshops to stimulate Irish food and agriculture firms prepare themselves to operate in new food-chain and food-consumption environments the emergence of which they had very little control over. In such cases, the scenario projects were also reported to have been relevant in furthering the aims of their commissioning organizations. It is apparent that in these cases the scenario project purpose was exploratory and adaptive (“how do we shape ourselves to survive and thrive in view of macro-externalities we don’t much control?”), and the scenario matrix purpose in each case was congruent to this.	Não encontrado: Caso não pertence ao modelo de treino e arquivo apresentou erro na conversão	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino
27	The prediction market Metaculus uses trades with in-platform credits allowing individuals to perform actions such as posing their own questions. It has set up a market to establish the probability of human extinction (source 10), although the market clearing price, which will represent its ‘final’ prediction, will not be available until it closes in 2030.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
28	For instance, one might reason that evidence about the adaptability of humans to environmental changes suggests a 30 % probability that we will survive the next 1000 years, but only a 10 % probability that we will not. ⁴ Combination functions can then be used to aggregate these imprecise probabilities to return the overall probability of extinction within this period.	Não encontrado: Fora do escopo do ano limite do TRM e arquivo apresentou	Não encontrado: Fora do escopo do ano limite do TRM	Não encontrado: Fora do escopo do ano limite do TRM

		erro na conversão		
29	By 2020-2030, annual growth rates are likely to increase and may exceed 10%-15%.	Encontrado	Encontrado	Encontrado
30	In recent times, the studies that cover the Russian oil and gas sector have been mainly dominated by institutional and macroeconomic perspectives. Paltsev (2014), for example, approaches the economic opportunities of Russia's natural gas exports through possible scenarios until 2050, including threats to Russian exports, such as the emerging shale gas technologies development across the globe, the EU energy policy (nuclear, non-nuclear, renewable), and the transformation of liquefied natural gas markets.	Encontrado	Encontrado	Encontrado
31	A sustainable development agenda for WIPO would surely encompass a broader understanding of development. The establishment of the UN's SDGs in 2015 as part of the 2030 Agenda for Sustainable Development is, in some respects, an effort, like the prior MDGs, to broaden the discourse and mandate of development beyond purely economic measures. Only three of the seventeen SDGs deal directly with economic or industrial development; others focus on education (4), gender equality (5), inequality (11), and peace and justice (16), among others.	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino
32	Given the rate and scale of economic change required to minimise climate risks (i.e. net-zero emissions by 2050 or sooner), in navigating the terrain ahead we should expect that knowledge systems and practices established to deal with past and even current change processes will at best provide partial guidance, and at worst be misguiding.	Não encontrado: Falha	Encontrado	Encontrado
33	It is estimated that artificial intelligence revolution will come into full force within the next twenty years.	Não encontrado: Caso não pertence ao modelo de treino e arquivo apresentou erro na conversão	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino

34	A study by Kallio, Salminen, and Sievänen (2013) created three scenarios for the year 2025 to compare the effects of wood energy on the carbon balance.	Encontrado	Encontrado	Encontrado
35	SDG 7 focuses on clean and affordable energy and aims for a substantial increase in the share of renewable energy in the global energy mix by 2030, whereas the sub-target of SDG13 addresses the national policies for climate change mitigation.	Encontrado	Encontrado	Encontrado
36	For example, Tapio, Rintamäki, Rikkonen, and Ruotsalainen (2017) created six energy scenarios for farms up to the year 2030. They showed that the role of wood burning will likely remain prominent in the future, whereas the role of other renewable energy sources varies in different scenarios.	Encontrado	Encontrado	Encontrado
37	Based on the National Energy and Climate Strategy for 2030 government report, the use of woodchips should increase significantly in the future due to increased use for heat and electricity production, as well as fuel production for transport (Huttunen, 2017).	Encontrado	Encontrado	Encontrado
38	The report estimates that in 2030, the use will be around 14–18 million m ³ (Huttunen, 2017). This would mean at least a 100% increase compared to the current level.	Não encontrado: Falha	Encontrado	Encontrado
39	However, the energy demand grows as well. Based on the existing energy policies and announced policy intentions, the amount of global energy consumption is projected to increase almost 80 % between 2016 and 2040.	Não encontrado: Falha	Encontrado	Encontrado
40	The concept of futures image has seldom been applied in energy research. We identified four studies published in academic journals that had constructed futures images on either directly or partly related to future energy use. Wallgren and Höjer (2009) studied energy use in the food system. Their end result was a futures image of sustainable food consumption in 2050, wherein energy efficiency was the key component.	Encontrado	Encontrado	Encontrado
41	a profound mind shift toward universal values that re-connect world development with a resilient Earth, recognise the right of all to development and promote a shift from materialistic lifestyles to the pursuit of well-being and fulfilment.' (It would also involve) three critical transformations ... decarbonising the world economy by 2050 to 2070, feeding the world through sustainable agriculture by 2050 and accelerating progress toward an economy of cyclic material flows.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado

42	This point was made amply clear in a 2018 workshop at NordiCHI, a biennial conference that functions as the main Nordic forum for human-computer interaction research. A multidisciplinary group of researchers, academics, philosophers and ethicists explored the role of AI and Human Computer Interaction (HCI) in the future of wisdom during the coming decades. ‘What will be the long-term consequences of HCI, AI, IoT, Big Data and Smart Technologies 50 years from now	Não encontrado: Falha	Encontrado	Encontrado
43	Brazil emits about 2.5 % of the world’s carbon dioxide and other polluting gases; moreover, Brazil’s overall emissions increased steeply from 2010 (CDIAC, 2017). In September 2016, the country ratified the Paris Agreement (Agriocone, 2016), committing to reduce emissions by 37 % by 2025 and by 43 % by 2030, based on 2005 reference values.	Não encontrado: Precisa de tratamento para o caso	Encontrado	Encontrado
44	The Brazilian population annual growth is about 0.8 %, slowing towards 0.5 % in 2030 and stabilizing at 240 million people around 2050.	Encontrado	Encontrado	Encontrado
45	Brazil is expected to become a major exporter of oil and to be the world’s 6th largest energy provider by 2035 (IEA, 2014).	Não encontrado: Precisa de tratamento para o caso	Encontrado	Encontrado
46	In view of the interviewees’ perceptions and previous analysis, “policies and regulations” can be considered as an uncertainty that could heavily affect the electricity market in general, and solar in particular, within the next 5 years.	Não encontrado: Falha	Não encontrado: Falha	Não encontrado: Falha
47	I would bet you anything that that the solar energy supply doesn’t reach 10 Giga[Watts] in Brazil in the next 10 years as the government claims. Money in Brazil is very costly, so if you take out a loan to purchase your PV system, you will never pay back that loan with the savings on your energy bill.	Não encontrado: Falha	Não encontrado: Falha	Não encontrado: Falha
48	Short-term decisions are in conflict with long-term ones. This happens not only in Brazil, but everywhere. A government generally thinks only about the next four to five years, a situation that results in sub-optimal decisions for the next 20, 30, or 50 years.	Não encontrado: Caso não pertence ao	Não encontrado: Caso não pertence ao	Não encontrado: Caso não pertence ao

		modelo de treino	modelo de treino	modelo de treino
49	Technological advances, particularly in computerization, encourage firms to robotize, replacing unskilled jobs that can be easily routinized by machines. It is a common view that new disruptive technologies will burn more jobs than those the same technology will be able to generate in several industries (Kotler, 2015). According to Frey and Osborne (2017), 47 % of total US employment is in the high-risk category, being these jobs expected to be automated relatively soon, perhaps by 2023 or 2033.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
50	Population by region, sex, age in 2050 is calculated based on population projection by the National Institute of Population and Social Security Research.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
51	Future scenarios of passenger travel demand within daily life area in Japan in 2050 were calculated by a combination of two levels of change of the corresponding term of Mega-trends	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
52	Passenger travel by non-motorized, motorized private, and motorized public transportation within daily life area in 2050 are 40 billion person-km/yr, 471 billion person-km/yr, and 377 billion person-km/yr in Population Change Only, 44 billion person-km/yr, 494 billion person-km/yr, and 411 billion person-km/yr in Socioeconomic Change Only, 88 billion person-km/yr, 356 billion person-km/yr, and 385 billion person-km/yr in Low Demand, and 29 billion person-km/yr, 852 billion person-km/yr, and 266 billion person-km/yr in High Accessibility, respectively.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
53	Travel demand for commuting, work, and private trips in 2050 are 407 billion person-km/yr, 78 billion person-km/yr, and 402 billion person-km/yr in Population Change Only, 462 billion person-km/yr, 78 billion person-km/yr, and 409 billion person-km/yr in Socioeconomic Change Only, 373 billion person-km/yr, 79 billion personkm/yr, and 377 billion person-km/yr in Low	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado

	Demand, and 545 billion person-km/yr, 98 billion person-km/yr, and 503 billion personkm/yr in High Accessibility, respectively.			
54	Average distance with respect to average travel time per capita within daily life by mode of transportation in 2015–16, and the Socioeconomic Change Only, Low Demand, and High Accessibility scenarios in 2050.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
55	Average per-capita travel distance and travel time by motorized public transportation by representative is in the range of 1–32 km/day and 1–66 min/day in 2015–16, 1–32 km/day and 1–65 min/day in Socioeconomic Change Only, 0–25 km/day and 0–52 min/day in Low Demand, and 0–22 km/day and 0–42 min/day in High Accessibility in 2050, respectively.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
56	Average per-capita travel distance and travel time by motorized private transportation by representative is in the range of 4–46 km/day and 10–72 min/day in 2015–16, 5–42 km/day and 11–65 min/day in Socioeconomic Change Only, 3–36 km/day and 8–55 min/day in Low Demand, and 9–59 km/day and 17–103 min/day in High Accessibility in 2050, respectively.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
57	Average per-capita travel distance and travel time by NMT by representative is in the range of 0–2 km/day and 3–23 min/day in 2015–16, 0–2 km/day and 4–23 min/day in Socioeconomic Change Only, 1–4 km/day and 10–44 min/day in Low Demand, and 0–2 km/day and 2–14 min/day in High Accessibility in 2050, respectively.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
58	In Population Change Only, travel demand by motorized private transportation decreases by 28 % because the rural population is projected to decrease selectively in 2050.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
59	In 2050, average distance and travel time within daily life in the Socioeconomic Change Only scenario shows little change except for women and those aged 65 and over compared to those in 2015–16.	Não encontrado: Precisa de	Encontrado	Encontrado

		tratamento para o caso e arquivo apresentou erro na conversão		
60	Contrasting future passenger travel demand scenarios in Japan in 2050, that is, Low Demand and High Accessibility, were developed in accordance with qualitative scenarios and quantitative assumptions of four mega-trends: population aging, gender equality, tertiary industrialization, and transport technology innovation.	Não encontrado: Erro na conversão do arquivo	Encontrado	Encontrado
61	In the Low Demand scenario, total travel demand by motorized private transportation is found to decrease by 45 % in 2050 compared to that in 2015–16.	Não encontrado: Precisa de tratamento para o caso e arquivo apresentou erro na conversão	Encontrado	Encontrado
62	Automation has generated expectations regarding increased productivity, efficiency, safety, and convenience for employers, but it also threatens the current job positions, salaries, skills, and talents of the workforce and, above all, the nature of the work. From the point of view of automation experts, at least 35 % of the jobs currently available will be automated over the next twenty years, and a study conducted by the Mckinsey Global Institute predicted that by 2025, robotics may replace up to 75 million jobs worldwide. Even in the case of partial automation within a particular job, automation could impact all hierarchical levels from gardeners to managers.	Não encontrado: Falha	Encontrado	Encontrado

63	The growth in the use of these platforms is rapid, and it is estimated that by 2025, nearly 540 million Internet users could use these platforms to find a job.	Não encontrado: Falha	Encontrado	Encontrado
64	According to the OECD projections, single-parent families will increase in absolute terms and proportion with respect to households with children. In fact, by 2030, it is expected that the rate of single-parent households will increase in all OECD countries.	Encontrado	Encontrado	Encontrado
65	The uncertainty of this axis focuses on the issue of whether by 2040 in Mexico the contemporary traditional family will have disappeared and been replaced by more radical and complex forms of family agreements in households.	Encontrado	Encontrado	Encontrado
66	This issue is considered one of the greatest uncertainties because while there is a clear trend towards entrepreneurship economics and automation worldwide, we do not know whether the economic conditions, infrastructure and social acceptance are a breeding ground or barrier to the proliferation of this type of economy in Mexico in 2040	Encontrado	Encontrado	Encontrado
67	In 2040, in Mexico, there are no working days, and everything is assigned by projects and resolved in the time that was committed for delivery.	Encontrado	Encontrado	Encontrado
68	In recent years, there has been an increase in life expectancy due to science and technology. The OECD predicts that life expectancy at birth will continue to increase until 80 and 90 years of age at the global level by 2050.	Encontrado	Encontrado	Encontrado
69	For example, discussing responsibility in management education, Henderson, Wersun, Wilson, Mo-ching Yeung, and Zhang (2019) equate responsible management with management for sustainability, in a context in which 'Humanity's very existence in 2068 may be considered to be uncertain'.	Não encontrado: Falha	Encontrado	Encontrado
70	The Good Anthropocene scenarios exhibit similarities with both of those narratives, though the GEO-6 scenarios are much more constrained by current trends and possibilities. Similarly, the vision put forth by Agenda 2063 for Africa (African Union Commission, 2015) is based on extrapolations of past and existing continental initiatives for sustainable growth and development, without significantly pushing the boundaries of what is possible in a truly uncertain and exponentially changing Anthropocene.	Não encontrado: Caso não pertence ao modelo de treino e arquivo apresentou	Não encontrado: Caso não pertence ao modelo de treino	Não encontrado: Caso não pertence ao modelo de treino

		erro na conversão		
71	The 2063 vision is focused on strengthening Africa's role in a globalizing world, and aims to achieve the SDGs mainly through standard economic growth.	Não encontrado: Caso não pertence ao modelo de treino e arquivo apresentou erro na conversão	Não encontrado: Caso não pertence ao modelo de treino	Encontrado
72	The 2063 vision is heavily embedded within conventional technological trends and does not emphasise the rich bio-cultural diversity of the continent.	Não encontrado: Caso não pertence ao modelo de treino e arquivo apresentou erro na conversão	Não encontrado: Caso não pertence ao modelo de treino	Encontrado
73	NASA plans have included sending astronauts to Mars by 2035, and China also intends to build a lunar colony followed by a Mars colony. These plans are enabled by technological innovation. New materials allow rockets to be lighter, stronger, and cheaper to launch, as does the profit-driven design philosophy of Musk and Bezos.	Encontrado	Encontrado	Encontrado
74	The worldbuilding and artefact development processes enabled participants to develop detailed pathways to 2038 and create an artefact which would bring to life an aspect of their envisioned future.	Não encontrado: Caso não pertence ao	Não encontrado: Caso não pertence ao	Encontrado

		modelo de treino	modelo de treino	
75	The dominance of Transform archetypes continues, as participants envisioned six Transform and two Collapse future worlds. The role of technology in enabling transformative change is a key precondition across all future pathways. Three of the future worlds developed narrative arcs centred around a global state of crisis along the pathway to 2038.	Encontrado	Encontrado	Encontrado
76	Perhaps nowhere as in the specific case of mathematical modelling one has elements to detect an excessive use of the technique, something which might be called a quantification hubris. I have already mentioned the computation of the cost of climate action/ inaction at the year 2100, which pales when compared to what the authors in (Pilkey & Pilkey-Jarvis, 2009) call “A Million Years of Certainty” relative to a model-based computation of the risk of a nuclear waste disposal one million years into the future.	Não encontrado: Precisa de tratamento para o caso e arquivo apresentou erro na conversão	Encontrado	Encontrado

2. Ciclo 1: Prospecções obtidas pelo NERMAP (*Journal Futures*)

ID	Data	Prospecção	Tipo de Prospecção	Encontrada manualmente
1	2021	Here are two examples of the narratives that emerged: Statement 1: Over the next years more people from UK and abroad would visit the centre of Chester ... For our sustainability, it is essential that long-term sustainable income streams are identified.	Completa	Não
2	2025	According to the vaccines market is estimated to rise from the present value of \$ 38 billion to over \$ 70 billion by 2025.	Completa	Sim
3	2025	The Beyond Current Horizons programme was a particularly apt framework to work within because it set out to explore: ... what society might look like in 2025 in order to anticipate the demands that will be placed on the UK education system, taking as a focus not the future in its entirety, but specifically the intersection between technological, educational and social futures.	Parcial	Não
4	2025	Vaccines market worth over \$ 70bn by 2025.	Completa	Não
5	2025	In 2025 this track of land housed an experimental atom modifier, which during its explosion caused a chemical spill in a nearby plant.	Completa	Sim
6	2025	A study by Kallio, Salminen, and Sievinen created three scenarios for the year 2025 to compare the effects of wood energy on the carbon balance.	Parcial	Sim
7	2025	Wind power in Finland up to the year 2025 -- soft scenarios based on expert views.	Parcial	Não
8	2025	China manufacturing locus in 2025: With a comparison of 'Made - in-China 2025 and 'Industry 4.0.	Parcial	Não
9	2030	A 2008 study found that southern Africa could lose 30% of its maize crop by 2030 due to the negative effects of climate change.	Completa	Sim
10	2030	Accelerating economic growth leads to higher GDP, more energy use, more CO2 emissions and more use of resources with 11 SDGs achieved by 2030 but with greater global pressures on planetary boundaries.	Completa	Sim
11	2030	Even if technological capacities for longer human missions are developed in next decades, the social and political context on Earth may become less conducive to space missions because of climate change, population growth in some parts of the globe, population decline in other parts,	Completa	Não

		and the world limited ability to distribute human resources equitably, leaving enough to ?nance manned space missions.		
12	2030	In a third scenario which sees a stronger policy focus on achieving the SDGs leads to more goals being attained by 2030 and a halt in the downward trend of environmental goals.	Completa	Sim
13	2030	Prioritizing climate change adaptation needs for food security in 2030.	Completa	Sim
14	2030	To deliver all the SDGs by 2030 will require a concerted global e ?ort, in which both time and monetary constraints must be overcome.	Completa	Sim
15	2030	The 17 goals lay out a clear plan of where humanity wants to go by 2030.	Parcial	Não
16	2030	By 2020 - 2030, annual growth rates are likely to increase and may exceed 10% -15 %.	Completa	Sim
17	2030	For example, Tapiro, Rintam?ki, Rikkonen, and Ruotsalainen created six energy scenarios for farms up to the year 2030.	Parcial	Sim
18	2030	Future images of meat consumption in 2030.	Parcial	Não
19	2030	Government report on the national energy and climate strategy for 2030.	Parcial	Não
20	2030	In fact, by 2030, it is expected that the rate of single-parent households will increase in all OECD countries.	Completa	Sim
21	2030	In terms of implications for their future careers, their re ?ections spanned a broad spectrum -- while for some it further con ?rmed the belief that the professional choice they are making is right , some participants re ?ected on the fact that their desired jobs as understood today might not exist at all in the next decades.	Completa	Não
22	2030	Introduction In this paper, we discuss using developed scenarios as a strategic tool to engage various stakeholders located in di ?erent countries as part of a larger survey for an international research project to gain a broad understanding of transport experts opinions regarding transport options and mobility-related commodity prices in the Asia-Paci ?c region in 2030.	Inválida	Não
23	2030	Lake management in 2030 -- Five future images based on an international Delphi study.	Parcial	Não
24	2030	Projected prevalence of car-sharing in four asian-paci ?c countries in 2030: What the experts think.	Completa	Não
25	2030	SDG 7 focuses on clean and a ?ordable energy and aims for a substantial increase in the share of renewable energy in the global energy mix by 2030, whereas the sub-target of SDG13 addresses the national policies for climate change mitigation.	Completa	Sim
26	2030	The future of families to 2030.	Parcial	Não

27	2030	Predicting the future of additive manufacturing: A Delphi study on economic and societal implications of 3D printing for 2030.	Inválida	Não
28	2035	NASA plans have included sending astronauts to Mars by 2035, and China also intends to build a lunar colony followed by a Mars colony.	Completa	Sim
29	2038	Three of the future worlds developed narrative arcs centred around a global state of crisis along the pathway to 2038.	Completa	Sim
30	2040	Contents lists available at ScienceDirect Futures journal homepage: www.elsevier.com/locate/futures Work/family life by 2040: Between a gig economy and traditional roles Guillermina Benavides Rinc?n *, Yolanda Montes Mart?nez Tecnologico de Monterrey, Mexico ARTICLE INFO Keywords: Scenarios Gender Work/life balance Work ?exibility ABSTRACT This research envisions four future scenarios of work/life balance for Mexican families by 2040.	Inválida	Não
31	2040	From this perspective, the use of a scenario approach is relevant for analyzing plausible futures to generate balance in work and family life by 2040.	Parcial	Não
32	2040	Future family and work/life balance in 2040.	Parcial	Não
33	2040	In 2040 oil & gas will still occupy 50% of the energy matrix.	Completa	Não
34	2040	In 2040, in Mexico, 50% of jobs are independent entrepreneurs who take on contracts with different companies simultaneously in such a way that they compromise the delivery times of each project but do not have a committed schedule with any employer.	Completa	Não
35	2040	In 2040, in Mexico, there are no working days, and everything is assigned by projects and resolved in the time that was committed for delivery.	Completa	Sim
36	2040	The following process was followed for the construction of the scenarios of the work/life balance by 2040: 1 A literature review was conducted to identify and categorize the main driving forces affecting the phenomenon under study.	Parcial	Não
37	2040	The uncertainty of this axis focuses on the issue of whether by 2040 in Mexico the contemporary traditional family will have disappeared and been replaced by more radical and complex forms of family agreements in households.	Completa	Sim
38	2040	This issue is considered one of the greatest uncertainties because while there is a clear trend towards entrepreneurship economics and automation worldwide, we do not know whether the economic	Completa	Sim

		conditions, infrastructure and social acceptance are a breeding ground or barrier to the proliferation of this type of economy in Mexico in 2040 .		
39	2040	This research considered the year 2040, which is slightly over 20 years in the future and when the children of the couples interviewed will face this duality of roles, i.e., having to reconcile work and family life, as a reference.	Parcial	Não
40	2050	According to Batissti : By 2050, under a typical middle-of-the-road emissions scenario, you looking at a doubling of the volatility for grains in the mid - latitudes.	Parcial	Sim
41	2050	Interview of Jared Diamond by David Wallace-Wells, `` Jared Diamond: There a 49 percent chance the world as we know it will end by 2050.	Parcial	Não
42	2050	Using a quantitative global simulation model called Earth 3 Randers, Rockstrom, Stoknes, Goluke, Collste, Cornell et al. have considered socio-economic and biophysical development through to 2050, and using one modelable indicator for each goal have assessed the number of SDGs likely to be achieved by 2030 and 2050.	Completa	Sim
43	2050	A public opinion survey of four future scenarios for Australia in 2050.	Parcial	Não
44	2050	Paltsev , for example, approaches the economic opportunities of Russia natural gas exports through possible scenarios until 2050, including threats to Russian exports, such as the emerging shale gas technologies development across the globe, the EU energy policy , and the transformation of lique ?ed natural gas markets.	Completa	Sim
45	2050	Period USAJapan Europe 2002 0.120.69 0.23 2003 0.130.30 0.20 2004 0.080.25 0.26 2005 0.290.32 0.33 2006 0.200.29 0.33 2007 0.490.31 0.39 2008 0.510.35 0.31 2009 0.570.35 0.36 2010 0.570.34 0.36 2020 0.670.36 0.39 2030 0.790.38 0.42 2040 0.930.40 0.45 2050 1.100.42 0.49 L. Aldieri, et al..	Inválida	Não
46	2050	Results Tables 2 and 3 show the results of the impact of R&D innovation on productivity over the period 2002 -- 2010 and the forecast value for 2020, 2030, 2040 and 2050.	Inválida	Não
47	2050	Scenarios for Russia natural gas exports to 2050.	Parcial	Não
48	2050	The Brazilian population annual growth is about 0.8 %, slowing towards 0.5% in 2030 and stabilizing at 240 million people around 2050.	Completa	Sim
49	2050	The OECD predicts that life ex - pectancy at birth will continue to increase until 80 and 90 years of age at the global level by 2050.	Completa	Sim
50	2050	The world in 2050 and the New Welfare scenario.	Parcial	Não

51	2050	Their end result was a futures image of sustainable food consumption in 2050, wherein energy efficiency was the key component.	Completa	Sim
52	2050	India: A Dharma based Country in 2050.	Inválida	Não
53	2052	2052: A global forecast for the next forty years.	Parcial	Não
54	2068	For this publication about a wiser future in 2068, I look for the babyhood, the infant, in the contributions that I received and explore what they might sound like if the conscious assumptions become that there is enough for everyone and that humans are driven to joy, and peace.	Inválida	Não
55	2068	Futures 115 102477 9 The Future in 2068 - An Alien Short StoryRafael Lourenco Prologue My father believed in aliens.	Inválida	Não
56	2068	Storytelling 2068 As part of the first survey process, the participants were asked to write a story about the world as they imagine it in 2068 and describe the role they would have played in that world.	Inválida	Não
57	2068	2038 Worldbuilding and artefact development The participants worked in teams and were facilitated through a design process to develop an artefact which would be of significance in 2038, halfway along the pathways towards 2068.	Inválida	Não
58	2068	On roles and agency in 2068 Alongside stories from 2068, the participants were also asked to describe their roles in those possible futures.	Inválida	Não
59	2068	Envisioning 2068 The next exercise took participants further into the future to 2068.	Inválida	Não
60	2068	Principles for responsible management education in 2068.	Inválida	Não
61	2068	The Summer Schools started with a series of future - casting exercises, followed by the development of a suite of detailed 2038 worlds as key markers on the pathways to 2068, as well as artefacts consistent with the logic of their speculative futures.	Inválida	Não
62	2068	There is a shared sense that technology will be fully embedded in our daily lives in 2068, that quality of life has improved lifestyles globally - from solving traffic problems, to rapid development of technology for good across healthcare, food supplies, energy production, as well as pollution management and ecosystem restoration.	Completa	Não
63	2068	They were asked to write a short story about the world as they imagine it in 2068.	Inválida	Não
64	2068	This issue focused on a theme of responsible futures, asking contributors to answer the question: what futures should humanity strive for? What might be considered responsible and wise in 2068, and why? 1.	Inválida	Não

65	2100	As Clark et al. put it: `` An evaluation of climate change risks that only considers the next 85 years of climate change impacts fails to provide essential information to stakeholders, the public and the political leaders who will ultimately be tasked with making decisions about policies on behalf of all, with impacts that will last for millennia.	Parcial	Não
66	2100	Climate models indicate that the Earth could warm by 3 °C -4 °C by the year 2100 and eventually by as much as 8 °C or more.	Completa	Sim
67	2100	Even with the current mitigation commitments and pledges fully implemented, there is roughly a 20 percent likelihood of exceeding 4 °C by 2100.	Completa	Sim
68	2100	Most projections of global warming focus on either the year 2100 or the effects of a doubling of CO ₂ .	Completa	Sim
69	2100	Unless draconian measures are taken to halt the increase in atmospheric CO ₂ , global temperature will likely increase by at least 3 °C above today by the year 2100 and could eventually increase by 8 °C or more.	Completa	Sim
70	2100	I wo be alive in 2100, but my grandchildren will ... so we are talking about people that are sharing life with us.	Inválida	Não

3. Ciclo 2: Prospecções obtidas pelo NERMAP (*Journal Futures*)

ID	Data	Prospecção	Tipo de Prospecção	Encontrada manualmente	Contida no Ciclo 1
1	2021	Here are two examples of the narratives that emerged: Statement 1: Over the next years more people from UK and abroad would visit the centre of Chester ... For our sustainability, it is essential that long-term sustainable income streams are identified.	Completa	Não	Sim
2	2025	In 2025 this track of land housed an experimental atom modifier, which during its explosion caused a chemical spill in a nearby plant.	Completa	Sim	Sim
3	2025	According to the vaccines market is estimated to rise from the present value of \$ 38 billion to over \$ 70 billion by 2025.	Completa	Sim	Sim
4	2025	The growth in the use of these platforms is rapid, and it is estimated that by 2025, nearly 540 million Internet users could use these platforms to find a job.	Completa	Sim	Não
5	2025	Vaccines market worth over \$ 70bn by 2025.	Completa	Não	Sim
6	2025	The Beyond Current Horizons programme was a particularly apt framework to work within because it set out to explore: ... what society might look like in 2025 in order to anticipate the demands that will be placed on the UK education system, taking as a focus not the future in its entirety, but specifically the intersection between technological, educational and social futures.	Parcial	Não	Sim
7	2025	China manufacturing locus in 2025: With a comparison of 'Made - in-China 2025 and 'Industry 4.0.	Parcial	Não	Sim
8	2025	From the point of view of automation experts, at least 35% of the jobs currently available will be automated over the next twenty years, and a study conducted by the McKinsey Global Institute predicted that by 2025, robotics may replace up to 75 million jobs worldwide.	Completa	Sim	Não
9	2025	A study by Kallio, Salminen, and Sievinen created three scenarios for the year 2025 to compare the effects of wood energy on the carbon balance.	Parcial	Sim	Sim
10	2025	Wind power in Finland up to the year 2025 -- soft scenarios based on expert views.	Parcial	Não	Sim

11	2030	By 2020 - 2030, annual growth rates are likely to increase and may exceed 10% -15 %.	Completa	Sim	Sim
12	2030	A 2008 study found that southern Africa could lose 30% of its maize crop by 2030 due to the negative effects of climate change.	Completa	Sim	Sim
13	2030	Accelerating economic growth leads to higher GDP, more energy use, more CO2 emissions and more use of resources with 11 SDGs achieved by 2030 but with greater global pressures on planetary boundaries.	Completa	Sim	Sim
14	2030	However because of the inherent complexity of the inter-related dilemmas involved, attaining improvement toward one SDG by 2030 may come at the expense of another .	Completa	Sim	Não
15	2030	In a third scenario which sees a stronger policy focus on achieving the SDGs leads to more goals being attained by 2030 and a halt in the downward trend of environmental goals.	Completa	Sim	Sim
16	2030	In fact, by 2030, it is expected that the rate of single-parent households will increase in all OECD countries.	Completa	Sim	Sim
17	2030	SDG 7 focuses on clean and affordable energy and aims for a substantial increase in the share of renewable energy in the global energy mix by 2030, whereas the sub-target of SDG13 addresses the national policies for climate change mitigation.	Completa	Sim	Sim
18	2030	To deliver all the SDGs by 2030 will require a concerted global effort, in which both time and monetary constraints must be overcome.	Completa	Sim	Sim
19	2030	The 17 goals lay out a clear plan of where humanity wants to go by 2030.	Parcial	Não	Sim
20	2030	Based on the National Energy and Climate Strategy for 2030 government report, the use of woodchips should increase significantly in the future due to increased use for heat and electricity production, as well as fuel production for transport .	Completa	Sim	Não
21	2030	Government report on the national energy and climate strategy for 2030.	Parcial	Não	Sim
22	2030	Predicting the future of additive manufacturing: A Delphi study on economic and societal implications of 3D printing for 2030.	Inválida	Não	Sim
23	2030	The Bookends Scenarios: Alternative Futures for the Public Library Network in NSW in 2030.	Parcial	Não	Não
24	2030	Future images of meat consumption in 2030.	Parcial	Não	Sim

25	2030	Introduction In this paper, we discuss using developed scenarios as a strategic tool to engage various stakeholders located in different countries as part of a larger survey for an international research project to gain a broad understanding of transport experts opinions regarding transport options and mobility-related commodity prices in the Asia-Pacific region in 2030.	Inválida	Não	Sim
26	2030	It has set up a market to establish the probability of human extinction , although the market clearing price, which will represent its final prediction, will not be available until it closes in 2030.	Parcial	Sim	Não
27	2030	Lake management in 2030 -- Five future images based on an international Delphi study.	Parcial	Não	Sim
28	2030	Prioritizing climate change adaptation needs for food security in 2030.	Completa	Não	Sim
29	2030	Projected prevalence of car-sharing in four asian-pacific countries in 2030: What the experts think.	Completa	Não	Sim
30	2030	The Bookends Scenarios: Alternative Futures for the Public Library Network in NSW in 2030.	Parcial	Não	Não
31	2030	The report estimates that in 2030, the use will be around 14 -- 18 million m ³ .	Parcial	Sim	Não
32	2030	Even if technological capacities for longer human missions are developed in next decades, the social and political context on Earth may become less conducive to space missions because of climate change, population growth in some parts of the globe, population decline in other parts, and the world limited ability to distribute human resources equitably, leaving enough to finance manned space missions.	Completa	Não	Sim
33	2030	In terms of implications for their future careers, their reflections spanned a broad spectrum -- while for some it further confirmed the belief that the professional choice they are making is right , some participants reflected on the fact that their desired jobs as understood today might not exist at all in the next decades.	Completa	Não	Sim
34	2030	In September 2016, the country ratified the Paris Agreement , committing to reduce emissions by 37% by 2025 and by 43% by 2030, based on 2005 reference values.	Completa	Sim	Não
35	2030	The future of families to 2030.	Parcial	Não	Sim
36	2030	Whilst pragmatically focusing on the short term through to 2030, the lack of long-term planning has been highlighted by the absence of global population goals and	Completa	Sim	Não

		boundaries in those targets considering environmental constraints, despite well-known connections between population growth and environmental degradation .			
37	2030	For example, Tapio, Rintam?ki, Rikkonen, and Ruotsalainen created six energy scenarios for farms up to the year 2030.	Parcial	Sim	Sim
38	2033	According to Frey and Osborne , 47% of total US employment is in the high-risk category, being these jobs expected to be automated relatively soon, perhaps by 2023 or 2033.	Completa	Sim	Não
39	2035	Brazil is expected to become a major exporter of oil and to be the world s 6th largest energy provider by 2035 .	Completa	Sim	Não
40	2035	NASA plans have included sending astronauts to Mars by 2035, and China also intends to build a lunar colony followed by a Mars colony.	Completa	Sim	Sim
41	2038	Three of the future worlds developed narrative arcs centred around a global state of crisis along the pathway to 2038.	Completa	Sim	Sim
42	2040	In 2040 ; oil & gas will still occupy 50% of the energy matrix.	Completa	Não	Sim
43	2040	In 2040, in Mexico, 50% of jobs are independent entrepreneurs who take on contracts with di ?erent companies simultaneously in such a way that they compromise the delivery times of each project but do not have a committed schedule with any employer.	Completa	Não	Sim
44	2040	In 2040, in Mexico, there are no working days, and everything is assigned by projects and resolved in the time that was committed for delivery.	Completa	Sim	Sim
45	2040	Based on the existing energy policies and announced policy intentions, the amount of global energy consumption is projected to increase almost 80% between 2016 and 2040 .	Completa	Sim	Não
46	2040	Contents lists available atScienceDirect Futures journal homepage: www.elsevier.com/locate/futures Work/family life by 2040: Between a gig economy and traditional roles Guillermmina Benavides Rinc?n *, Yolanda Montes Mart?nez Tecnologico de Monterrey, Mexico ARTICLE INFO Keywords: Scenarios Gender Work/life balance Work ?exibility ABSTRACT This research envisions four future scenarios of work/life balance for Mexican families by 2040.	Inválida	Não	Sim

47	2040	From this perspective, the use of a scenario approach is relevant for analyzing plausible futures to generate balance in work and family life by 2040.	Parcial	Não	Sim
48	2040	The following process was followed for the construction of the scenarios of the work/life balance by 2040: 1 A literature review was conducted to identify and categorize the main driving forces affecting the phenomenon under study.	Parcial	Não	Sim
49	2040	The uncertainty of this axis focuses on the issue of whether by 2040 in Mexico the contemporary traditional family will have disappeared and been replaced by more radical and complex forms of family agreements in households.	Completa	Sim	Sim
50	2040	Future family and work/life balance in 2040.	Parcial	Não	Não
51	2040	The scenario methodology chosen to analyze the possible future work/life balance in 2040 is the deductive method of the intuitive logic scenario approach, which is commonly referred to as the `` Shell model ''.	Parcial	Não	Não
52	2040	This issue is considered one of the greatest uncertainties because while there is a clear trend towards entrepreneurship economics and automation worldwide, we do not know whether the economic conditions, infrastructure and social acceptance are a breeding ground or barrier to the proliferation of this type of economy in Mexico in 2040.	Completa	Sim	Sim
53	2040	This research considered the year 2040, which is slightly over 20 years in the future and when the children of the couples interviewed will face this duality of roles, i.e., having to reconcile work and family life, as a reference.	Parcial	Não	Sim
54	2050	Period USA Japan Europe 2002 0.120.69 0.23 2003 0.130.30 0.20 2004 0.080.25 0.26 2005 0.290.32 0.33 2006 0.200.29 0.33 2007 0.490.31 0.39 2008 0.510.35 0.31 2009 0.570.35 0.36 2010 0.570.34 0.36 2020 0.670.36 0.39 2030 0.790.38 0.42 2040 0.930.40 0.45 2050 1.100.42 0.49 L. Aldieri, et al..	Inválida	Não	Sim
55	2050	According to Batissti : By 2050, under a typical middle-of-the-road emissions scenario, you looking at a doubling of the volatility for grains in the mid - latitudes.	Parcial	Sim	Sim
56	2050	The Brazilian population annual growth is about 0.8 %, slowing towards 0.5% in 2030 and stabilizing at 240 million people around 2050.	Completa	Sim	Sim
57	2050	Given the rate and scale of economic change required to minimise climate risks , in navigating the terrain ahead we should expect that knowledge systems and practices	Completa	Sim	Não

		established to deal with past and even current change processes will at best provide partial guidance, and at worst be misguiding.			
58	2050	Interview of Jared Diamond by David Wallace-Wells, `` Jared Diamond: There a 49 percent chance the world as we know it will end by 2050.	Parcial	Não	Sim
59	2050	Using a quantitative global simulation model called Earth 3 Randers, Rockstrom, Stoknes, Goluke, Collste, Cornell et al. have considered socio-economic and biophysical development through to 2050, and using one modelable indicator for each goal have assessed the number of SDGs likely to be achieved by 2030 and 2050.	Completa	Sim	Sim
60	2050	The OECD predicts that life expectancy at birth will continue to increase until 80 and 90 years of age at the global level by 2050.	Completa	Sim	Sim
61	2050	Results Tables 2 and 3 show the results of the impact of R&D innovation on productivity over the period 2002 -- 2010 and the forecast value for 2020, 2030, 2040 and 2050 .	Inválida	Não	Sim
62	2050	A public opinion survey of four future scenarios for Australia in 2050.	Parcial	Não	Sim
63	2050	Average distance and travel time within daily life per capita by all modes, motorized public, motorized private, and NMT for 2015 -- 16, and Socioeconomic Change Only, Low Demand, and High Accessibility scenarios in 2050.	Completa	Sim	Não
64	2050	Average per-capita travel distance and travel time by NMT by representative is in the range of 0 ? 2 km/day and 3 ? 23 min/day in 2015 ? 16, 0 ? 2 km/day and 4 ? 23 min/day in Socioeconomic Change Only, 1 ? 4 km/day and 10 ? 44 min/day in Low Demand, and 0 ? 2 km/day and 2 ? 14 min/day in High Accessibility in 2050, respectively.	Completa	Sim	Não
65	2050	Average per-capita travel distance and travel time by motorized private transportation by representative is in the range of 4 ? 46 km/day and 10 ? 72 min/day in 2015 ? 16, 5 ? 42 km/day and 11 ? 65 min/day in Socioeconomic Change Only, 3 ? 36 km/day and 8 ? 55 min/day in Low Demand, and 9 ? 59 km/day and 17 ? 103 min/day in High Accessibility in 2050, respectively.	Completa	Sim	Não
66	2050	Average per-capita travel distance and travel time by motorized public transportation by representative is in the range of 1 ? 32 km/day and 1 ? 66 min/day in 2015 ? 16, 1 ? 32 km/day and 1 ? 65 min/day in Socioeconomic Change Only, 0 ? 25 km/day and	Completa	Sim	Não

		0 ? 52 min/day in Low Demand, and 0 ? 22 km/day and 0 ? 42 min/day in High Accessibility in 2050, respectively.			
67	2050	Commissioning Organization Title of Scenario Set Scenarios Location Principal Sector Addressed Client Type Columbia Basin Trust Shaping Our Future Together British Columbia, Canada Environment Agency Arctic Council Arctic Marine Navigation in 2050 Arctic region Trade & Industry Association Association of Research Libraries The ARL 2030 Scenarios Washington DC, USA Research libraries Association Trade Knowledge Network Trade and Agriculture in Southern Africa Zimbabwe Trade Network Office of the Presidency South Africa Scenarios 2014 South Africa National development Government Conservation International / CSIRO Exploring Ecotourism in Milne Bay-Papua New Guinea Papua New Guinea Tourism Agency / Research 9 Public Library Network, New South Wales The Bookends Scenarios New South Wales, Australia Public libraries Network Nautilus Institute North Korean Nuclear Crisis North Korea International Relations Think tank Cofisa: Collaboration Finland-South Africa Biotechnology Scenarios for the Eastern Cape Eastern Cape, South Africa Biotechnology Agency / Government Bord Bia The Future of the Irish Food & Drink Market 2020 Ireland Food & Agriculture Trade association United Parcel Service Migration to Open Source Systems United States Software & Logistics Corporation UN Food & Agriculture Organization Food for Thought: Discovering Common Ground Egypt Food & Agriculture Development World Economic Forum Future of the Global Financial System International Finance Think tank Futures 115 102479 A.V. Gordon Futures 115 102479 Appendix B Table B1 Interview Schedule.	Inválida	Não	Não
68	2050	Contrasting future passenger travel demand scenarios in Japan in 2050, that is, Low Demand and High Accessibility, were developed in accordance with qualitative scenarios and quantitative assumptions of four mega-trends: population aging, gender equality, tertiary industrialization, and transport technology innovation.	Completa	Sim	Não
69	2050	Discussion In Population Change Only, travel demand by motorized private transportation decreases by 28% because the rural population is projected to decrease selectively in 2050.	Completa	Sim	Não

70	2050	Fig. 5 shows average distance with respect to average travel time per capita within daily life by mode of transportation in 2015 ? 16, and the Socioeconomic Change Only, Low Demand, and High Accessibility scenarios in 2050.	Parcial	Não	Não
71	2050	Future scenarios of passenger travel demand for daily life in Japan Future scenarios of passenger travel demand within daily life area in Japan in 2050 were calculated by a combination of two levels of change of the corresponding term of Mega-trends shown in Table 2.	Parcial	Sim	Não
72	2050	In 2050, average distance and travel time within daily life in the Socioeconomic Change Only scenario shows little change except for women and those aged 65 and over compared to those in 2015 ? 16.	Parcial	Sim	Não
73	2050	In the Low Demand scenario, total travel demand by motorized private transportation is found to decrease by 45% in 2050 compared to that in 2015 ? 16.	Completa	Sim	Não
74	2050	Population by region, sex, age in 2050 is calculated based on -- population projection by the National Institute of Population and Social Security Research , National Institute of Population and Social Security Research -RRB-.	Parcial	Sim	Não
75	2050	The scenarios stalled in use .2 For the 13 scenario projects studied, the table of matrix purpose, project purpose, and reported success-in-use, as defined in section 2 above, is as follows: Commissioning organization Scenario project Scenario matrix Scenario project Congruence of Reported Success purpose purpose purposes in use Columbia Basin Trust Shaping Our Future Together Normative Normative Yes Yes Arctic Council Arctic Marine Navigation in 2050 Normative Normative Yes Yes Association of Research Libraries The ARL 2030 Scenarios Adaptive Adaptive Yes Yes Trade Knowledge Network Trade and Agriculture in Southern Africa Normative Normative Yes Yes Office of the Presidency South Africa Scenarios 2014 Adaptive Normative No No Conservation International / CS - Exploring Ecotourism in Milne Bay-Papua Adaptive Adaptive Yes Yes IRO New Guinea Public Library Network, New So - The Bookends Scenarios Adaptive Normative No No uth Wales Nautilus Institute North Korean Nuclear Crisis Normative Adaptive No No Cofisa: Collaboration Finland-So - Biotechnology Scenarios for the Eastern Adaptive Normative No No uth Africa Cape Bord Bia The Future of the Irish Food & Drink	Inválida	Não	Não

		Adaptive Adaptive Yes Yes Market 2020 United Parcel Service Migration to Open Source Systems Adaptive Adaptive Yes Yes UN Food & Agriculture Organiz - Food for Thought: Discovering Common Normative Normative Yes No ation Ground World Economic Forum Future of the Global Financial System Normative Normative Yes Yes In all cases where there was congruence of implied scenario purpose and scenario project purpose, there was success in use, as defined -LRB- with one exception: UN FAO, where the `` Arab Spring \' altogether overwhelmed the implementation of scenarios.			
76	2050	The world in 2050 and the New Welfare scenario.	Parcial	Não	Sim
77	2050	Their end result was a futures image of sustainable food consumption in 2050, wherein energy efficiency was the key component.	Completa	Sim	Sim
78	2050	Total passenger travel demand within daily life area in 2050.	Parcial	Não	Não
79	2050	India: A Dharma based Country in 2050.	Parcial	Não	Não
80	2050	As shown in Fig. 4, passenger travel by non-motorized, motorized private, and motorized public transportation within daily life area in 2050 are 40 billion person-km/yr, 471 billion person-km/yr, and 377 billion person-km/yr in Population Change Only, 44 billion person-km/yr, 494 billion person-km/yr, and 411 billion person-km/yr in Socioeconomic Change Only, 88 billion person-km / yr, 356 billion person-km/yr, and 385 billion person-km/yr in Low Demand, and 29 billion person-km/yr, 852 billion person-km/yr, and 266 billion person-km/yr in High Accessibility, respectively.	Completa	Sim	Não
81	2050	Travel demand for commuting, work, and private trips in 2050 are 407 billion person-km/yr, 78 billion person-km/yr, and 402 billion person-km/yr in Population Change Only, 462 billion person-km / yr, 78 billion person-km/yr, and 409 billion person-km/yr in Socioeconomic Change Only, 373 billion person-km/yr, 79 billion person - km/yr, and 377 billion person-km/yr in Low Demand, and 545 billion person-km/yr, 98 billion person-km/yr, and 503 billion person - km/yr in High Accessibility, respectively.	Completa	Sim	Não
82	2050	Arctic Marine Navigation to 2050.	Parcial	Não	Não
83	2050	Scenarios for Russia natural gas exports to 2050.	Parcial	Não	Sim

84	2050	Paltsev, for example, approaches the economic opportunities of Russia natural gas exports through possible scenarios until 2050, including threats to Russian exports, such as the emerging shale gas technologies development across the globe, the EU energy policy , and the transformation of lique ?ed natural gas markets.	Completa	Sim	Sim
85	2052	2052: A global forecast for the next forty years.	Parcial	Não	Não
86	2068	3.1.4 2038 Worldbuilding and artefact development The participants worked in teams and were facilitated through a design process to develop an artefact which would be of sig - ni ?cance in 2038, halfway along the pathways towards 2068.	Inválida	Não	Sim
87	2068	4.2.5 On roles and agency in 2068 Alongside stories from 2068, the participants were also asked to describe their roles in those possible futures.	Inválida	Não	Sim
88	2068	For example, discussing responsibility in man - agement education, Henderson, Wersun, Wilson, Mo-ching Yeung, and Zhang equate responsible management with man - agement for sustainability, in a context in which ` Humanity very existence in 2068 may be considered to be uncertain.	Completa	Sim	Não
89	2068	For this publication about a wiser future in 2068, I look for the babyhood, the infant, in the contributions that I received and explore what they might sound like if the conscious assumptions become that there is enough for everyone and that humans are driven to joy, and peace.	Inválida	Não	Sim
90	2068	Future visions summer school design The purpose of the week -- long workshop was to develop a future narrative , and then work in teams to design and prototype an artefact .	Inválida	Não	Não
91	2068	Futures 115 102477 9 The Future in 2068 - An Alien Short StoryRafael Lourenco Prologue My father believed in aliens.	Inválida	Não	Sim
92	2068	Principles for responsible management education in 2068.	Inválida	Não	Sim
93	2068	The present research aims to explore young people images of the future through design futures methods.	Inválida	Não	Não
94	2068	There is a shared sense that technology will be fully embedded in our daily lives in 2068, that quality of life has improved lifestyles globally -- from solving tra? c problems, to rapid development of technology for good across healthcare, food	Completa	Não	Sim

		supplies, energy production, as well as pollution management and ecosystem restoration.			
95	2068	They were asked to write a short story about the world as they imagine it in 2068.	Inválida	Não	Sim
96	2068	This issue focused on a theme of ` responsible futures, asking contributors to answer the question: what futures should humanity strive for ? What might be considered responsible and wise in 2068, and why ? 1.	Inválida	Não	Sim
97	2068	This issue focused on a theme of ` responsible futures, asking contributors to answer the questions: what futures should humanity strive for ? What might be considered responsible and wise in 2068, and why ? The central theme of the issue was inspired by Max Tegmark call for debate on the Futures of Life, with explicit reference to artificial intelligence, gods and ` apocalypse.	Inválida	Não	Não
98	2068	What will be the long-term consequences of HCI, AI, IoT, Big Data and Smart Technologies 50 years from now -- in 2068 ? .	Parcial	Sim	Não
99	2068	3.1.2 Storytelling 2068 As part of the ?rst survey process, the participants were asked to write a story about the world as they imagine it in 2068 and describe the role they would have played in that world.	Inválida	Não	Sim
100	2068	T alternative futures set in 2038 and 2068 that were developed through design futures methods during two week -- long Global Summer Schools at Imperial College London in July and August 2018.	Inválida	Não	Não
101	2068	Envisioning 2068 The next exercise took participants further into the future to 2068.	Inválida	Não	Sim
102	2068	The Summer Schools started with a series of future -- casting exercises, followed by the development of a suite of detailed 2038 worlds as key markers on the pathways to 2068, as well as artefacts consistent with the logic of their speculative futures.	Inválida	Não	Sim
103	2068	London: Penguin, 2005, p. 76-77 -RRB- Introduction Maaike Engelen Question: Which practices will contribute to a wiser future ? Picture the year 2068.	Inválida	Não	Não
104	2070	Three critical transformations... decarbonising the world economy by 2050 to 2070, feeding the world through sustainable agriculture by 2050 and accelerating progress toward an economy of cyclic material flows	Completa	Sim	Não
105	2100	Even with the current mitigation commitments and pledges fully implemented, there is roughly a 20 percent likelihood of exceeding 4 °C by 2100.	Completa	Sim	Sim

106	2100	Such a warming level and associated sea-level rise of 0.5 -- 1 meter, or more, by 2100 would not be the end point: a further warming to levels over 6 °C, with several meters of sea-level rise, would likely occur over the following centuries.	Completa	Sim	Não
107	2100	Compounded over the 21 st century, the annual growth projections result in a three to eight-fold increase in global per-capita income by 2100 regardless of biophysical constraints .	Completa	Não	Não
108	2100	I wo be alive in 2100, but my grandchildren will ... so we are talking about people that are sharing life with us.	Inválida	Não	Sim
109	2100	As Clark et al. put it: `` An evaluation of climate change risks that only considers the next 85 years of climate change impacts fails to provide essential information to stakeholders, the public and the political leaders who will ultimately be tasked with making decisions about policies on behalf of all, with impacts that will last for millennia.	Parcial	Não	Sim
110	2100	I have already mentioned the computation of the cost of climate action / inaction at the year 2100, which pales when compared to what the authors in call `` A Million Years of Certainty \' relative to a model-based computation of the risk of a nuclear waste disposal one million years into the future.	Completa	Sim	Não
111	2100	Climate models indicate that the Earth could warm by 3 °C -4 °C by the year 2100 and eventually by as much as 8 °C or more.	Completa	Sim	Sim
112	2100	Most projections of global warming focus on either the year 2100 or the e ffects of a doubling of CO 2 .	Completa	Sim	Sim
113	2100	Unless draconian measures are taken to halt the increase in atmospheric CO 2 , global temperature will likely increase by at least 3 °C above today by the year 2100 and could eventually increase by 8 °C or more .	Completa	Sim	Sim
114	2100	If incomes continue to rise in most countries and richer people have fewer children, then world population should peak at 9 -- 11 billion around the year 2100.	Completa	Sim	Não
115	2100	The idea that numbers as different as ? A consumer price index ? A credit rating ? The price of a bundle of derivatives ? The number of planets needed by humanity in one year ? The gross domestic product to the year 2100 ? The numerical risk of an	Inválida	Não	Não

		industrial product or practice need to come under some sort of collective ethical umbrella or even policing may sound as a topic for science fiction.			
116	2300	Longer-term projections are highly speculative and show everything from runaway growth to a population crash down to 2.3 billion in 2300.	Parcial	Não	Não
117	2300	The lack of attention to the very long run is a serious shortcoming, since integrated carbon-climate models project that if CO ₂ from current in situ fossil fuel resources continues to be released into the atmosphere, the peak concentration of atmospheric CO ₂ could exceed 1400 ppm by the year 2300 and the average global temperature could warm by 8 °C or more.	Completa	Sim	Não

4. Ciclo 3: Prospecções obtidas pelo NERMAP (*Journal Futures*)

ID	Data	Prospecção	Tipo de Prospecção	Encontrada manualmente	Contida no Ciclo 2
1	2021	Here are two examples of the narratives that emerged: Statement 1: Over the next years more people from UK and abroad would visit the centre of Chester ... For our sustainability, it is essential that long-term sustainable income streams are identified.	Completa	Não	Sim
2	2025	In 2025 this track of land housed an experimental atom modifier, which during its explosion caused a chemical spill in a nearby plant.	Completa	Sim	Sim
3	2025	According to the vaccines market is estimated to rise from the present value of \$ 38 billion to over \$ 70 billion by 2025.	Completa	Sim	Sim
4	2025	The growth in the use of these platforms is rapid, and it is estimated that by 2025, nearly 540 million Internet users could use these platforms to find a job.	Completa	Sim	Sim
5	2025	Vaccines market worth over \$ 70bn by 2025.	Completa	Não	Sim
6	2025	The Beyond Current Horizons programme was a particularly apt framework to work within because it set out to explore: ... what society might look like in 2025 in order to anticipate the demands that will be placed on the UK education system, taking as a focus not the future in its entirety, but specifically the intersection between technological, educational and social futures.	Parcial	Não	Sim
7	2025	China manufacturing locus in 2025: With a comparison of 'Made - in-China 2025 and 'Industry 4.0.	Parcial	Não	Sim
8	2025	Food scenarios 2025: Drivers of change between global and regional.	Parcial	Não	Não
9	2025	From the point of view of automation experts, at least 35% of the jobs currently available will be automated over the next twenty years, and a study conducted by the McKinsey Global Institute predicted that by 2025, robotics may replace up to 75 million jobs worldwide.	Completa	Sim	Sim
10	2025	A study by Kallio, Salminen, and Sievinen created three scenarios for the year 2025 to compare the effects of wood energy on the carbon balance.	Parcial	Sim	Sim

11	2025	Wind power in Finland up to the year 2025 -- soft scenarios based on expert views.	Parcial	Não	Sim
12	2030	By 2020 - 2030, annual growth rates are likely to increase and may exceed 10% -15 %.	Completa	Sim	Sim
13	2030	The 2030 agenda emphasises the SDGs are interlinked, and recognises the importance of ensuring integration across all the sectors they represent, without indicating ways in which this might be achieved.	Completa	Sim	Não
14	2030	A 2008 study found that southern Africa could lose 30% of its maize crop by 2030 due to the negative effects of climate change.	Completa	Sim	Sim
15	2030	Accelerating economic growth leads to higher GDP, more energy use, more CO2emissions and more use of resources with 11 SDGs achieved by 2030 but with greater global pressures on planetary boundaries.	Completa	Sim	Sim
16	2030	However because of the inherent complexity of the inter-related dilemmas involved, attaining improvement toward one SDG by 2030 may come at the expense of another .	Completa	Sim	Sim
17	2030	In a third scenario which sees a stronger policy focus on achieving the SDGs leads to more goals being attained by 2030 and a halt in the downward trend of environmental goals.	Completa	Sim	Sim
18	2030	In fact, by 2030, it is expected that the rate of single-parent households will increase in all OECD countries.	Completa	Sim	Sim
19	2030	SDG 7 focuses on clean and a ?ordable energy and aims for a substantial increase in the share of renewable energy in the global energy mix by 2030, whereas the sub-target of SDG13 addresses the national policies for climate change mitigation.	Completa	Sim	Sim
20	2030	To deliver all the SDGs by 2030 will require a concerted global effort, in which both time and monetary constraints must be overcome.	Completa	Sim	Sim
21	2030	The 17 goals lay out a clear plan of where humanity wants to go by 2030.	Parcial	Não	Sim
22	2030	Based on the National Energy and Climate Strategy for 2030 government report, the use of woodchips should increase significantly in the future due to increased use for heat and electricity production, as well as fuel production for transport .	Completa	Sim	Sim
23	2030	Government report on the national energy and climate strategy for 2030.	Parcial	Não	Sim

24	2030	Predicting the future of additive manufacturing: A Delphi study on economic and societal implications of 3D printing for 2030.	Inválida	Não	Sim
25	2030	The Bookends Scenarios: Alternative Futures for the Public Library Network in NSW in 2030.	Parcial	Não	Sim
26	2030	Future images of meat consumption in 2030.	Parcial	Não	Sim
27	2030	Introduction In this paper, we discuss using developed scenarios as a strategic tool to engage various stakeholders located in different countries as part of a larger survey for an international research project to gain a broad understanding of transport experts opinions regarding transport options and mobility-related commodity prices in the Asia-Pacific region in 2030.	Inválida	Não	Sim
28	2030	It has set up a market to establish the probability of human extinction , although the market clearing price, which will represent its final prediction, will not be available until it closes in 2030.	Parcial	Sim	Sim
29	2030	Lake management in 2030 -- Five future images based on an international Delphi study.	Parcial	Não	Sim
30	2030	Prioritizing climate change adaptation needs for food security in 2030.	Completa	Não	Sim
31	2030	Projected prevalence of car-sharing in four asian-pacific countries in 2030: What the experts think.	Completa	Não	Sim
32	2030	The Bookends Scenarios: Alternative Futures for the Public Library Network in NSW in 2030.	Parcial	Não	Sim
33	2030	The report estimates that in 2030, the use will be around 14 -- 18 million m ³ .	Parcial	Sim	Sim
34	2030	Even if technological capacities for longer human missions are developed in next decades, the social and political context on Earth may become less conducive to space missions because of climate change, population growth in some parts of the globe, population decline in other parts, and the world limited ability to distribute human resources equitably, leaving enough to finance manned space missions.	Completa	Não	Sim
35	2030	In terms of implications for their future careers, their reflections spanned a broad spectrum -- while for some it further confirmed the belief that the professional choice they are making is right , some participants reflected on the fact that their desired jobs as understood today might not exist at all in the next decades.	Completa	Não	Sim

36	2030	In September 2016, the country ratified the Paris Agreement, committing to reduce emissions by 37% by 2025 and by 43% by 2030, based on 2005 reference values.	Completa	Sim	Sim
37	2030	A second example, the Association of Research Libraries 2030 scenarios, is represented by the matrix: constraints on individual researchers, and research enterprise aggregation.	Parcial	Sim	Não
38	2030	The 2030 target is largely political, set for consistency to match the 15-year timeframe of the Millennium Development Goals.	Completa	Sim	Não
39	2030	The future of families to 2030.	Parcial	Não	Sim
40	2030	Whilst pragmatically focusing on the short term through to 2030, the lack of long-term planning has been highlighted by the absence of global population goals and boundaries in those targets considering environmental constraints, despite well-known connections between population growth and environmental degradation.	Completa	Sim	Sim
41	2030	For example, Tapiro, Rintamäki, Rikkonen, and Ruotsalainen created six energy scenarios for farms up to the year 2030.	Parcial	Sim	Sim
42	2033	According to Frey and Osborne, 47% of total US employment is in the high-risk category, being these jobs expected to be automated relatively soon, perhaps by 2023 or 2033.	Completa	Sim	Sim
43	2035	Brazil is expected to become a major exporter of oil and to be the world's 6th largest energy provider by 2035.	Completa	Sim	Sim
44	2035	NASA plans have included sending astronauts to Mars by 2035, and China also intends to build a lunar colony followed by a Mars colony.	Completa	Sim	Sim
45	2038	The worldbuilding and artefact development processes enabled participants to develop detailed pathways to 2038 and create an artefact which would bring to life an aspect of their envisioned future.	Completa	Sim	Não
46	2038	Three of the future worlds developed narrative arcs centred around a global state of crisis along the pathway to 2038.	Completa	Sim	Sim
47	2040	In 2040 oil & gas will still occupy 50% of the energy matrix.	Completa	Não	Sim
48	2040	In 2040, in Mexico, 50% of jobs are independent entrepreneurs who take on contracts with different companies simultaneously in such a way that they compromise the	Completa	Não	Sim

		delivery times of each project but do not have a committed schedule with any employer.			
49	2040	In 2040, in Mexico, there are no working days, and everything is assigned by projects and resolved in the time that was committed for delivery.	Completa	Sim	Sim
50	2040	Based on the existing energy policies and announced policy intentions, the amount of global energy consumption is projected to increase almost 80% between 2016 and 2040 .	Completa	Sim	Sim
51	2040	Contents lists available at ScienceDirect Futures journal homepage: www.elsevier.com/locate/futures Work/family life by 2040: Between a gig economy and traditional roles Guillermina Benavides Rinc?n *, Yolanda Montes Mart?nez Tecnologico de Monterrey, Mexico ARTICLE INFO Keywords: Scenarios Gender Work/life balance Work ?exibility ABSTRACT This research envisions four future scenarios of work/life balance for Mexican families by 2040.	Inválida	Não	Sim
52	2040	From this perspective, the use of a scenario approach is relevant for analyzing plausible futures to generate balance in work and family life by 2040.	Parcial	Não	Sim
53	2040	The following process was followed for the construction of the scenarios of the work/life balance by 2040: 1 A literature review was conducted to identify and categorize the main driving forces affecting the phenomenon under study.	Parcial	Não	Sim
54	2040	The uncertainty of this axis focuses on the issue of whether by 2040 in Mexico the contemporary traditional family will have disappeared and been replaced by more radical and complex forms of family agreements in households.	Completa	Sim	Sim
55	2040	Future family and work/life balance in 2040.	Parcial	Não	Sim
56	2040	The scenario methodology chosen to analyze the possible future work/life balance in 2040 is the deductive method of the intuitive logic scenario approach, which is commonly referred to as the `` Shell model ''.	Parcial	Não	Sim
57	2040	This issue is considered one of the greatest uncertainties because while there is a clear trend towards entrepreneurship economics and automation worldwide, we do not know whether the economic conditions, infrastructure and social acceptance are a breeding ground or barrier to the proliferation of this type of economy in Mexico in 2040 .	Completa	Sim	Sim

58	2040	This research considered the year 2040, which is slightly over 20 years in the future and when the children of the couples interviewed will face this duality of roles, i.e., having to reconcile work and family life, as a reference.	Parcial	Não	Sim
59	2050	Socio-technical scenarios as a methodological tool to explore social and political feasibility in low-carbon transitions: Bridging computer models and the multi-level perspective in UK electricity generation.	Parcial	Não	Não
60	2050	According to Batissti : By 2050, under a typical middle-of-the-road emissions scenario, you looking at a doubling of the volatility for grains in the mid - latitudes.	Parcial	Sim	Sim
61	2050	The Brazilian population annual growth is about 0.8 %, slowing towards 0.5% in 2030 and stabilizing at 240 million people around 2050.	Completa	Sim	Sim
62	2050	Given the rate and scale of eco - nomic change required to minimise climate risks , in navigating the terrain ahead we should expect that knowledge systems and practices established to deal with past and even current change processes will at best provide partial guidance, and at worst be misguiding.	Completa	Sim	Sim
63	2050	Interview of Jared Diamond by David Wallace-Wells, `` Jared Diamond: There a 49 percent chance the world as we know it will end by 2050.	Parcial	Não	Sim
64	2050	Using a quantitative global simulation model called Earth 3 Randers, Rockstrom, Stoknes, Goluke, Collste, Cornell et al. have considered socio-economic and biophysical development through to 2050, and using one modelable indicator for each goal have assessed the number of SDGs likely to be achieved by 2030 and 2050.	Completa	Sim	Sim
65	2050	The OECD predicts that life ex - pectancy at birth will continue to increase until 80 and 90 years of age at the global level by 2050.	Completa	Sim	Sim
66	2050	Results Tables 2 and 3 show the results of the impact of R&D innovation on productivity over the period 2002 -- 2010 and the forecast value for 2020, 2030, 2040 and 2050 .	Inválida	Não	Sim
67	2050	A public opinion survey of four future scenarios for Australia in 2050.	Parcial	Não	Sim
68	2050	Average distance and travel time within daily life per capita by all modes, motorized public, motorized private, and NMT for 2015 -- 16, and Socioeconomic Change Only, Low Demand, and High Accessibility scenarios in 2050.	Completa	Sim	Sim

69	2050	Average per-capita travel distance and travel time by NMT by representative is in the range of 0 ? 2 km/day and 3 ? 23 min/day in 2015 ? 16, 0 ? 2 km/day and 4 ? 23 min/day in Socioeconomic Change Only, 1 ? 4 km/day and 10 ? 44 min/day in Low Demand, and 0 ? 2 km/day and 2 ? 14 min/day in High Accessibility in 2050, respectively.	Completa	Sim	Sim
70	2050	Average per-capita travel distance and travel time by motorized private transportation by representative is in the range of 4 ? 46 km/day and 10 ? 72 min/day in 2015 ? 16, 5 ? 42 km/day and 11 ? 65 min/day in Socioeconomic Change Only, 3 ? 36 km/day and 8 ? 55 min/day in Low Demand, and 9 ? 59 km/day and 17 ? 103 min/day in High Accessibility in 2050, respectively.	Completa	Sim	Sim
71	2050	Average per-capita travel distance and travel time by motorized public transportation by representative is in the range of 1 ? 32 km/day and 1 ? 66 min/day in 2015 ? 16, 1 ? 32 km/day and 1 ? 65 min/day in Socioeconomic Change Only, 0 ? 25 km/day and 0 ? 52 min/day in Low Demand, and 0 ? 22 km/day and 0 ? 42 min/day in High Accessibility in 2050, respectively.	Completa	Sim	Sim
72	2050	Commissioning Organization Title of Scenario Set Scenarios Location Principal Sector Addressed Client Type Columbia Basin Trust Shaping Our Future Together British Columbia, Canada Environment Agency Arctic Council Arctic Marine Navigation in 2050 Arctic region Trade & Industry Association Association of Research Libraries The ARL 2030 Scenarios Washington DC, USA Research libraries Association Trade Knowledge Network Trade and Agriculture in Southern Africa Zimbabwe Trade Network Office of the Presidency South Africa Scenarios 2014 South Africa National development Government Conservation International / CSIRO Exploring Ecotourism in Milne Bay-Papua New Guinea Papua New Guinea Tourism Agency / Research 9 Public Library Network, New South Wales The Bookends Scenarios New South Wales, Australia Public libraries Network Nautilus Institute North Korean Nuclear Crisis North Korea International Relations Think tank Cofisa: Collaboration Finland-South Africa Biotechnology Scenarios for the Eastern Cape Eastern Cape, South Africa Biotechnology Agency / Government Bord Bia The Future of the Irish Food & Drink Market 2020 Ireland Food & Agriculture Trade	Inválida	Não	Sim

		association United Parcel Service Migration to Open Source Systems United States Software & Logistics Corporation UN Food & Agriculture Organization Food for Thought: Discovering Common Ground Egypt Food & Agriculture Development World Economic Forum Future of the Global Financial System International Finance Think tank Futures 115 102479 A.V. Gordon Futures 115 102479 Appendix B Table B1 Interview Schedule.			
73	2050	Contrasting future passenger travel demand scenarios in Japan in 2050, that is, Low Demand and High Accessibility, were developed in accordance with qualitative scenarios and quantitative assumptions of four mega-trends: population aging, gender equality, tertiary industrialization, and transport technology innovation.	Completa	Sim	Sim
74	2050	Discussion In Population Change Only, travel demand by motorized private transportation decreases by 28% because the rural population is projected to decrease selectively in 2050.	Completa	Sim	Sim
75	2050	Fig. 5 shows average distance with respect to average travel time per capita within daily life by mode of transportation in 2015 ? 16, and the Socioeconomic Change Only, Low Demand, and High Accessibility scenarios in 2050.	Parcial	Não	Sim
76	2050	Future scenarios of passenger travel demand for daily life in Japan Future scenarios of passenger travel demand within daily life area in Japan in 2050 were calculated by a combination of two levels of change of the corresponding term of Mega-trends shown in Table 2.	Parcial	Sim	Sim
77	2050	In 2050, average distance and travel time within daily life in the Socioeconomic Change Only scenario shows little change except for women and those aged 65 and over compared to those in 2015 ? 16.	Parcial	Sim	Sim
78	2050	In the Low Demand scenario, total travel demand by motorized private transportation is found to decrease by 45% in 2050 compared to that in 2015 ? 16.	Completa	Sim	Sim
79	2050	Population by region, sex, age in 2050 is calculated based on -- population projection by the National Institute of Population and Social Security Research , National Institute of Population and Social Security Research -RRB-.	Parcial	Sim	Sim
80	2050	The scenarios stalled in use .2 For the 13 scenario projects studied, the table of matrix purpose, project purpose, and reported success-in-use, as defined in section 2 above,	Inválida	Não	Sim

		is as follows: Commissioning organization Scenario project Scenario matrix Scenario project Congruence of Reported Success purpose purpose purposes in use Columbia Basin Trust Shaping Our Future Together Normative Normative Yes Yes Arctic Council Arctic Marine Navigation in 2050 Normative Normative Yes Yes Association of Research Libraries The ARL 2030 Scenarios Adaptive Adaptive Yes Yes Trade Knowledge Network Trade and Agriculture in Southern Africa Normative Normative Yes Yes Office of the Presidency South Africa Scenarios 2014 Adaptive Normative No No Conservation International / CS - Exploring Ecotourism in Milne Bay-Papua Adaptive Adaptive Yes Yes IRO New Guinea Public Library Network, New So - The Bookends Scenarios Adaptive Normative No No uth Wales Nautilus Institute North Korean Nuclear Crisis Normative Adaptive No No Cofisa: Collaboration Finland-So - Biotechnology Scenarios for the Eastern Adaptive Normative No No uth Africa Cape Bord Bia The Future of the Irish Food & Drink Adaptive Adaptive Yes Yes Market 2020 United Parcel Service Migration to Open Source Systems Adaptive Adaptive Yes Yes UN Food & Agriculture Organiz - Food for Thought: Discovering Common Normative Normative Yes No ation Ground World Economic Forum Future of the Global Financial System Normative Normative Yes Yes In all cases where there was congruence of implied scenario purpose and scenario project purpose, there was success in use, as defined -LRB- with one exception: UN FAO, where the `` Arab Spring \' altogether overwhelmed the implementation of scenarios.			
81	2050	The world in 2050 and the New Welfare scenario.	Parcial	Não	Sim
82	2050	Their end result was a futures image of sustainable food consumption in 2050, wherein energy efficiency was the key component.	Completa	Sim	Sim
83	2050	Total passenger travel demand within daily life area in 2050.	Parcial	Não	Sim
84	2050	India: A Dharma based Country in 2050.	Parcial	Não	Sim
85	2050	As shown in Fig. 4, passenger travel by non-motorized, motorized private, and motorized public transportation within daily life area in 2050 are 40 billion person-km/yr, 471 billion person-km/yr, and 377 billion person-km/yr in Population Change Only, 44 billion person-km/yr, 494 billion person-km/yr, and 411 billion person-	Completa	Sim	Sim

		km/yr in Socioeconomic Change Only, 88 billion person-km / yr, 356 billion person-km/yr, and 385 billion person-km/yr in Low Demand, and 29 billion person-km/yr, 852 billion person-km/yr, and 266 billion person-km/yr in High Accessibility, respectively.			
86	2050	Travel demand for commuting, work, and private trips in 2050 are 407 billion person-km/yr, 78 billion person-km/yr, and 402 billion person-km/yr in Population Change Only, 462 billion person-km / yr, 78 billion person-km/yr, and 409 billion person-km/yr in Socioeconomic Change Only, 373 billion person-km/yr, 79 billion person - km/yr, and 377 billion person-km/yr in Low Demand, and 545 billion person-km/yr, 98 billion person-km/yr, and 503 billion person - km/yr in High Accessibility, respectively.	Completa	Sim	Sim
87	2050	Arctic Marine Navigation to 2050.	Parcial	Não	Sim
88	2050	Scenarios for Russia natural gas exports to 2050.	Parcial	Não	Sim
89	2050	Paltsev , for example, approaches the economic opportunities of Russia natural gas exports through possible scenarios until 2050, including threats to Russian exports, such as the emerging shale gas technologies development across the globe, the EU energy policy , and the transformation of lique ?ed natural gas markets.	Completa	Sim	Sim
90	2052	2052: A global forecast for the next forty years.	Parcial	Não	Sim
91	2063	Agenda 2063: The Africa we want.	Parcial	Não	Não
92	2063	The 2063 vision is focused on strengthening Africa role in a globalizing world, and aims to achieve the SDGs mainly through standard economic growth.	Completa	Sim	Não
93	2063	The 2063 vision is heavily embedded within conventional technological trends and does not emphasise the rich bio-cultural diversity of the continent.	Completa	Sim	Não
94	2068	4.2.6 Summary The process of envisioning 2068 has surfaced a number of key assumptions participants are making about how change might happen over a long -- time horizon.	Inválida	Não	Não
95	2068	3.1.4 2038 Worldbuilding and artefact development The participants worked in teams and were facilitated through a design process to develop an artefact which would be of sig - ni ?cance in 2038, halfway along the pathways towards 2068.	Inválida	Não	Sim

96	2068	For example, discussing responsibility in management education, Henderson, Wersun, Wilson, Mo-ching Yeung, and Zhang equate responsible management with management for sustainability, in a context in which ` Humanity very existence in 2068 may be considered to be uncertain.	Completa	Sim	Sim
97	2068	For this publication about a wiser future in 2068, I look for the babyhood, the infant, in the contributions that I received and explore what they might sound like if the conscious assumptions become that there is enough for everyone and that humans are driven to joy, and peace.	Inválida	Não	Sim
98	2068	Future visions summer school design The purpose of the week -- long workshop was to develop a future narrative , and then work in teams to design and prototype an artefact .	Inválida	Não	Sim
99	2068	Futures 115 102477 9 The Future in 2068 - An Alien Short StoryRafael Lourenco Prologue My father believed in aliens.	Inválida	Não	Sim
100	2068	Principles for responsible management education in 2068.	Inválida	Não	Sim
101	2068	The present research aims to explore young people images of the future through design futures methods.	Inválida	Não	Sim
102	2068	There is a shared sense that technology will be fully embedded in our daily lives in 2068, that quality of life has improved lifestyles globally -- from solving traffic problems, to rapid development of technology for good across healthcare, food supplies, energy production, as well as pollution management and ecosystem restoration.	Completa	Não	Sim
103	2068	They were asked to write a short story about the world as they imagine it in 2068.	Inválida	Não	Sim
104	2068	This issue focused on a theme of ` responsible futures, asking contributors to answer the question: what futures should humanity strive for ? What might be considered responsible and wise in 2068, and why ? 1.	Inválida	Não	Sim
105	2068	This issue focused on a theme of ` responsible futures, asking contributors to answer the questions: what futures should humanity strive for ? What might be considered responsible and wise in 2068, and why ? The central theme of the issue was inspired by Max Tegmark call for debate on the Futures of Life, with explicit reference to artificial intelligence, gods and ` apocalypse.	Inválida	Não	Sim

106	2068	What will be the long-term consequences of HCI, AI, IoT, Big Data and Smart Technologies 50 years from now -- in 2068 ? .	Parcial	Sim	Sim
107	2068	3.1.2 Storytelling 2068 As part of the ?rst survey process, the participants were asked to write a story about the world as they imagine it in 2068 and describe the role they would have played in that world.	Inválida	Não	Sim
108	2068	T alternative futures set in 2038 and 2068 that were developed through design futures methods during two week -- long Global Summer Schools at Imperial College London in July and August 2018.	Inválida	Não	Sim
109	2068	Envisioning 2068 The next exercise took participants further into the future to 2068.	Inválida	Não	Sim
110	2068	The Summer Schools started with a series of future -- casting exercises, followed by the development of a suite of detailed 2038 worlds as key markers on the pathways to 2068, as well as artefacts consistent with the logic of their speculative futures.	Inválida	Não	Sim
111	2068	London: Penguin, 2005, p. 76-77 -RRB- Introduction Maaike Engelen Question: Which practices will contribute to a wiser future ? Picture the year 2068.	Inválida	Não	Sim
112	2070	Three critical transformations ... decarbonising the world economy by 2050 to 2070, feeding the world through sustainable agriculture by 2050 and accelerating progress toward an economy of cyclic material flows	Completa	Sim	Sim
113	2100	Even with the current mitigation commitments and pledges fully implemented, there is roughly a 20 percent likelihood of exceeding 4 °C by 2100.	Completa	Sim	Sim
114	2100	Compounded over the 21 st century, the annual growth projections result in a three to eight-fold increase in global per-capita income by 2100 regardless of biophysical constraints .	Completa	Não	Sim
115	2100	Such a warming level and associated sea-level rise of 0.5 -- 1 meter, or more, by 2100 would not be the end point: a further warming to levels over 6 °C, with several meters of sea-level rise, would likely occur over the following centuries.	Completa	Sim	Sim
116	2100	I wo be alive in 2100, but my grandchildren will ... so we are talking about people that are sharing life with us.	Inválida	Não	Sim
117	2100	As Clark et al. put it: `` An evaluation of climate change risks that only considers the next 85 years of climate change impacts fails to provide essential information to stakeholders, the public and the political leaders who will ultimately be tasked with	Parcial	Não	Sim

		making decisions about policies on behalf of all, with impacts that will last for millennia.			
118	2100	I have already mentioned the computation of the cost of climate action / inaction at the year 2100, which pales when compared to what the authors in call `` A Million Years of Certainty ' relative to a model-based computation of the risk of a nuclear waste disposal one million years into the future.	Completa	Sim	Sim
119	2100	Climate models indicate that the Earth could warm by 3 °C -4 °C by the year 2100 and eventually by as much as 8 °C or more.	Completa	Sim	Sim
120	2100	Most projections of global warming focus on either the year 2100 or the effects of a doubling of CO ₂ .	Completa	Sim	Sim
121	2100	Unless draconian measures are taken to halt the increase in atmospheric CO ₂ , global temperature will likely increase by at least 3 °C above today by the year 2100 and could eventually increase by 8 °C or more.	Completa	Sim	Sim
122	2100	If incomes continue to rise in most countries and richer people have fewer children, then world population should peak at 9 -- 11 billion around the year 2100.	Completa	Sim	Sim
123	2100	The idea that numbers as different as ? A consumer price index ? A credit rating ? The price of a bundle of derivatives ? The number of planets needed by humanity in one year ? The gross domestic product to the year 2100 ? The numerical risk of an industrial product or practice need to come under some sort of collective ethical umbrella or even policing may sound as a topic for science fiction.	Inválida	Não	Sim
124	2300	Longer-term projections are highly speculative and show everything from runaway growth to a population crash down to 2.3 billion in 2300.	Parcial	Não	Sim
125	2300	The lack of attention to the very long run is a serious shortcoming, since integrated carbon-climate models project that if CO ₂ from current in situ fossil fuel resources continues to be released into the atmosphere, the peak concentration of atmospheric CO ₂ could exceed 1400 ppm by the year 2300 and the average global temperature could warm by 8 °C or more.	Completa	Sim	Sim