

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
DEPARTAMENTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA**

JULIANA HIROKO KOWATA

**UMA ABORDAGEM COMPUTACIONAL PARA
CONSTRUÇÃO DE MAPAS CONCEITUAIS A PARTIR
DE TEXTOS EM LÍNGUA PORTUGUESA DO BRASIL**

**VITÓRIA
2010**

JULIANA HIROKO KOWATA

**UMA ABORDAGEM COMPUTACIONAL PARA
CONSTRUÇÃO DE MAPAS CONCEITUAIS A PARTIR
DE TEXTOS EM LÍNGUA PORTUGUESA DO BRASIL**

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Departamento de Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Informática, na área de concentração em Informática na Educação.

Orientador: Prof. DSc. Davidson Cury.

Co-orientadora: Prof^a DSc. Maria Claudia Silva Boeres.

VITÓRIA

2010

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

K88a Kowata, Juliana Hiroko, 1975-
Uma abordagem computacional para construção de mapas
conceituais a partir de textos em língua portuguesa do Brasil /
Juliana Hiroko Kowata. – 2010.
102 f. : il.

Orientador: Davidson Cury.
Co-Orientadora: Maria Claudia Silva Boeres.
Dissertação (mestrado) – Universidade Federal do Espírito
Santo, Centro Tecnológico.

1. Ensino auxiliado por computador. 2. Lingüística -
Processamento de dados. 3. Mapas conceituais. I. Cury,
Davidson. II. Boeres, Maria Claudia Silva. III. Universidade
Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 004

JULIANA HIROKO KOWATA

**UMA ABORDAGEM COMPUTACIONAL PARA
CONSTRUÇÃO DE MAPAS CONCEITUAIS A PARTIR
DE TEXTOS EM LÍNGUA PORTUGUESA DO BRASIL**

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Departamento de Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Informática, na área de concentração de Informática na Educação.

Aprovada em 23 de agosto de 2010.

COMISSÃO EXAMINADORA

Prof. DSc. Davidson Cury
Universidade Federal do Espírito Santo
Presidente

Prof^a DSc. Maria Claudia da Silva Boeres
Universidade Federal do Espírito Santo

Prof. DSc. Alexandre Ibrahim Direne
Universidade Federal do Paraná

Prof. DSc. Orivaldo Lira Tavares
Universidade Federal do Espírito Santo

Prof^a DSc. Tânia Gava
Universidade Federal do Espírito Santo

Para Helmo...

*“Os olhos já não podem ver
Coisas que só o coração pode entender
Fundamental é mesmo o amor
É impossível ser feliz sozinho...”*

Tom Jobim

“Wave”

AGRADECIMENTOS

*“Quero, um dia, poder dizer às pessoas que nada foi em vão...
[...] que a vida é bela sim, e que eu sempre dei o melhor de mim... e que valeu a pena.”*

Mário Quintana
“Certezas”

Nesses vários meses, ora de intenso movimento, ora de inexplicável inércia, não pretendia muito mais do que respirar novos ares.

Chegada a esta parada, de tudo que passou, restou uma trilha. Trilha singela, mas consistente o suficiente para abrigar outros caminhantes. Esta é a minha contribuição para a Ciência: deixo a crítica para os críticos qualificados.

Uma longa jornada não prescinde de companhias. Muitos foram os que se dispuseram a apoiar a pesquisa, com sugestões, comentários e críticas, proporcionando-me momentos únicos de aprendizado. Para estas pessoas, registro meus sinceros agradecimentos:

Ao orientador desta pesquisa, professor *Davidson Cury (Dede)*, a quem sempre serei grata pelo modo como me acolheu no PPGI: sem receios em me ofertar oportunidades de aprendizado e sem reservas em depositar sua confiança em meu trabalho.

À co-orientadora, professora *Maria Cláudia da Silva Boeres (Claudinha)*, uma pessoa generosa, com quem aprendi muito sobre o ofício de pesquisadora e a quem respeito, por conservar-se humilde apesar de seu vasto conhecimento.

Ao professor *Ricardo Falbo*, pelo apoio decisivo para meu ingresso no PPGI e de cujas disciplinas ministradas derivaram meu interesse em pesquisar sobre a construção e a representação do conhecimento, fio condutor deste trabalho.

Ao professor *Crediné Menezes* por suas contribuições que auxiliaram na concepção deste trabalho e enriqueceram a pesquisa.

Ao professor *Orivaldo Lira Tavares*, que literalmente “abriu” para mim as portas do Laboratório de Informática da Educação (LIE) e com quem, desde então, tenho tido o privilégio de aprender, mesmo que em rápidas conversas pelos corredores do departamento.

Aos professores *Elias Silva Oliveira*, *Giancarlo Guizzardi* e *Sérgio Freitas* que apoiaram a pesquisa apontando caminhos para o trabalho com linguagem natural e representação de conhecimento, elementos que tornaram viáveis a construção teórica e a realização do estudo de caso.

Ao professor *Pablo Gamallo*, da Universidade de Santiago de Compostela, por sua gentileza em ceder valiosas informações sobre a ferramenta de *tagging* utilizada neste projeto.

Ao professor *Roger Nkambou*, da Université de Montreal, que cordialmente nos enviou materiais sobre suas mais recentes pesquisas sobre mapas conceituais e ontologias.

À *Cláudia Camerini Corrêa Pérez*, professora da Universidade Federal do Pampa e à *Renata Vieira*, professora da PUC-RS, que também já pesquisaram sobre o tema e que de modo prestativo enviaram dissertação sobre o assunto.

À *Luana Morellato*, colega do PPGI, que prontamente compartilhou sua dissertação de Mestrado, inédita à época, sobre reconhecimento de Sintagmas Nominais.

Ao *Helmo*, companheiro de todas as horas, com quem primeiro partilhei, partilho e partilharei minhas muitas dúvidas e poucas certezas, por sua paciência, contribuições e apoio incondicional.

E agradeço especialmente a Deus, por permitir que tudo isso fosse possível e sem quem não teria conseguido “chegar a lugar nenhum”.

Juliana H. Kowata.

Vitória, agosto de 2010.

*“Chega mais perto e contempla as palavras.
Cada uma
tem mil faces secretas sob a face neutra
e te pergunta, sem interesse pela resposta,
pobre ou terrível que lhe deres:
Trouxeste a chave?”*

*Repara:
ermas de melodia e conceito
elas se refugiaram na noite, as palavras.
Ainda úmidas e impregnadas de sono,
rolam num rio difícil e se transformam em
desprezo.”*

Carlos Drummond Andrade
“A procura da Poesia”

RESUMO

Um mapa conceitual é um recurso esquemático para representar e organizar um conjunto de significados em uma estrutura proposicional. No processo de aprendizagem, os atos de fazer e refazer mapas conceituais podem ser considerados como meios para identificar conceitos e seus significados, dando origem ao conhecimento de forma explícita. A aplicação de softwares para a construção de mapas conceituais estabeleceu um novo patamar na experiência de construção.

Nesta década, houve um crescente interesse da comunidade acadêmica pela aplicação de recursos computacionais para a construção de mapas conceituais a partir de um conjunto de dados previamente definido, por meio da extração de conceitos e de relações. Esta dissertação propõe uma abordagem focada no reconhecimento de elementos para a construção de mapas conceituais a partir de textos, e adicionalmente, expõe os resultados de experimentos conduzidos com a Língua Portuguesa do Brasil.

ABSTRACT

A concept map is a schematic resource to represent and organize a set of meanings in a propositional structure. In a learning process, the practice of making and remaking concept maps might be considered as an effort to find out concepts and their meanings, giving rise to the knowledge in an explicit way.

Computer aided programs can make the choice to use concept maps easier than before. Over the last few years, many basic functions in concept maps building, such as publishing and sharing, were facilitated by computer aided programs. An increasing interest in applying computational resources to automatically build concept maps from data sources by extracting concepts and linking words has emerged recently. In this thesis, we propose an approach focused on the recognition of concept maps' core elements in texts and, in addition, we expose the results of the experiment conducted in the Brazilian Portuguese Language.

LISTAS DE FIGURAS

Figura 1. Planejamento de Fases da Pesquisa Científica	22
Figura 2. Exemplo de Mapa Conceitual (Fonte: NOVAK; CAÑAS; 2006a)	37
Figura 3. Metamodelo da linguagem de Mapas Conceituais.....	38
Figura 4. Perspectivas e Classes para Análise de Abordagens de Construção de Mapas Conceituais.....	44
Figura 5. Análise das Abordagens Investigadas	53
Figura 6. Modelo de Processo para Construção de Mapas Conceituais a partir de Textos	60
Figura 7. Módulos do Protótipo e Atividades do Processo	69
Figura 8. Documento de entrada para o protótipo.....	75
Figura 9. Saída do Tokenizer	75
Figura 10. Trecho da Saída Intermediária do Tagger.....	76
Figura 11. Saída Definitiva do Tagger.....	76
Figura 12. Trecho da Saída do Chunker	77
Figura 13. Trecho da Saída do GraphBuilder – estrutura do grafo.....	77
Figura 14. Saída do GraphBuilder – imagem do grafo	78
Figura 15. Saída do CmapBuilder – Proposições do Mapa Conceitual.....	78
Figura 16. Saída do CmapBuilder – Imagem do Mapa Conceitual.....	79
Figura 17. Redução na transformação em grafo	84
Figura 18 - Aumento do número de tokens	85

LISTA DE TABELAS

Tabela 1. Perspectivas e Questões Norteadoras	44
Tabela 2. Caracterização das Abordagens para Construção de Mapas Conceituais com o uso de recursos computacionais	52
Tabela 3. Padrões Linguísticos para Reconhecimento de Sintagmas Nominais.....	65
Tabela 4. Exemplos de padrões linguísticos baseados em expressões regulares....	70
Tabela 5. Elementos reconhecidos pelo <i>Chunker</i>	72
Tabela 6. Correspondência de <i>chunk</i> para elemento de grafo.....	73
Tabela 7. Saídas produzidas pelos módulos de Text2Cmap	74
Tabela 8. Quantitativos e Médias dos Elementos Identificados	81
Tabela 9. Quantitativo de tokens do mapa conceitual comparado ao documento original.....	82
Tabela 10. Formação de grupos de acordo com a quantidade de tokens do documento original e do mapa conceitual construído	83
Tabela 11. <i>Chunks</i> não reconhecidos	84
Tabela 12. Detalhamento dos elementos dos grupos	86
Tabela 13. Exemplos de falhas no processo que ocasionam perda de tokens	88
Tabela 14. Síntese da Abordagem Proposta.....	93
Tabela 15. Ambiente de Desenvolvimento	101
Tabela 16. Relação de Padrões para reconhecimento de <i>tokens</i>	101
Tabela 17. Trecho do código para reconhecimento de <i>chunks</i>	101
Tabela 18. Regras de Compatibilidade entre Nós Candidatos.....	102
Tabela 19. Busca de Elemento Subsunçor a partir das folhas	103

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Mapas Conceituais como Instrumentos de Ressignificação	17
1.2	Problematização	19
1.2.1	Questões Norteadoras.....	20
1.2.2	Hipóteses.....	21
1.3	Metodologia.....	22
1.3.1	Planejamento.....	22
1.3.2	Referencial metodológico	23
1.4	Produção Científica	25
1.5	Estrutura da dissertação	25
2	MAPAS CONCEITUAIS	27
2.1	Breve Histórico	28
2.2	Fundamentação Teórica dos Mapas Conceituais	29
2.3	Definições	31
2.4	Organização de Conhecimento por meio de Mapas Conceituais.....	33
2.5	Representação de Conhecimento por meio de Mapas Conceituais.....	36
2.6	Considerações Finais.....	39
3	REVISÃO E PROPOSTA DE CLASSIFICAÇÃO DAS ABORDAGENS COMPUTACIONAIS PARA CONSTRUÇÃO DE MAPAS CONCEITUAIS.....	41
3.1	Motivações	42
3.2	Caracterização para Abordagens de Construção de Mapas Conceituais	43
3.2.1	Perspectiva baseada nos Objetivos da Pesquisa	45
3.2.2	Perspectiva baseada nas Fontes de Dados	46
3.2.3	Perspectiva baseada nos Mapas Construídos	50
3.3	Análise de Resultados.....	52

3.4	Considerações Finais.....	54
4	UMA ABORDAGEM COMPUTACIONAL PARA CONSTRUÇÃO DE MAPAS CONCEITUAIS A PARTIR DE TEXTOS EM LÍNGUA PORTUGUESA DO BRASIL	56
4.1	Especificação do Problema.....	57
4.2	Visão Geral da Abordagem	58
4.3	Atividades.....	61
4.3.1	Normalização de Texto.....	62
4.3.2	Segmentação de Texto.....	62
4.3.3	Tokenização	62
4.3.4	Análise Lexical.....	63
4.3.5	Reconhecimento de Elementos Centrais.....	63
4.3.6	Interpretação de Dependências.....	65
4.3.7	Construção de Mapas Conceituais	66
4.4	Considerações Finais.....	66
5	ESTUDO DE CASO	68
5.1	Protótipo Computacional	69
5.1.1	Módulo <i>Tokenizer</i>	69
5.1.2	Módulo <i>Tagger</i>	70
5.1.3	Módulo <i>Chunker</i>	71
5.1.4	Módulo <i>GraphBuilder</i>	72
5.1.5	Módulo <i>CmapBuilder</i>	74
5.2	Exemplo	74
5.3	Experimentação com Corpus	79
5.3.1	Mac-Morpho.....	80
5.3.2	Definição da Amostra	80
5.3.3	Resultados do Experimento.....	81
5.4	Análise dos Resultados Obtidos	86

6	CONCLUSÕES E TRABALHOS FUTUROS.....	90
6.1	Trabalhos Futuros	93
7	REFERÊNCIAS.....	95
	APÊNDICES.....	100

1 INTRODUÇÃO

As sociedades atuais são todas, pouco ou muito, sociedades da informação nas quais o desenvolvimento de tecnologias pode criar um ambiente cultural e educativo suscetível à diversificação das fontes do conhecimento e do saber (DELORS et al., 1999, p. 186).

Nas sociedades da informação, a ubiquidade e a importância das novas tecnologias dão ensejo ao desenvolvimento de habilidades e competências específicas, os quais deveriam estar presentes nos currículos escolares. A educação nesse cenário exige uma abordagem em que o componente tecnológico não pode ser ignorado (MERCADO, 2002, p.1) (DOWBOR, 2001, p.1), visto que suas inovações e o seu impacto revestem-se de uma dimensão não puramente tecnológica, mas essencialmente econômica e social (DELORS et al., 1999, p.186).

Vencer os desafios tecnológicos, portanto, passou a ser motivo de inquietação em sistemas educativos no mundo inteiro (DELORS et al., 1999, p. 168). O desenvolvimento de currículos que permitam o uso efetivo da tecnologia em um ambiente educacional é uma escolha educativa que reflete uma opção de sociedade pela inclusão digital.

Dessa maneira, ao recomendar o uso efetivo da tecnologia em ambientes educacionais, a UNESCO estabelece relação entre educação, tecnologia e bem estar social, afirmando que “para viver, aprender e trabalhar bem em uma sociedade cada vez mais complexa, rica em informação e baseada em conhecimento, alunos e professores **devem usar a tecnologia de forma efetiva** [...]” (UNESCO, 2009, p.1, grifo nosso). Além disso, a UNESCO também explicita os benefícios do uso da tecnologia na formação de estudantes, quando declara que:

[...] o uso da tecnologia pode permitir que os alunos tornem-se usuários qualificados da informação; audiência crítica da informação; solucionadores de problemas e tomadores de decisão; usuários criativos e efetivos de ferramentas de produtividade; comunicadores, colaboradores, editores e produtores de informação; cidadãos informados, responsáveis e capazes de contribuir com a sociedade (UNESCO, 2009, p.1)

A partir das questões expostas, entendemos que as experiências para construção de conhecimento no processo de ensino e aprendizagem podem e devem ser aprimoradas por meio do emprego de recursos tecnológicos adequados, desde que o domínio do técnico não ocorra dissociado do pedagógico e que novas tecnologias convivam em harmonia com as tecnologias existentes. Ademais, as facilidades técnicas oferecidas pelos computadores possibilitam a exploração de um leque extenso de ações pedagógicas, proporcionando maior diversidade de atividades para professores e alunos (SILVA, 2005, p.33) (VALENTE, 2002, p.23).

Cabe à tecnologia o papel de apoio, principalmente, no espaço da construção do conhecimento, criando condições para que o aluno, sujeito ativo da aprendizagem, aprenda ao fazer, levantar e testar ideias, experimentar, aplicar conhecimentos e representar o pensamento. Logo, a integração de diferentes tecnologias no currículo escolar – prática que fomenta o surgimento de novas maneiras de ensinar e aprender – fundamenta-se tanto na introdução de novas tecnologias quanto na ressignificação das chamadas tecnologias convencionais, cabendo ao professor decidir quais veículos e linguagens privilegiar, levando os estudantes à melhor compreensão dos conteúdos (SILVA, 2005, p. 36).

Particularmente, interessamo-nos pelo uso de recursos de Tecnologia da Informação e Comunicação no apoio às práticas educacionais voltadas para a aprendizagem significativa. De forma mais específica, pesquisamos o uso de mapas conceituais como recursos de apoio ao processo de ensino e aprendizagem e como instrumentos para ressignificação de conteúdos expressos por meio de tecnologias convencionais, notadamente, pela escrita.

1.1 Mapas Conceituais como Instrumentos de Ressignificação

Os mapas conceituais têm por objetivo a representação de relações significativas entre conceitos na forma de proposições. Uma proposição consiste em dois ou mais termos conceituais ligados por palavras de modo a formar uma unidade semântica (NOVAK; GOWIN, 1984).

Os elementos que constituem os mapas conceituais são denominados de conceitos e ligações. Os conceitos normalmente são representados por círculos ou retângulos contendo rótulos internos e as proposições se formam por meio de segmentos de reta nomeados entre dois conceitos (NOVAK; CAÑAS, 2006a).

Embora lápis e papel sejam suficientes para a construção de mapas conceituais, as interações sem o apoio de ferramentas computacionais são, em geral, complexas e exaustivas. Portanto, a introdução de programas de software destinados ao desenho de mapas conceituais estabeleceu um novo patamar de qualidade na experiência de construção. A popularização da Internet possibilitou um novo salto qualitativo, permitindo o surgimento de uma geração de ferramentas verdadeiramente colaborativas, ampliando as perspectivas para uso de mapas conceituais (NOVAK; CAÑAS, 2006a).

Inicialmente concebidos como instrumentos para representar a compreensão conceitual de crianças nos anos iniciais de sua vida escolar (NOVAK; CAÑAS, 2006b), atualmente, os mapas conceituais são empregados em diversas áreas, apoiando o desenvolvimento de competências profissionais, as atividades de modelagem de sistemas e processos, a condução de avaliações de aprendizagem e a organização de pesquisas científicas (DALEY et al., 2008).

Ao longo da última década, houve um crescente interesse da comunidade acadêmica pela aplicação de recursos computacionais para a construção de mapas conceituais a partir de um conjunto de dados previamente definido. As pesquisas científicas nesta área concentram-se na independência do auxílio humano no processo de construção de mapas conceituais, buscando por alternativas para a superação das dificuldades de construção a partir do “zero” (CHANG et al., 2008) e para a redução de tempo e esforços despendidos na aquisição de conhecimento (LEE et al., 2009) (TSENG et al., 2007), sobretudo, em domínios extensos (VALERIO; LEAKE, 2006) e dependentes de especialistas (CHANG et al., 2008).

As principais abordagens para a construção de mapas conceituais a partir de conteúdos presentes em fontes de dados são motivadas por questões de ordem educacional: como instrumentos no processo de ensino e aprendizagem (ALVES; PEREIRA; CARDOSO, 2001) (CLARIANA; KOUL, 2004) (LAU et al., 2008), na

definição de estratégias pedagógicas (CHANG et al., 2008) (CHEN et al., 2008) (BAI; CHEN, 2008) (LEE et al., 2009) (TSENG et al., 2007) e na avaliação de desempenho de estudantes (GRAUDINA; GRUNDSPENKIS, 2008) (VILLALON; CALVO, 2008). Outras pesquisas são inspiradas pela necessidade de: representar conhecimentos sob uma perspectiva favorável à análise de especialistas de domínio (KUMAZAWA et al., 2009), como forma de aquisição de conhecimentos de domínio (ZOUAQ; NKAMBOU, 2008, 2009) (PÉREZ; VIEIRA, 2004, 2005), para resumir o conteúdo de extensas bibliotecas digitais (RICHARDSON; FOX, 2007) e para exploração e compartilhamento de documentos digitais (GAINES; SHAW, 1994) (VALERIO; LEAKE, 2006).

O conjunto das pesquisas citadas revela o amplo espectro de finalidades que estimulam a investigação com foco na utilização de recursos computacionais na construção de mapas conceituais a partir de conteúdos existentes.

Considerando as possibilidades de investigação, definimos para a nossa pesquisa o viés educacional, pautado na aplicação de recursos de Tecnologia da Informação e Comunicação no apoio às práticas educacionais voltadas para a aprendizagem significativa, designadamente, às práticas apoiadas pelo uso de mapas conceituais como recursos de apoio ao processo de ensino e aprendizagem.

Nesta pesquisa, portanto, delimitamos o problema ao uso de recursos tecnológicos adequados com o intuito de promover a ressignificação de conteúdos estáticos, essencialmente na forma de textos, em conteúdos dinâmicos, na forma de mapas conceituais.

1.2 Problematização

Os esforços presentes nesta pesquisa, bem como o objeto de sua investigação, devem ser entendidos à luz da Educação. São, antes de serem exercícios computacionais, ações que buscam alternativas favoráveis à integração de diferentes tecnologias e mídias a fim de criar oportunidades para a descoberta de maneiras de ensinar, aprender e desenvolver o currículo.

Partimos do pressuposto de que a construção, avaliação e evolução dos mapas conceituais são mecanismos que instigam o aluno a tornar-se sujeito ativo do processo de aprendizagem (NOVAK, 2000). Consideramos que as dinâmicas com mapas conceituais permitem que o currículo seja tecido na ação colaborativa entre professores e alunos. Ademais, presumimos ser possível o uso de recursos computacionais na construção de mapas conceituais que permitam a ressignificação de conteúdos textuais.

Pelas razões reveladas, restringimos o universo desta pesquisa às abordagens para construção de mapas conceituais a partir de fontes de dados não estruturadas, apoiada no uso de recursos computacionais. As abordagens baseadas em fontes não estruturadas utilizam documentos escritos em linguagem natural como fonte primária para a extração de elementos do mapa conceitual. São exemplos de fontes de dados não estruturadas: textos produzidos por especialistas de domínio (ALVES; PEREIRA; CARDOSO, 2001) (GAINES; SHAW, 1994) (VALERIO; LEAKE, 2006) (ZOUAQ; NKAMBOU, 2008, 2009), textos produzidos por estudantes em resposta a questionários dissertativos (CHANG et al., 2008) ou temas propostos (CLARIANA; KOUL, 2004), artigos acadêmicos (CHEN et al., 2008), resumos de teses e dissertações eletrônicas (RICHARDSON; FOX, 2007), mensagens postadas em fóruns de discussão (LAU ET AL., 2008), textos jornalísticos e didáticos (PÉREZ; VIEIRA, 2004, 2005).

Orientados pela ideia da construção automática de mapas conceituais a partir de textos e inspirados pelas recomendações da UNESCO que incentivam a promoção da diversidade linguística e de recursos de aprendizagem em língua materna (UNESCO, 2009b, p.14), estabelecemos como objetivo de investigação desta pesquisa a **definição de uma abordagem computacional que permita a construção de mapas conceituais a partir de textos em Língua Portuguesa do Brasil.**

1.2.1 Questões Norteadoras

De acordo com Silva e Menezes (2001, p.19), a pesquisa científica visa encontrar respostas para as indagações propostas. Logo, as ações da pesquisa concentram-

se em descobrir as respostas para as ditas indagações. A seguir, estão listadas as principais questões que foram investigadas por esta pesquisa:

- i. É possível construir mapas conceituais a partir de textos em linguagem natural?
- ii. Como extrair os elementos para a construção de mapas conceituais a partir de textos sem que haja perda semântica no processo?
- iii. Quais são os métodos e as técnicas utilizadas por pesquisas similares?
- iv. É possível definir algum mapeamento entre constituintes de mapas conceituais e constituintes de uma sentença em linguagem natural?
- v. Como os mapas conceituais podem ser construídos a partir de textos em conformidade com a Teoria da Aprendizagem Significativa?

1.2.2 Hipóteses

A partir do nosso interesse em despender esforços no entendimento mais profundo do processo de construção de mapas conceituais a partir de textos, procuramos, ao longo da pesquisa, por resultados que nos permitiram reforçar ou descartar as seguintes hipóteses:

- i. O uso de métodos e técnicas de processamento de linguagem natural é indicado para endereçar o problema;
- ii. Elementos constituintes de mapas conceituais podem ser correlacionados aos elementos constituintes de uma sentença em linguagem natural;
- iii. A gramática sintagmática pode ser utilizada como meio para a identificação de elementos constituintes de mapas conceituais; e
- iv. Grafos são estruturas de dados adequadas para serem utilizadas no processo de construção de mapas conceituais a partir de textos.

1.3 Metodologia

A Associação Brasileira de Normas Técnicas (2005) define a dissertação como:

[...] o resultado de um trabalho experimental ou exposição de um estudo científico retrospectivo, de tema único e bem delimitado em sua extensão, **com o objetivo de reunir, analisar e interpretar informações. Deve evidenciar o conhecimento de literatura existente** sobre o assunto [...] (ABNT NBR 14724, 2005, grifo nosso).

Portanto, o conteúdo esperado de uma dissertação – produto final de nossa pesquisa – é composto por um “trabalho experimental ou exposição de estudo científico”, acrescido de “análise e interpretação das informações”, conduzido a partir de “conhecimento de literatura existente”, nas palavras de Eco (2007), com um ponto de apoio como “anões às costas de gigantes”.

Considerando tais recomendações, descrevemos, a seguir, o planejamento desta pesquisa.

1.3.1 Planejamento

Planejamos a pesquisa em três fases distintas, de acordo com os resultados esperados em cada um dos três momentos: Concepção, Elaboração e Consolidação. As fases foram orquestradas de modo que os resultados produzidos em uma fase fossem insumos para a fase seguinte, conforme ilustrado na Figura 1.

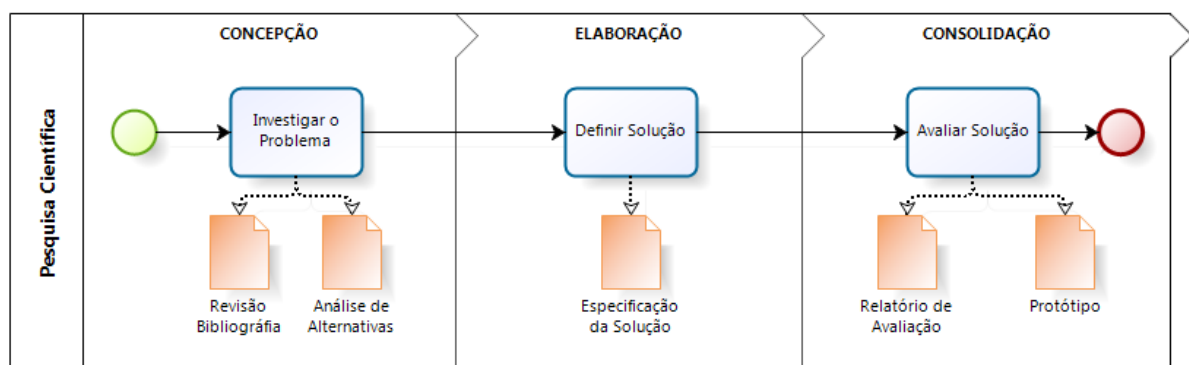


Figura 1. Planejamento de Fases da Pesquisa Científica

O conjunto das saídas de cada fase foi planejado para subsidiar a escrita dos capítulos desta dissertação.

Na fase de **Concepção**, definimos um conjunto de hipóteses e, adicionalmente, propusemos como resultado esperado a revisão bibliográfica de abordagens similares e a comparação destas, com o intuito de produzir um referencial que admitisse a análise das melhores alternativas de solução. Na ausência de parâmetros adequados para a comparação das abordagens, optamos por definir uma que contivesse os principais pontos que comportassem a caracterização das abordagens investigadas. Os estudos dessa fase constituíram a base para a publicação de um artigo no XX Simpósio Brasileiro de Informática na Educação¹.

Na fase de **Elaboração**, propusemos como resultado esperado a especificação de uma solução que consentisse a construção automática de mapas conceituais a partir de textos em Língua Portuguesa do Brasil. Para isso, utilizamos como insumo os resultados produzidos na fase de Concepção, que auxiliaram na seleção dos métodos e técnicas mais adequados aos objetivos desta pesquisa e confirmar algumas das hipóteses levantadas sobre o uso de métodos linguísticos. Esses estudos foram temas de duas publicações na quarta edição do “*Concept Mapping Conference*”.

Com a solução especificada, na fase de **Consolidação**, propusemos como resultado esperado sua avaliação por meio de estudo de caso com corpus² da Língua Portuguesa. Definimos como pré-requisito da avaliação a implementação da solução na forma de um protótipo.

1.3.2 Referencial metodológico

O entendimento desta pesquisa segundo o ponto de vista de seus objetivos requer dupla designação: **pesquisa descritiva** e **exploratória**.

Ao entendê-la como uma **pesquisa descritiva** (GIL, 1999) consideramos que o resultado esperado da solução do problema definido se traduziu, inevitavelmente, na especificação de uma solução para o problema da construção de mapas conceituais segundo as premissas propostas. A solução, portanto, detalha ações para

¹ Para maiores detalhes sobre as produções decorrentes desta pesquisa, consultar a Subseção 1.4.

² Corpus é um subconjunto de uma coleção de textos eletrônicos em formato padronizado com certas convenções relacionadas ao conteúdo e selecionadas sem restrições rigorosas, construído em conformidade com critérios explícitos de design para um propósito específico (ATKINS, CLEAR, OSTLER, 1999).

determinar as relações entre a linguagem natural e os elementos de mapas conceituais.

Podemos, também, considerá-la como uma **pesquisa exploratória** (GIL, 1999), na medida em que significativa parcela da pesquisa concentrou-se no estudo bibliográfico, buscando uma maior aproximação com as questões centrais do problema. Durante o desenvolvimento da pesquisa, constatou-se a dificuldade de comparar abordagens diversas, o que resultou na acentuação do estudo de caráter exploratório, de que resultaram parâmetros para tabulação e análise de diferentes pesquisas.

Do ponto de vista da forma de abordagem do problema, Silva e Menezes (2001) consideram que uma pesquisa pode abordar o problema de forma quantitativa ou qualitativa. A abordagem qualitativa utiliza-se da descrição, com foco no processo e seu significado, materializado de forma descritiva. A abordagem quantitativa considera a quantificação como seu principal instrumento para classificação e análise das informações. Novamente, a pesquisa descrita por esta dissertação apresenta duas formas de abordagem: **qualitativa e quantitativa**.

Decorrente do problema de natureza descritiva, tratamos a solução por meio de uma abordagem qualitativa, focada na proposta de uma solução na forma de um processo. Entretanto, os resultados produzidos por essa abordagem qualitativa, quando submetidos a um estudo de caso, foram abordados de forma quantitativa, por meio de coleta e análise de dados.

Do ponto de vista dos procedimentos técnicos empregados na pesquisa, consideramos importante lançar mão dos seguintes recursos:

- i. a pesquisa bibliográfica, na fase de concepção;
- ii. a pesquisa experimental, na fase de elaboração;
- iii. o estudo de caso, na fase de consolidação.

Quanto ao método, adotamos prioritariamente, o método indutivo. As constatações particulares dos fenômenos da linguagem, em casos concretos, levaram à elaboração de generalizações para construção do mapa conceitual.

1.4 Produção Científica

Além desta dissertação, os seguintes artigos foram frutos desta pesquisa:

- i. **“Caracterização das Abordagens para Construção (Semi) Automática de Mapas Conceituais”**, publicado no XX Simpósio Brasileiro de Informática na Educação, 2009;
- ii. **“Concept Maps Core Elements Candidates Recognition from Text”**, aceito para publicação no 4th Concept Mapping Conference, 2010.
- iii. **“A Review of Semi-Automatic Approaches to Build Concept Maps”**, aceito para publicação no 4th Concept Mapping Conference, 2010.
- iv. **“Uma abordagem computacional para a construção de mapas conceituais a partir de textos em Língua Portuguesa do Brasil”**, submetido ao XXI Simpósio Brasileiro de Informática na Educação, 2010.

1.5 Estrutura da dissertação

Esta dissertação está estruturada na seguinte forma:

Neste Capítulo, de caráter introdutório, apresentamos uma visão geral da dissertação, destacando o contexto, as motivações, o problema e a metodologia de trabalho observada pela pesquisa.

No Capítulo 2, estabelecemos as bases para a discussão sobre Mapas Conceituais, propondo subsídios à compreensão do tema sob a ótica histórica e teórica. Em um segundo momento, destacamos os mecanismos que permitem o uso de mapas conceituais na organização e na representação de conhecimento.

No Capítulo 3, conduzimos uma revisão bibliográfica das abordagens computacionais para a construção de mapas conceituais. Propomos a análise das

abordagens similares à nossa pesquisa, a partir de um conjunto de características, propiciando o entendimento do estado da arte nesta área de pesquisa a partir de um conjunto de informações quantitativas.

No Capítulo 4, cerne da pesquisa, descrevemos a solução proposta para a construção de mapas conceituais a partir de textos em Língua Portuguesa do Brasil, por meio do uso de recursos computacionais. Especificamos a solução como um processo no qual as atividades são coordenadas para a resolução do problema de construção de mapas a partir de textos.

No Capítulo 5, relatamos os resultados do estudo de caso orientado para a avaliação da solução proposta, por meio do desenvolvimento de um protótipo computacional e da coleta de dados a partir de uma amostra extraída do corpus linguístico Mac-Morpho, recurso em português do Brasil.

No Capítulo 6, apresentamos as considerações finais e os principais direcionamentos para pesquisas futuras.

2 MAPAS CONCEITUAIS

*“[...] Giro um simples compasso
E num círculo eu faço o mundo...”*

*Toquinho; Vinicius de Moraes; G. Morra; M. Fabrizio
“Aquarela”*

O objetivo deste capítulo é pontuar as características singulares dos “mapas conceituais”, tema capital para o desenvolvimento da presente dissertação.

Iniciamos pela discussão das origens dos mapas conceituais. Delineamos brevemente, na Seção 2.1, as circunstâncias e as motivações que levaram ao seu surgimento e na Seção 2.2, abordamos a Teoria da Aprendizagem Significativa, fundamento teórico dos mapas conceituais, ressaltando como a teoria estabeleceu os principais elementos e a dinâmica de construção de mapas conceituais.

Em seguida, aprofundamos no entendimento de mapas conceituais, apresentando, na Seção 2.3, a definição de mapa conceitual, interpretado no contexto das linguagens de especificação conceitual, e introduzimos informações importantes para destacar a distinção entre o mapa conceitual na acepção de modelo, de especificação e também da linguagem de especificação.

Na Seção 2.4, descrevemos mapas conceituais como meio para organização do conhecimento. Para tanto, enumeramos os elementos que permitem a organização do conhecimento por meio de um mapa conceitual, enfatizando que na compreensão destes elementos reside a chave para a construção de mapas conceituais.

Na Seção 2.5, abordamos o mapa conceitual sob o viés da representação de conhecimento. Descrevemos os formalismos que devem ser observados na representação do conhecimento por meio de mapas conceituais, com o objetivo de fornecer elementos para a construção de mapas conceituais bem formados.

Encerramos este capítulo na Seção 2.6 no qual apresentamos nossas considerações finais.

2.1 Breve Histórico

Entre as década de 1960 e 1970, o educador norte-americano Joseph Novak conduziu um programa de pesquisa para compreender a habilidade em crianças de seis a oito anos na aquisição de conceitos abstratos de ciências e os efeitos desse primeiro aprendizado ao longo de suas vidas escolares (NOVAK, 2000, p. 27). Ao investigar sobre os fatores limitadores do aprendizado, o programa de pesquisa atuou em várias frentes de trabalho, desde a elaboração de recursos didáticos na forma de tutoriais em áudio à coleta de evidências que pudessem de alguma forma, revelar como a estrutura cognitiva da criança era alterada no processo de aprendizagem.

Entrevistando crianças, antes e após o uso dos recursos de áudio no processo de aprendizagem, a pesquisa acumulou, inevitavelmente, um grande volume de registros. A princípio, os registros eram transcritos a partir de cada entrevista, com intuito de prover dados para análise posterior. Entretanto, o método de transcrição apresentou-se inadequado, sobretudo, por não favorecer a observação explícita das mudanças na compreensão conceitual das crianças (NOVAK, 2006b, p. 4), como pode ser constatado nas reminiscências do próprio Novak:

[...] Era extremamente difícil analisar estes registros e encontrar padrões ou regularidades que nos pudessem ajudar a compreender o como e o porquê das crianças aprenderem ou não o novo assunto (NOVAK, 2000, p. 27, tradução nossa).

[...] À medida que transcrevíamos as fitas, nós pudemos observar que as proposições usadas pelos estudantes normalmente melhoravam em relevância, número e qualidade, **mas ainda era difícil observar especificamente como suas estruturas cognitivas estavam mudando** (NOVAK, 2006b, p. 4, tradução nossa, grifo nosso).

Dada à fragilidade do método de transcrição, os pesquisadores enveredaram na busca por alternativas que pudessem levar ao aprimoramento na análise das entrevistas. Dentre as opções examinadas, ganharam proeminência as representações em forma de diagramas.

Ao revisitar a teoria cognitiva de Ausubel, a investigação, já direcionada para a abordagem diagramática, lançou as ideias seminais da ferramenta que hoje conhecemos como “mapa conceitual” (Novak, 2006a, p. 4), como relatado no extrato a seguir:

[...] Quando trabalhamos com a teoria da aprendizagem significativa de Ausubel [...], decidimos examinar transcrições de entrevistas para estudarmos palavras e proposições conceituais dadas pelos alunos, pois estas iriam indicar conhecimentos anteriores e conhecimentos pós-instrução. **Depois de tentarmos várias formas de organizar as palavras e proposições conceituais, surgiu ao meu grupo de investigação a ideia de mapa conceitual** (NOVAK, 2000, p. 27, tradução nossa, grifo nosso).

Desde então, os mapas conceituais são usados em uma diversidade de áreas, tendo maior destaque nas aplicações de fins educativos.

2.2 Fundamentação Teórica dos Mapas Conceituais

A *Teoria da Aprendizagem Significativa* proposta nos anos 60 pelo psicólogo norte-americano David Ausubel é considerada a perspectiva cognitiva clássica da aprendizagem significativa (MOREIRA, 2006, p.1). A abordagem de Ausubel destina-se, em primeiro lugar, à aprendizagem cognitiva, ou seja, à aquisição de conhecimentos, mas também apresenta importantes implicações na aprendizagem afetiva e motora (NOVAK, 2000, p. 51).

O cerne da teoria é a **aprendizagem significativa**, contraponto da aprendizagem mecânica. Ausubel define a aprendizagem significativa como uma opção de aprendizagem feita por um indivíduo, na qual ele aprende relacionando novos conhecimentos às proposições e conceitos conhecidos, e privilegia uma interação cognitiva “não arbitrária” e “não literal” (MOREIRA, 2006, p.1). A aprendizagem mecânica, por sua vez, demanda a memorização e resulta na incorporação arbitrária da nova informação à estrutura de conhecimentos de um indivíduo (NOVAK; GOWIN, 1984, p. 23).

A aprendizagem significativa requer um esforço maior do indivíduo em processo de aprendizagem, impelindo-o a buscar em seus conhecimentos pregressos os conceitos mais adequados aos quais serão ligadas as novas informações. Os

conceitos existentes, especificamente relevantes, que serão utilizados como âncoras para as novas informações são denominados de subsunçores. Deste modo, o conhecimento preexistente, ou seja, a estrutura cognitiva é o arcabouço para o aprendizado, levando à afirmação de Ausubel de que: **“o fator mais importante que influencia a aprendizagem é o que o aluno já sabe”** (apud NOVAK; GOWIN, 1984, p. 56, grifo nosso).

Novak e Gowin (1984, p. 23) consideram que todo o trabalho de pesquisa com mapas conceituais têm nos conceitos e nas proposições os elementos centrais na estrutura do conhecimento e na construção de significado. Declaram, categoricamente, que “a melhor teoria da aprendizagem que se foca nos conceitos e na aprendizagem proposicional como base em que assenta a construção pelos indivíduos dos seus próprios significados idiossincrásicos é a [...] proposta por David Ausubel” (NOVAK; GOWIN, 1984, p. 23). E, em obra posterior, ao justificar sua predileção pela Teoria da Aprendizagem Significativa, Novak afirma que embora muitas novas teorias tenham tentado explicar a aprendizagem humana, ele ainda considera “a Teoria de Ausubel, talvez com algumas alterações e novas ideias [...] a mais abrangente e consistente” (NOVAK 2000, p. 5), “a integração de muitas observações sobre a aprendizagem numa teoria única e coerente” (NOVAK 2000, p. 51).

Nota-se, portanto, que a aderência teórica dos Mapas Conceituais à Teoria da Aprendizagem de Ausubel não é fruto do mero acaso. Ao contrário, Novak propositadamente arquitetou a construção de Mapas Conceituais em três pilares da Teoria de Ausubel: a) o desenvolvimento de novos significados como construções sobre conceitos e proposições importantes e já conhecidos; b) a estrutura cognitiva, como um espaço organizado hierarquicamente, com conceitos mais gerais e mais inclusivos no topo da hierarquia e conceitos mais específicos e menos inclusivos ancorados nos conceitos mais gerais; c) a aprendizagem significativa, definindo relacionamentos entre conceitos mais explícitos, mais precisos e melhor integrados com outros conceitos e proposições.

Das premissas teóricas, derivaram-se as características estruturais e a dinâmica de construção dos Mapas Conceituais: a) organização semi-hierárquica dos conceitos, em que os conceitos específicos “ancoram-se” aos conceitos gerais; b) definição de

conceitos por meio de um número reduzido de palavras que definem “uma regularidade percebida em objetos e eventos”; c) a relação entre dois conceitos rotulada de modo que uma proposição seja elaborada.

Portanto, a compreensão da Teoria da Aprendizagem Significativa de Ausubel é o fundamento inicial para a fluência na técnica para construção de um mapa conceitual.

2.3 Definições

Nesta seção, procuramos definir o termo “mapa conceitual” no contexto das linguagens de modelagem conceitual.

Para Guizzardi (2005, p. 2), modelo é a entidade abstrata, instância de uma dada porção da realidade articulada por meio de um conjunto de conceitos elencados a partir de uma perspectiva específica de um dado domínio. A perspectiva de observação, denominada de “conceitualização”, auxilia na definição de conceitos e circunstâncias considerados relevantes na abordagem do domínio, estabelecendo um universo de discurso. O modelo quando representado como um artefato concreto apoia a comunicação, o aprendizado e a análise de aspectos relevantes do domínio (GUIZZARDI, 2005 p. XI).

O modelo quando representado por meio de um artefato concreto deve ser expresso em uma linguagem adequada. Tal linguagem é denominada de “linguagem de especificação”, também conhecida como “linguagem de modelagem”, que determina os modelos de especificação, ou seja, as representações conceituais gramaticalmente válidas para um modelo de domínio. No caso de mapas conceituais, o modelo reflete as escolhas dos conceitos e das relações representativas do domínio, cuja conceitualização é determinada por meio da “questão focal”. O modelo, como artefato concreto, é representado por meio de um diagrama, observando as especificações válidas para o modelo, instanciadas da linguagem de especificação, cujo alfabeto é formado por símbolos pictográficos, especificamente por retângulos e segmentos de reta, ambos devidamente rotulados.

O termo “mapa conceitual” é utilizado indiscriminadamente na literatura, referindo-se ora a um processo de construção, ora a uma técnica, ora a um diagrama e também a uma linguagem de especificação. A variedade de usos do termo dificulta a sua compreensão, levando muitas vezes à criação de novas acepções a partir das existentes. A seguir, esclarecemos como esta dissertação entende o termo “mapa(s) conceitual(ais)”.

Novak (2006b, p.1), o pioneiro, define mapas conceituais como “ferramentas gráficas para organização e representação de conhecimento”. Da afirmação, não se pode observar diretamente se o termo “ferramenta” faz menção ao modelo em si, à especificação do modelo ou à linguagem de especificação. Entretanto, no decurso de sua explanação, ao orientar sobre a seleção de conceitos, encontramos elementos que associam os mapas conceituais a uma conceitualização e quando discorre sobre as regras a serem observadas para a construção de mapas bem formados, associamos que suas observações são aderentes ao modelo, descrito de acordo com uma especificação válida instanciada de uma linguagem de especificação. Portanto, entendemos que de modo implícito, o autor se refere aos mapas conceituais como modelos, representados por meio de artefatos concretos.

Outro emprego usual do termo “mapas conceituais” refere-se à lógica subjacente aos mapas conceituais, ou seja, às regras observadas em sua construção. Nesta acepção, há a aderência com a definição de especificação, no qual consta a abordagem de “mapas conceituais” a partir da ótica de uma especificação válida, instanciada de uma linguagem de especificação. Neste caso, a linguagem e o modelo de especificação são anteriores ao modelo de domínio.

Ao longo desta dissertação, o uso do termo “mapa(s) conceitual(ais)” estará associado ao modelo de domínio, representado pelo artefato concreto, na forma de diagrama, que observa um modelo de especificação válido, derivado da linguagem de especificação. Este será o sentido padrão, salvo em ocasiões que mencionemos de forma explícita outro uso para o termo.

Como modelo, o objetivo primário de um mapa conceitual é o de representar as conceitualizações de um dado domínio, na forma de proposições (NOVAK, 1984, p. 31).

Observado como um modelo concreto e diagramático, o mapa conceitual contém a conceitualização de um dado domínio expressa essencialmente por proposições e representadas por meio de um conjunto de segmentos de reta, ou arcos, que ligam palavras delimitadas dentro de retângulos ou círculos.

Apreende-se, portanto, que o entendimento de mapas conceituais, enquanto modelo, passa pela compreensão de como é realizada a conceitualização do domínio – com a observação das nuances que os distinguem de outros diagramas em redes –, quanto pela linguagem utilizada na representação do modelo.

2.4 Organização de Conhecimento por meio de Mapas Conceituais

A conceitualização nos mapas conceituais é organizada por meio de conceitos e de relações, expressas por meio de proposições.

O conceito em um mapa conceitual deve ser compreendido como “uma regularidade apreendida dos acontecimentos ou objetos, ou registro dos acontecimentos ou objetos, designados por um símbolo” (NOVAK 2000, p. 21, tradução nossa).

Ao apresentar a definição de conceito, Novak estabelece o ponto de vista clássico em que se destacam as referências aos “objetos” e aos “acontecimentos”. Ao objeto, Novak (1984, p. 45) associa uma existência, mentalmente acessível por meio de uma “imagem”, e fisicamente realizada por meio de átomos e moléculas. São exemplos de objetos: cadeiras, gatos, pedras, etc. Os acontecimentos descrevem fenômenos de “trocas de energia”, tais como leitura, corrida, pensamento. Novak adverte que nem sempre os objetos podem ser materializados, pois os conceitos podem ser derivados de observações de registros, e de descrições que levam à formação de um conceito (Novak 2000, p. 21) não associado a uma entidade do mundo que conhecemos, ou ao nosso tempo, tais como, saci, curupira, dinossauro.

A codificação do conceito ocorre por meio de símbolos, normalmente palavras ou sinais. Por exemplo, a palavra “cadeira”, composta pela escrita e pelo som, remete a

uma imagem qualquer de “cadeira” e ao entendimento de que se trata de um artefato com pernas que sustentam um assento.

A definição clássica de conceito, proposta em 1984 (NOVAK; GOWIN, 1984, p, 41) não considera que nomes próprios de pessoas, localidades e coisas específicas sejam símbolos representativos de conceitos, nem símbolos que não remetam a imagens mentais. Por isso, esta concepção abarca uma série de limitações, sobretudo para mapas conceituais situados em domínios biográficos, geográficos, científicos. Não seria possível, por exemplo, criar um mapa conceitual sobre a vida de “Castro Alves”, a construção de “Brasília” ou sobre o “Humanismo”. Em obra recente, Novak (2000, p. 21) deixa clara a existência de conceitos primários e secundários, como proposto por Ausubel. Os conceitos primários, originados da observação direta dos objetos, acontecimentos – ou registros destes – e os conceitos secundários, frutos do processo de amadurecimento, de assimilação de conceitos e “sem exemplares visíveis”, tais como amor, história, capitalismo (NOVAK, 2000, p. 41). Em crítica à definição hermética de “conceito” em Mapas Conceituais, Dutra e outros (1994, p.1) assumem a defesa de um posicionamento dependente do sujeito do processo de construção do conhecimento:

A concepção piagetiana de conceito situa a representação através de mapas conceituais num paradigma diferente daquele estabelecido pela visão cognitivista de Ausubel e Novak. [...] as palavras [escolhidas para serem conceitos] dos mapas conceituais (em geral substantivo) não são necessariamente, na perspectiva do sujeito, os conceitos. (DUTRA et al., 2004, p. 1)

Deixamos clara nossa posição em adotar as observações feitas por Dutra e outros sobre a necessidade de que a conexão entre o conceito e as palavras que o designam seja adequada na perspectiva do sujeito da ação.

Cientes da impossibilidade de esgotar a questão em um único capítulo, arbitramos, para fins desta pesquisa, que “conceito” é uma representação abstrata de certos aspectos de entidades que existem em um domínio, incluindo ainda, sua designação genérica ou de um exemplar particular desta entidade ou ainda, de uma característica passível de quantificação ou qualificação.

Com a acepção de “conceito” demarcada, é possível introduzir o termo “proposição”. A proposição é uma unidade semântica, ou seja, uma unidade de significado, composta pela relação entre dois conceitos, cuja assinatura é formada por um

conjunto de palavras denominadas de palavras de ligação³. Nas palavras de Novak e Cañas, “proposições são declarações sobre algum objeto ou evento no universo, seja de ocorrência natural ou construída” (NOVAK; CAÑAS, 2006a, p.1, tradução nossa).

O mapa conceitual normalmente é arranjado em torno de uma questão focal, levando ao destaque certo número de conceitos. A questão focal é uma situação ou evento que se tenta compreender por meio da organização de ideias na forma de mapa conceitual e que provê o contexto para a elaboração deste mesmo mapa (NOVAK; GOWIN, 2006a, p.2).

Outro aspecto que merece destaque em Mapas Conceituais é a existência de um traçado hierárquico entre os conceitos, situando os conceitos mais genéricos e inclusivos no topo da hierarquia. Adicionalmente, podem ocorrer hierarquias de mapas, quando um conceito necessita ser trabalhado em um Mapa Conceitual específico (NOVAK, 2000, p. 218). Deve-se observar que a organização hierárquica entre conceitos e mesmo entre mapas é uma variável dependente do contexto de elaboração do Mapa Conceitual. Os mesmos conceitos poderão ser vistos em diferentes posições hierárquicas em função dos enfoques adotados (NOVAK, 2000, p. 20).

A adoção da organização hierárquica em um Mapa Conceitual está longe de ser uma prática pacificada. Ruiz-Primo e Shavelson (1996) questionam a necessidade de existência de estruturas hierárquicas, afirmando que a organização depende mais do conteúdo do conhecimento que está sendo organizado, e menos de questões metodológicas e conceituais. Safayeni e outros (2003) defendem o uso de Mapas Conceituais para a representação de relacionamentos estáticos e propõem os Mapas Conceituais Cíclicos, uma extensão de Mapas Conceituais, nos quais a existência de ligações não hierárquicas permitiria a representação de relacionamentos funcionais ou dinâmicos entre conceitos, tais como causalidade, probabilidade, correlação.

³ O termo “palavras de ligação” é a tradução literal do original em inglês “linking words”, presente na tradução portuguesa de Novak e Gowin (1984) e Novak (2000). Em artigos de outros autores, é muito comum encontrar o termo “frase de ligação” como sinônimo de “palavras de ligação”.

Nos mapas conceituais hierárquicos, os conceitos tendem a se agrupar abaixo de alguns conceitos centrais, configurando um domínio específico. Os conceitos que compartilham um mesmo conceito ancestral pertencem ao mesmo domínio.

É necessário acentuar que a organização do conhecimento por meio do mapa conceitual não visa o compartilhamento da aprendizagem, pois a aprendizagem, como processo interno e de responsabilidade individual, não está em condição de ser compartilhada. O mapa conceitual atua no compartilhamento de significados dos conceitos, que precisam ser discutidos, negociados e sujeitos a consenso (NOVAK; GOWIN, 1984, p. 36).

2.5 Representação de Conhecimento por meio de Mapas Conceituais

O mapa conceitual, como modelo, é materializado na forma de um diagrama composto de retângulos⁴ e segmentos de reta, ou arcos, ambos nomeados. Essa representação advém de um modelo de especificação instanciado da linguagem de especificação de mapas conceituais.

Embora a literatura não cite explicitamente a existência de uma linguagem de especificação de mapas conceituais, e de um modelo de especificação, é evidente que ambos são indiretamente mencionados na construção dos mapas conceituais, tendo em vista que a elaboração de um mapa é realizada em consonância com um conjunto de regras gramaticais e a partir de primitivas definidas em uma linguagem de especificação.

As primitivas da linguagem para especificação de mapas conceituais são os retângulos e os arcos. Os retângulos representam os conceitos e os arcos representam a existência de relação entre conceitos. Os conceitos devem ser expressos por palavras contidas dentro de um retângulo. As relações entre conceitos, representada por meio de um arco, também são expressas por palavras

⁴ Os retângulos são a representação geométrica mais comumente vista nos mapas conceituais, mas existem mapas nos quais a representação é realizada por meio de círculos, elipses, hexágonos, ou outras formas geométricas.

que devem ser posicionadas próximo ao arco. A Figura 2 ilustra um mapa conceitual.

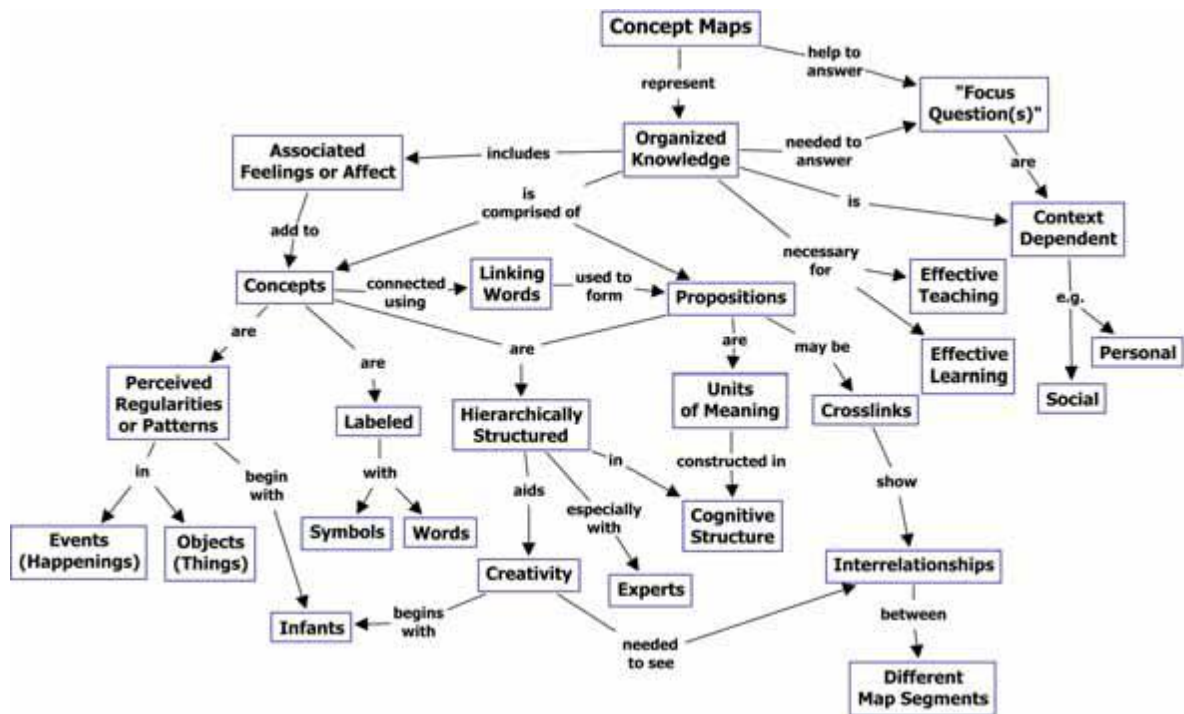


Figura 2. Exemplo de Mapa Conceitual (Fonte: NOVAK; CAÑAS; 2006a)

Formalmente, mapas conceituais podem ser definidos como um conjunto $MC = \{C, R, G\}$ em que C é um conjunto de conceitos $C = \{c_1, c_2, \dots, c_n\}$, R é um conjunto de relações entre conceitos $R = \{r_1, r_2, \dots, r_k\}$ e $G = g_1, g_2, \dots, g_m$ é um conjunto ordenado de níveis de generalização. Cada conceito c_i corresponde a uma palavra, frase e é única em C . Cada relação r_i é uma tripla da forma $r_i = (c_p, c_q, l_i)$ em que c_p e c_q são conceitos de C e l_i é o rótulo para a relação r_i que também corresponde a uma palavra ou frase. Cada nível de generalização $g_i = c_1, c_2, \dots, c_s$ que compartilha o mesmo nível de generalização, o conjunto é ordenado no sentido que para dois níveis g_i e g_j , g_i é mais geral do que g_j se e somente se $i < j$ (VILLALON; CALVO, 2008).

Neste trabalho, estendemos a definição de Villalon e Calvo (2008), propondo o metamodelo da linguagem de mapas conceituais.

A Figura 3 ilustra como nossa pesquisa entende o metamodelo da linguagem de mapas conceituais⁵, i.e., ela mostra os elementos que podemos utilizar ao construir

⁵ Um metamodelo é composto pela linguagem e pelas regras que definem o processo e os elementos para modelagem.

um mapa conceitual. Nós definimos que todos os elementos centrais derivam de um mesmo ancestral denominado “Element”. “Element” é um conjunto completo e disjunto, e uma instância de “Element” deve mandatoriamente, ser uma instância “Concept” ou uma instância de “Relationship”.

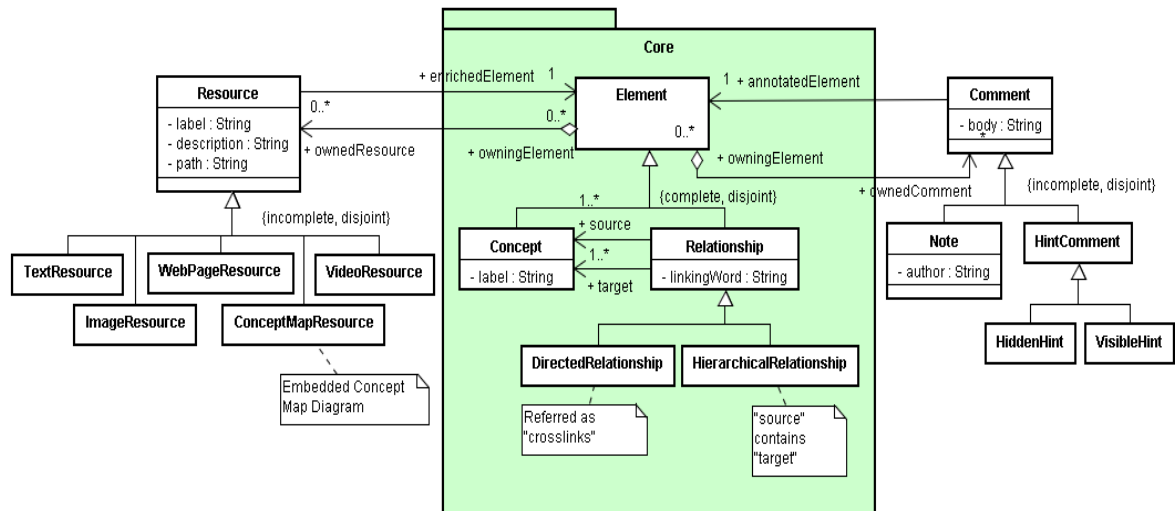


Figura 3. Metamodelo da linguagem de Mapas Conceituais

Os elementos “Concept” e “Relationship” são dois elementos presentes em qualquer mapa conceitual. Em nossa pesquisa, optamos por denominá-los como “Elementos Centrais” (“*Core Elements*”), para distingui-los de outros recursos introduzidos por ferramentas de apoio à construção de mapas conceituais. Os recursos adicionais atuam como elementos complementares que enriquecem o mapa conceitual, dentre os quais se destacam as imagens, os vídeos, as páginas de internet, os comentários e os próprios mapas conceituais embutidos em conceitos.

Os elementos centrais podem ser enriquecidos por recursos e também podem ser anotados por comentários. A diferença ontológica entre relações e conceitos está associada com os significados assumidos no mundo real. Enquanto conceitos representam “coisas e eventos”, relações representam “ações e processos”. Além disso, relações são completamente dependentes de conceitos: em um mapa conceitual, não devem existir relações desassociadas de conceitos “target” e conceitos “source”.

O metamodelo da linguagem de especificação de mapas conceituais explicita os elementos chaves para a construção de mapas conceituais e esclarece como esses elementos podem ser combinados.

O entendimento do mapa conceitual como modelo, e da linguagem de especificação de mapas conceituais, bem como das especificações válidas na construção de mapas conceituais são o fundamento para a compreensão dos próximos capítulos desta dissertação.

2.6 Considerações Finais

Para fins desta pesquisa, adotamos a aceção de mapas conceituais como modelo de domínio, representado pelo artefato concreto, na forma de diagrama, que observa um modelo de especificação válido, derivado da linguagem de especificação.

Expresso por um artefato concreto, o mapa conceitual é uma ferramenta que auxilia no processo de ensino e aprendizagem, permitindo a conciliação de diversas conceitualizações sobre um mesmo assunto.

Concebido para o trabalho com crianças, é natural que as simbologias utilizadas na linguagem de representação de mapas conceituais sejam simples. Na mesma linha de raciocínio, é igualmente compreensível o reduzido número de restrições para a sua construção. A existência de regras sofisticadas poderia tornar a construção de mapas conceituais uma atividade pouco atrativa para o universo infantil e inapropriada para o processo de aprendizagem.

Ao deixar o ambiente educacional, os mapas conceituais se deparam com públicos qualificados, normalmente adultos, que buscam uma forma de expressão para o conhecimento especializado. Neste ambiente, a facilidade de especificação festejada pelas crianças converte-se em fragilidade, no qual concessões lógicas, representações ambíguas e inexatas não são aceitáveis.

Por isso, qualquer pesquisa com mapas conceituais deve considerar os fins de sua construção para possibilitar a avaliação dos ônus associados. Naturalmente, atividades que exigem maior rigor na especificação não devem considerar a

linguagem de especificação de mapas conceituais como primeira candidata. Por outro lado, se o conteúdo não impuser limitações à forma, privilegiando a compreensibilidade para humanos, estruturas simples podem ser as mais adequadas, pois não requerem esforço extra no aprendizado das regras da própria linguagem de especificação. Declarações sobre a inadequação de mapas conceituais como artefatos ou como linguagem de especificação não cabem se desprovidas desse contexto.

No próximo Capítulo, analisaremos pesquisas que tenham como alvo a construção de mapas conceituais a partir de dados já existentes com o uso de recursos computacionais.

3 REVISÃO E PROPOSTA DE CLASSIFICAÇÃO DAS ABORDAGENS COMPUTACIONAIS PARA CONSTRUÇÃO DE MAPAS CONCEITUAIS

“Precisamos de encontrar um ponto de apoio [...]. Os homens medievais, que tinham um respeito exagerado pela autoridade dos autores antigos, diziam que os modernos, embora ao seu lado fossem «anões», apoiando-se neles tornavam-se «anões às costas de gigantes» e, deste modo, viam mais além do que os seus predecessores.”

Umberto Eco

“Como se faz uma tese”

*“O mar anterior a nós, teus medos
Tinham coral e praias e arvoredos.
Desvendadas a noite e a cerração,
As tormentas passadas e o mistério,
Abria em flor o Longe, e o Sul sidério
'Splendia sobre as naus da iniciação.”*

Fernando Pessoa

“Il Parte – Mar Portuguez”

O objetivo deste capítulo é prover informações que subsidiem o entendimento do estado da arte na construção de mapas conceituais a partir de fonte de dados, por meio de recursos computacionais.

Na Seção 3.1 elencamos as motivações que legitimam a pesquisa na construção automática de mapas conceituais.

Na Seção 3.2 propomos a caracterização da pesquisa neste campo apoiados na revisão bibliográfica de publicações do meio acadêmico ocorridas entre os anos de 1994 e 2009 que, explicitamente, declaram a intenção de produzir mapas conceituais com algum nível de automação. A caracterização é definida por um modelo que propusemos para análise das abordagens de construção automática e semiautomática de mapas conceituais (KOWATA; CURY; BOERES, 2009), baseado em um conjunto de classes de análise, agrupada em diferentes perspectivas.

Na Seção 3.3 apresentamos números que permitem compreender as principais escolhas das abordagens e inferir tendências no tratamento da construção automática de mapas conceituais.

Finalmente, na Seção 3.4, encerramos com as considerações finais deste capítulo.

3.1 Motivações

Nesta década, houve um crescente interesse da comunidade acadêmica pela aplicação de recursos computacionais para a aquisição de conhecimento de textos e de outras fontes de dados e na sua representação em forma de mapas conceituais.

As pesquisas desta área encontram nos mapas conceituais um caminho para ressignificação do conteúdo disponível na web, em bibliotecas digitais, enciclopédias, jornais online e outros repositórios de dados. Nesta área de pesquisa, os principais desafios residem na busca pela independência do auxílio humano no processo de construção de mapas conceituais, como forma de superar as dificuldades de construção a partir do “zero” (CHANG et al., 2008) e na redução de tempo e esforços para a aquisição de conhecimento (LEE et al., 2009) (TSENG et al., 2007), sobretudo, em domínios de conhecimento extensos (VALERIO; LEAKE, 2006) e dependentes de especialistas (CHANG et al., 2008).

As principais abordagens para construção de mapas conceituais a partir de fontes de dados encontram motivações em:

- Questões de ordem didática, como o uso de mapas conceituais como instrumentos no processo de ensino e aprendizagem (ALVES; PEREIRA; CARDOSO, 2001) (CLARIANA; KOUL, 2004) (LAU et al., 2008), na definição de estratégias pedagógicas (CHANG et al., 2008) (CHEN et al., 2008) (BAI; CHEN, 2008) (LEE et al., 2009) (TSENG et al.; 2007), na avaliação de desempenho de estudantes (GRAUDINA; GRUNDSPENKIS, 2008) (VILLALON; CALVO, 2008);
- Questões associadas à organização de conhecimento, utilizando mapas conceituais na apresentação de ontologias sob uma perspectiva favorável à análise de especialistas de domínio (KUMAZAWA et al., 2009);

- Questões concernentes à aquisição de conhecimento de domínios, que propõe a aplicação de mapas conceituais como ferramenta de intermediação da aquisição (ZOUAQ; NKAMBOU, 2008, 2009) (PÉREZ; VIEIRA, 2004, 2005); e
- Questões decorrentes do atendimento a requisitos de exploração e compartilhamento de documentos digitais (GAINES; SHAW, 1994) (VALERIO; LEAKE, 2006) e no resumo desse conteúdo (RICHARDSON; FOX, 2007).

Nesta pesquisa, interessamo-nos pela aplicação de mapas conceituais como recursos de apoio ao processo de ensino e aprendizagem, dentro de um contexto educacional cujas práticas estejam alicerçadas na aprendizagem significativa. Consideramos que a ressignificação de conteúdos por meio de mapas conceituais favorecem a comunicação com leitores humanos.

Na próxima seção, propomos a caracterização das abordagens da área.

3.2 Caracterização para Abordagens de Construção de Mapas Conceituais

A análise das abordagens de construção de mapas conceituais apoiada em recursos computacionais requer a escolha de um conjunto de características que serão utilizadas como base para uniformização das principais características de cada uma das abordagens.

Entretanto, ao longo de nossos estudos, constatamos a ausência de referências consistentes que permitissem a caracterização das abordagens para a construção automática e semiautomática de mapas conceituais.

Portanto, com o intuito de preencher esta lacuna, propomos, nesta seção, a caracterização da pesquisa neste campo apoiados na revisão bibliográfica de publicações do meio acadêmico ocorridas entre os anos de 1994 e 2009 que, explicitamente, declaram a intenção de produzir mapas conceituais de modo automático ou semiautomático.

Para tanto, propomos um conjunto de classes, definidas a partir de diferentes perspectivas de análise das abordagens. Cada uma destas perspectivas de análise

pretende responder a um conjunto de questões norteadoras, conforme descritas na Tabela 1.

Tabela 1. Perspectivas e Questões Norteadoras

Perspectivas	Questões Norteadoras
Objetivos da Pesquisa	<ul style="list-style-type: none"> Quais são os objetivos da pesquisa? Quem é o público-alvo dos mapas conceituais produzidos pela abordagem?
Fonte de Dados Referenciada	<ul style="list-style-type: none"> Quais tipos de fontes de dados são utilizados pela abordagem? Quais são os métodos e técnicas empregados para manipulá-los? A fonte de dados pertence a um domínio específico?
Mapas Conceituais Construídos	<ul style="list-style-type: none"> Qual é a aparência dos mapas conceituais resultantes da abordagem? Quais foram as ferramentas utilizadas na construção dos mapas conceituais? Como as abordagens avaliam as saídas resultantes?

As perspectivas e respectivas classes são ilustradas pela Figura 4. Ao longo desta seção, definimos cada uma das perspectivas e também conduziremos uma revisão das publicações acadêmicas relacionadas com a construção de mapas conceituais de forma automática e semiautomática.

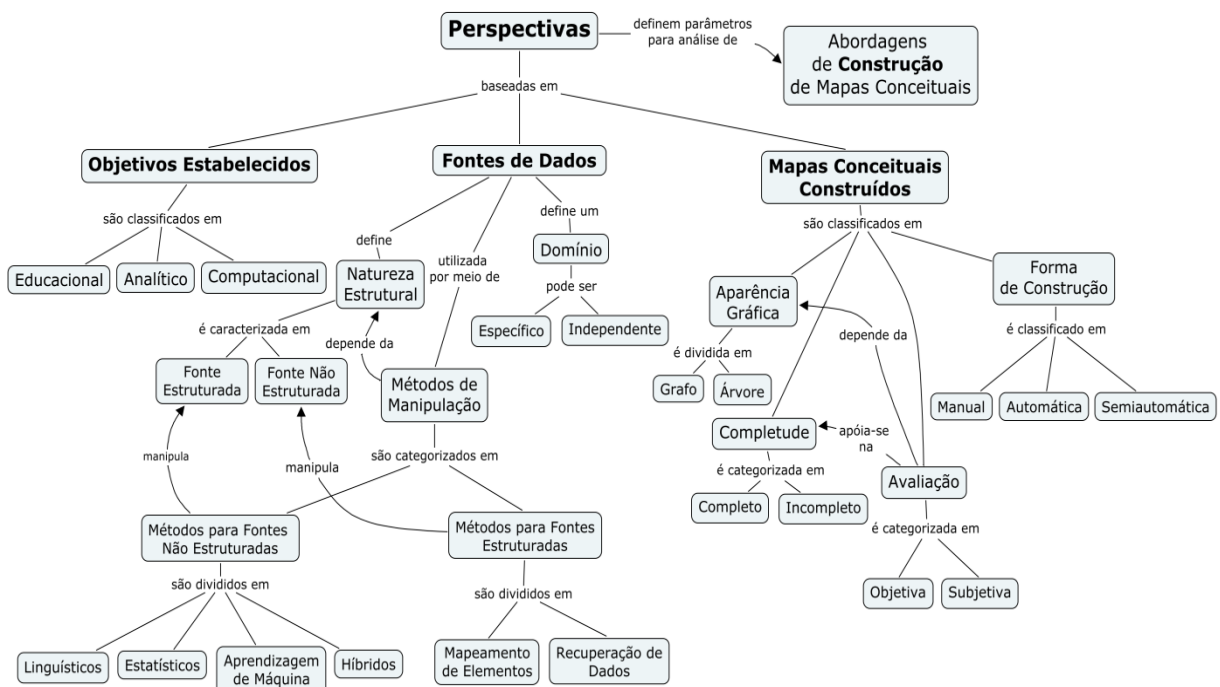


Figura 4. Perspectivas e Classes para Análise de Abordagens de Construção de Mapas Conceituais

Nas próximas seções, explicaremos cada uma das perspectivas e classes de análise.

3.2.1 Perspectiva baseada nos Objetivos da Pesquisa

Os objetivos definidos pela pesquisa para a construção de mapas conceituais direcionam a escolha da fonte de dados e definem uma variedade de requisitos, tais como: o público-alvo, o nível de legibilidade, a riqueza das proposições, entre outros.

Propomos as seguintes classes de acordos com a perspectiva baseada nos objetivos da pesquisa: **classe educacional**, **classe analítica** e **classe computacional**.

Na **classe educacional**, o principal objetivo é o uso de mapas conceituais como recursos complementares no processo de aprendizagem. Entre as abordagens que se enquadram nesta classe, mencionamos aquelas que:

- Promovem a construção interativa de mapas conceituais (ALVES; PEREIRA; CARDOSO, 2001) (CLARIANA; KOUL, 2004);
- Permitem que professores acompanhem a compreensão de alunos em tópicos específicos apresentados em ambientes de aprendizagem tradicionais (GRAUDINA; GRUNDSPEKIS, 2008) ou em ambientes virtuais de aprendizagem (LAU et al., 2008); e
- Resumem e compartilham conteúdo digital (RICHARDSON; FOX, 2007) (RICHARDSON et al.; 2008).

A **classe analítica** é caracterizada pelo uso de mapas conceituais como ferramentas para especialistas de domínio envolvidos em atividades de exploração, análise e reconhecimento de padrões em fontes de dados considerados extensas. Esse é o contexto das abordagens que:

- Exploram registros históricos de alunos, buscando por padrões de aprendizagem (BAI; CHEN, 2008) (CHANG et al., 2008) (CHEN et al., 2008) (LEE et al., 2009) (TSENG et al., 2007);
- Exibem informações para os especialistas de domínio (GAINES; SHAW, 1994) (KUMAZAWA et al., 2009) (PÉREZ; VIEIRA, 2004, 2005); e

- Propõem o uso de mapas conceituais como ferramentas para acesso, busca e análise de dados em bibliotecas digitais (VALERIO; LEAKE, 2006).

Na **classe computacional**, o fio condutor consiste na manipulação de mapas conceituais para a aquisição de conhecimento por meio de aplicações de software. Nesta classe estão as pesquisas que utilizam os mapas conceituais como representações intermediárias entre o texto e as ontologias de domínio (ZOUAQ; NKAMBOU, 2008, 2009).

3.2.2 Perspectiva baseada nas Fontes de Dados

A fonte de dados⁶ usada por uma abordagem impacta todo o processo de construção de mapas conceituais. Como uma entrada, a natureza estrutural da fonte de dados atua tanto como parâmetro de escolha do método de sua manipulação, quanto como uma restrição que determina o nível de dependência do domínio. Abaixo da perspectiva da fonte de dados, determinamos pontos de vista mais específicos, e definimos: **(i) natureza estrutural; (ii) método de manipulação, e (iii) dependência de domínio.**

3.2.2.1 Perspectiva da Natureza Estrutural

A natureza estrutural da fonte de dados é definida pela organização dos dados. A partir deste ponto de vista, as fontes de dados usadas nas abordagens são organizadas em duas classes: na classe das fontes de dados **estruturadas** e na classe das fontes de dados **não estruturadas**. Na classe das fontes de dados não estruturadas, textos em linguagem natural são as fontes primitivas para extração dos elementos de mapas conceituais. Entre as abordagens que usam dados não estruturados, podemos mencionar aqueles que trabalham com:

- Textos produzidos por especialistas de domínio (ALVES; PEREIRA; CARDOSO, 2001) (GAINES; SHAW, 1994) (VALERIO; LEAKE, 2006) (ZOUAQ; NKAMBOU, 2008, 2009);

⁶ Quando houver mais de uma fonte de dados referenciada, devemos considerar somente as fontes de dados primárias, i. e., a fonte de dados usada pela abordagem para a extração dos conceitos e das relações. Portanto, devemos ignorar as características de fontes de dados auxiliares, geralmente utilizadas com propósitos lexicais, e consultados em atividades complementares como, por exemplo, na desambiguação de palavras, na identificação de sinônimos e no *part-of-speech tagging*.

- Textos produzidos por estudantes em resposta a questionários (CHANG et al., 2008) ou a partir de um tema proposto (CLARIANA; KOUL, 2004);
- Textos científicos como, por exemplo, artigos acadêmicos (CHANG et al., 2008), resumos de teses e dissertações (RICHARDSON et al., 2008);
- Mensagens de fóruns de discussão (LAU et al., 2008);
- Textos jornalísticos e didáticos (PÉREZ; VIEIRA, 2004, 2005).

A classe das **fontes de dados estruturadas** apresenta dados organizados com algum nível de formalismo. Implementações para modelos conceituais e de ontologias de domínio são exemplos de fontes de dados estruturadas e estão presentes nas abordagens que utilizam respostas tabuladas de estudantes (BAI; CHEN, 2008) (LEE et al., 2009) (TSENG et al., 2007) e nas que utilizam ontologias de domínio (GRAUDINA; GRUNDSPENKIS, 2008) (KUMAZAWA et al., 2009).

3.2.2.2 Perspectiva do Método de Manipulação

O vínculo entre os métodos e técnicas computacionais aplicados na construção automática e semiautomática de mapas conceituais e a natureza estrutural da fonte de dados determinam a distinção dos métodos para fontes de dados não estruturadas dos métodos para fontes de dados estruturadas.

Os métodos que utilizam fontes de dados não estruturadas guardam um estreito relacionamento com o campo do Processamento de Linguagem Natural e todos eles poderiam, sob este critério, ser igualmente denominados de métodos baseados em métodos linguísticos. Entretanto, a presença de técnicas de manipulação linguística não é condição suficiente na definição da natureza técnica da abordagem (GÓMEZ-PÉREZ; MANZANO-MACHO, 2005). Por isso, a caracterização considera a principal técnica utilizada na descoberta de conhecimento de textos. As classes para os métodos de manipulação são qualificadas em classe de **Métodos Linguísticos**, classe de **Métodos Estatísticos**, classe de **Métodos de Aprendizagem de Máquina** e, na impossibilidade de uma classificação única, há a classe dos **Métodos Híbridos** (ZOUAQ; NKAMBOU, 2009).

A classe dos **Métodos Linguísticos** é caracterizada pelo uso de técnicas de linguística computacional (GÓMEZ-PÉREZ; MANZANO-MACHO, 2005) que utilizam recursos de análise morfológica, sintática, semântica, pragmática e do discurso. Tais técnicas são extremamente relacionadas com as características estruturais dos textos e baseiam-se, principalmente, na identificação de padrões linguísticos, na detecção de termos por meio de gramáticas que definem regras sintáticas, na definição de papéis temáticos e no reconhecimento de entidades nomeadas. De forma geral, os métodos linguísticos são mais precisos do que os Métodos Estatísticos, mas normalmente requerem o auxílio de bases de conhecimentos externos, tais como dicionários, thesaurus e banco de dados lexicais (ZOUAQ; NKAMBOU, 2009). Neste grupo estão abordagens de Pérez e Vieira (2004; 2005), Richardson e Fox (2007) e Richardson e outros (2008).

Na classe dos **Métodos Estatísticos**, pode-se observar a utilização de técnicas baseadas em indicadores quantitativos (GÓMEZ-PÉREZ; MANZANO-MACHO, 2005). De forma geral, tais técnicas produzem informações que permitem a análise da frequência de um termo e coocorrências entre termos em documentos ou *corpus*. As técnicas mais populares são a análise de frequência de repetição de termos ou padrões de palavras, o cálculo de pesos que indicam a relevância de termos em um conjunto de documentos (TF-IDF) e técnicas de agrupamento de documentos (clusterização). A grande contribuição dos métodos estatísticos está na relativa simplicidade de manipulação do documento, independente de conhecimento sobre as estruturas linguísticas das sentenças. O aspecto negativo a ser destacado é a imprevisibilidade dos resultados e a inadequação na representação semântica das sentenças (ZOUAQ; NKAMBOU, 2009). Neste grupo estão abordagens de Gaines e Shaw (1994) e de Clariana e Koul (2004).

Na classe dos **Métodos de Aprendizagem de Máquina** utilizam-se técnicas de aprendizagem de máquina para a extração de elementos do texto. São técnicas que aplicam vários algoritmos para apoiar a descoberta de conceitos e suas relações (GÓMEZ-PÉREZ; MANZANO-MACHO, 2005) e são normalmente utilizados em conjunto com métodos estatísticos. Exemplos de métodos de aprendizagem de máquina podem ser encontrados nos algoritmos utilizados na descoberta de elementos coocorrentes e que podem caracterizar regras de associação, na

identificação de palavras-chaves e taxonomias entre elementos (ZOUAQ; NKAMBOU, 2009).

A classe dos **Métodos Híbridos** caracteriza-se pela combinação de técnicas linguísticas, técnicas de aprendizagem de máquina e técnicas estatísticas de forma que não se pode precisar a técnica dominante. As abordagens híbridas estão presentes em Alves et al. (2001), Lau et al. (2008), Valerio e Leake (2006) e Zouaq e Nkambou (2008; 2009)

Os métodos que utilizam fontes de dados estruturadas geralmente aplicam métodos de Mapeamento de Elementos ou Métodos de Recuperação de Dados. Esses métodos, na maioria das vezes, utilizam técnicas de mapeamento de elementos e técnicas de recuperação de dados. Baseado nestas técnicas, identificamos duas grandes classes para manipulação de fontes de dados estruturadas: a classe dos **Métodos de Mapeamento de Elementos** e dos **Métodos de Recuperação de Dados**.

As abordagens de Graudina e Grundspenkis (2008) e Kumazawa e outros (2009) podem ser enquadradas na **classe dos métodos de Elementos de Mapeamento** porque elas tratam do mapeamento de conceitos, propriedades e associações existentes em ontologias de domínio para elementos de mapas conceituais. As abordagens de Bai e Chen (2008), Chen e outros (2008), Lee e outros (2009), Tseng e outros (2007) não requerem grande esforço na identificação de conceitos de um mapa conceitual porque os definem no momento em que estabelecem uma questão focal. Portanto, estas abordagens podem ser enquadradas dentro da classe de **Métodos de Recuperação de Dados** para recuperação de conceitos armazenados na fonte de dados.

3.2.2.3 Perspectiva da Dependência de Domínio

Segundo o domínio, observamos que as abordagens podem ser distribuídas na classe **independente de domínio** ou na classe **específica de domínio**.

As abordagens da classe **independente de domínio** não apresentam restrição para a construção de mapas conceituais a um domínio específico e nem requerem

conhecimento prévio sobre o mesmo. Nesta categoria se enquadram os trabalhos de Alves e outros (2001), Pérez e Vieira (2004) (2005) e Valerio e Leake (2006).

As abordagens da classe **específica de domínio** manipulam fontes de dados pertencentes a um domínio específico, tais como: *E-learning* (CHEN et al.; 2008) e Informática aplicada à Educação (ZOUAQ; NKAMBOU, 2008, 2009), Fisiologia e Funcionamento do Coração Humano (CLARIANA; KOUL, 2004), Sistemas de Manufatura apoiados por Computador (GAINES; SHAW, 1994), Ciência da Computação (GRAUDINA; GRUNDSPENKIS, 2008) (RICHARDSON; FOX, 2007) (RICHARDSON et al., 2008), Ciência da Sustentabilidade (KUMAZAWA et al., 2009). Os trabalhos de Bai e Chen (2008), Chang e outros (2008), Lau e outros (2008), Lee e outros (2009) e Tseng e outros (2007) não são, necessariamente, direcionados a um domínio específico, entretanto a independência de conhecimento prévio do domínio é comprometida quando a abordagem adota o planejamento prévio de questões focais e a identificação de conceitos associados.

3.2.3 Perspectiva baseada nos Mapas Construídos

Considerando a perspectiva de saída da abordagem, o mapa conceitual construído pode ser analisado considerando (i) a aparência gráfica, (ii) a completude em relação à sua definição, (iii) a forma de construção e (iv) os mecanismos propostos para avaliação de sua qualidade. A aparência gráfica de um mapa conceitual pode ser diretamente observada a partir do mapa conceitual. Caso exista uma topologia explícita, o mapa conceitual pertence à classe dos mapas em forma de **árvore**, senão, à classe da estrutura na forma de **grafos**. As definições de ambas as estruturas matemáticas podem ser encontradas na Teoria de Grafos (DIESTEL, 2005).

Quanto à completude, um mapa conceitual pode ser analisado em relação à aderência aos elementos definidos na Teoria de Mapas Conceituais, caracterizando-se classes de mapas **completos** ou **incompletos**. Em mapas conceituais ditos incompletos, proposições não podem ser identificadas, pois não há rótulos para as relações ou sua semântica é insuficiente para o propósito de gerar proposições, ao contrário dos mapas completos nos quais as proposições são claramente definidas.

Quanto à forma de construção, as abordagens variam na proposta para a construção de representações gráficas. Algumas apresentam recursos automáticos e nativos, outras utilizam recursos para integração a ferramentas de terceiros ou simplesmente não entram no mérito da construção gráfica. Na construção de representação da **classe automática**, a abordagem apresenta um recurso nativo que constrói o mapa conceitual visualmente, sem a intervenção humana nesta construção. Dentre as abordagens que se enquadram nesta categoria estão a de Bai e Chen (2008), Chang e outros (2008), Chen (2008), Kumazawa (2009), Lau e outros (2008), Lee e outros (2009), Tseng e outros (2007) e Zouaq e Nkambou (2008, 2009). Nas abordagens da **classe semiautomática**, um conjunto de proposições são geradas e importadas por ferramentas de representação visual, como observado em Alves e outros (2001), Clariana e Koul (2004), Gaines e Shaw (1994), Graudina e Grundspenkis (2008), Richardson e Fox (2007), Richardson e outros (2008). Já na classe de construção **manual**, as abordagens extraem apenas os conceitos e as relações, mas não definem proposições. Por isso, o mapa deve ser construído manualmente em uma ferramenta apropriada, como em Valerio e Leake (2006).

Quanto à Avaliação, que consiste na validação de proposições no mapa conceitual e de elementos estruturais do mapa conceitual construído, as classes são denominadas de classe **Subjetiva** e classe **Objetiva**. A avaliação na classe subjetiva é conduzida por meio de especialistas humanos que utilizam seus próprios critérios na validação do mapa conceitual, impossibilitando uma avaliação uniforme e independente do especialista. As abordagens analisadas utilizam quase que exclusivamente esta abordagem. A avaliação na classe objetiva pode ser conduzida tanto por especialistas humanos quanto por recursos computacionais, pois se baseiam em critérios de pontuação, originalmente definidos por Novak e Gowin, para proposições válidas, níveis de hierarquia, número de ramificações, existência de cross-links e exemplos específicos (NOVAK; GOWIN, 1984).

Uma síntese das abordagens pode ser encontrada na Tabela 2.

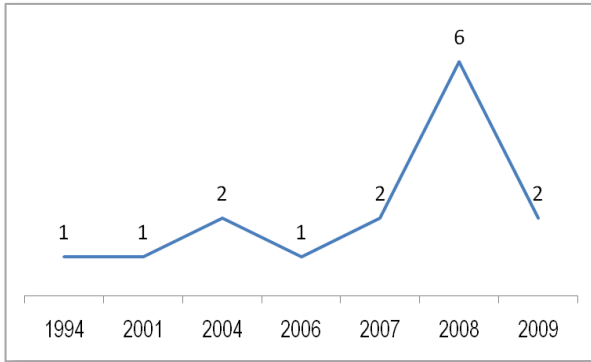
Tabela 2. Caracterização das Abordagens para Construção de Mapas Conceituais com o uso de recursos computacionais

Abordagem		Fonte de Dados			Mapa Construído		
Referência	Objetivo	Domínio	Estrutura	Método	Construção	Leiaute Completa	Avaliação
Alves e outros (2001)	Educacional	Independente de Domínio	Não estruturada	Método Híbrido	Semi	Grafo Completo	Subjetiva
Bai et al (2008)	Analítico	Específica de Domínio	Estruturada	Recuperação de Dados	Automática	Grafo Incompleto	Informação não disponível
Chang e outros (2008)	Analítico	Específica de Domínio	Não estruturada	Método Estatístico	Automática	Árvore Incompleto	Informação não disponível
Chen e outros (2008)	Analítico	Específica de Domínio	Não estruturada	Método Estatístico	Automática	Grafo Incompleto	Subjetiva
Clariana e Koul (2004)	Educacional	Específica de Domínio	Não estruturada	Método Estatístico	Semi	Grafo Incompleto	Objetiva
Gaines e Shaw (1994)	Analítico	Específica de Domínio	Não estruturada	Método Estatístico	Semi	Grafo Incompleto	Subjetiva
Graudina e Grundspenkis (2008)	Educacional	Específica de Domínio	Estruturada	Mapeamento de Elementos	Semi	Grafo Incompleto	Subjetiva
Kumazawa e outros (2009)	Analítico	Específica de Domínio	Estruturada	Mapeamento de Elementos	Automática	Grafo Incompleto	Subjetiva
Lau e outros (2008)	Educacional	Específica de Domínio	Não estruturada	Método Híbrido	Automática	Grafo Incompleto	Subjetiva
Lee e outros (2009)	Analítico	Específica de Domínio	Estruturada	Recuperação de Dados	Automática	Grafo Incompleto	Subjetiva
Pérez e Vieira (2004, 2005)	Educacional	Independente de Domínio	Não estruturada	Método Linguístico	Informação não disponível	Grafo Completo	Subjetiva
Richardson e Fox (2007); Richardson e outros (2008)	Educacional	Específica de Domínio	Não estruturada	Método Linguístico	Semi	Grafo Completo	Subjetiva
Tseng e outros (2007)	Educacional	Específica de Domínio	Estruturada	Recuperação de Dados	Automática	Grafo Incompleto	Subjetiva
Valerio e Leake (2006)	Analítico	Independente de Domínio	Não estruturada	Método Híbrido	Manual	Grafo Completo	Subjetiva
Zouaq e Nkambou (2008, 2009)	Computacional	Específica de Domínio	Não estruturada	Método Híbrido	Automática	Grafo Completo	Subjetiva

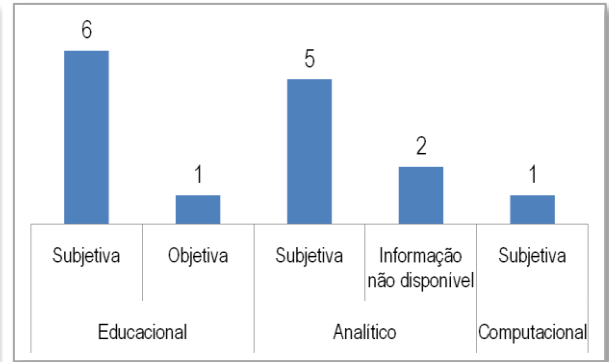
3.3 Análise de Resultados

A partir do uso das características definidas para análise das abordagens, como exibido na Tabela 2, tornamos possível uma compreensão mais aprofundada das pesquisas na área de construção automática e semiautomática de mapas conceituais, por meio de uma série de gráficos, ilustrados na Figura 5.

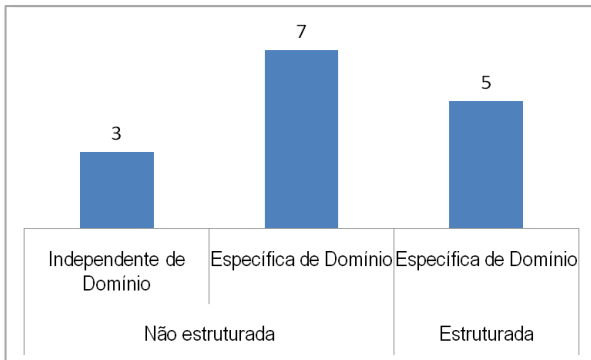
Entre as observações relevantes para esta pesquisa, constatamos na Figura 5(a) que a maioria das abordagens analisadas é recente: 10 das 15 abordagens (66,67%) foram publicadas nos últimos três anos.



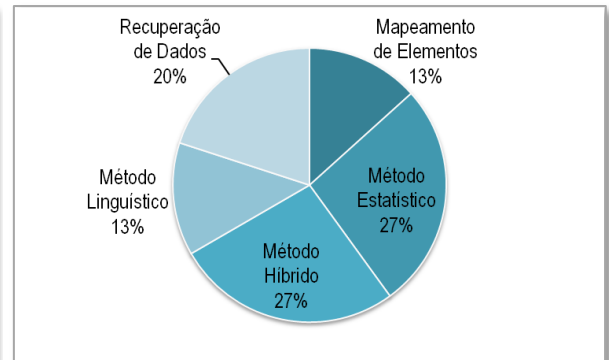
(a) Número de Abordagens por Ano



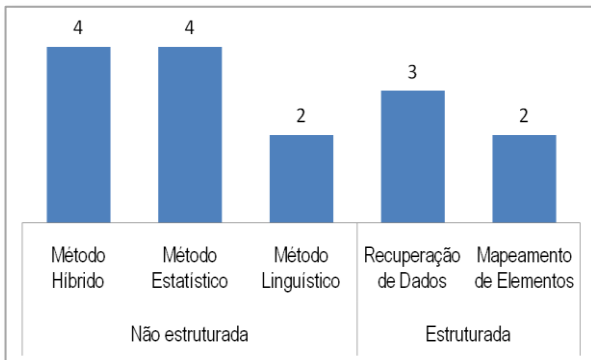
(b) Objetivos x Avaliação



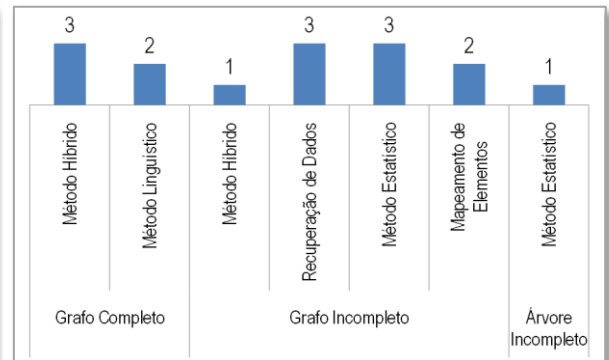
(c) Natureza Estrutural x Domínio



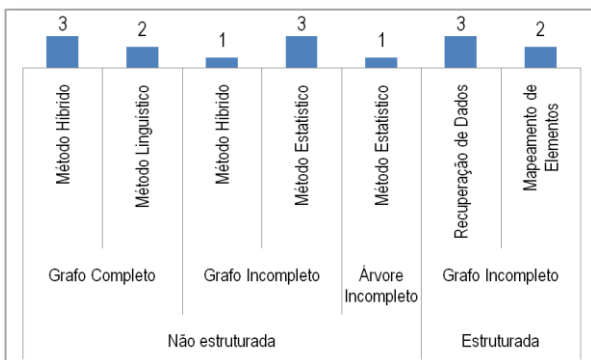
(d) Métodos de Manipulação



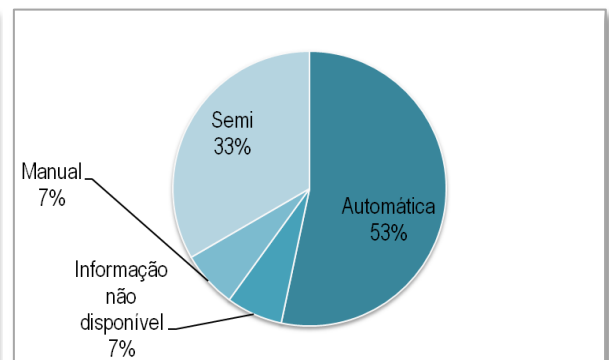
(e) Natureza Estrutural x Método



(f) Leiaute/Completeness x Método



(g) Natureza Estrutural x Leiaute/Completeness x Método



(h) Forma de Construção

Figura 5. Análise das Abordagens Investigadas

Na Figura 5(b) é possível observar que 14 (93,34%) das 15 abordagens destinam-se à produção de mapas conceituais para leitores humanos, considerando-se a soma dos objetivos analíticos (7 \approx 46,67%) e educacionais (7 \approx 46,67%). Importante notar, neste mesmo gráfico, que a avaliação subjetiva do mapa conceitual construído está presente em 12 abordagens (80%).

Na Figura 5(c), os números revelam o interesse significativo pela utilização de fontes de dados não estruturadas que estão presentes em 10 abordagens (66,67%). Aliás, são nas fontes de dados não estruturadas que encontramos as únicas três abordagens (2,00%) que tratam da construção de mapas conceituais desvinculados de um domínio específico, uma vez que a maioria das abordagens, 12 (80%), é dependente de domínio.

Os métodos híbridos (26,66%) e estatísticos (26,66%) são adotados em mais abordagens do que os métodos puramente linguísticos (13,33%), conforme Figura 5(d). Em uma análise mais criteriosa, a partir da Figura 5(e), notamos que os métodos híbridos, estatísticos e linguísticos respondem pelas fontes de dados não estruturadas. Na observação das Figuras 5(f) e 5(g) fica evidente que mapas conceituais ditos completos foram obtidos somente na manipulação de fontes de dados não estruturados por meio de métodos híbridos e métodos linguísticos, mas representam apenas 5 abordagens (33,33%) do universo das 15 analisadas e apresentam-se na forma de grafos.

Por fim, na Figura 5(h) observa-se que 53% das abordagens oferecem alternativas para construção automática de mapas conceituais a partir da manipulação da fonte de dados; que 33% das abordagens ainda necessitam de algum nível de intervenção humana neste processo e ainda há 7% das abordagens que não apresentam nenhuma proposta neste sentido.

3.4 Considerações Finais

O interesse pela pesquisa na construção automática de mapas conceituais a partir de fontes de dados é relativamente recente. Consideramos que as perspectivas e classes apresentadas ao longo deste capítulo contribuirão para ampliar o entendimento das soluções existentes. Embora considerável parcela destas pesquisas se autodenomine de “construções automáticas”, entendemos como mais

adequado o emprego da expressão “construção semiautomática” dado o momento em que se encontram tais abordagens, ainda significativamente dependentes de intervenções humanas ao longo do processo de construção.

Outro ponto que merece destaque é a relação entre a construção de mapas conceituais e os dados não estruturados. A manipulação de dados não estruturados é um desafio tangível para as aplicações computacionais. Grande massa de dados não estruturados está disponível na web e em bibliotecas digitais e seu conteúdo aumenta diariamente. Por isso, é compreensível que 10 das 15 abordagens (66,67%) estejam ocupadas na definição de métodos para dados não estruturados.

Observações curiosas relacionadas às ontologias de domínio também surgiram em nossas pesquisas: duas abordagens usam mapas conceituais para representar ontologias para leitores humanos (GRAUDINA; GRUNDSPENKIS, 2008) (KUMAZAWA et al., 2009). Por outro lado, existem pesquisas que usam mapas conceituais como representações intermediárias entre dados não estruturados e ontologias de domínio (ZOUAQ; NKAMBOU, 2008, 2009). Podemos assumir que, embora mapas conceituais sejam normalmente subestimados devido à simplicidade de sua estrutura e também por causa da sua falta de formalismo, eles ainda são muito úteis para a compreensão humana.

Estamos cientes de que o conjunto de classes propostas neste capítulo é declaradamente incompleto e passível de extensões futuras, dentre as quais se destacam os aspectos relativos à qualidade, à completude e aos contextos de uso de mapas conceituais. Portanto, este mapeamento não deve ser entendido como um assunto exaurido; ao contrário, propusemos apenas um ponto de partida para iniciar o debate e, principalmente, para fundamentar o trabalho descrito por esta dissertação.

No próximo capítulo definimos uma abordagem híbrida para manipular textos em Língua Portuguesa (do Brasil) para produzir mapas conceituais completos, com o intuito de utilizá-los como recursos de apoio ao processo de ensino e aprendizagem. Para alcançar este objetivo, desenvolvemos técnicas para reconhecimento de elementos centrais dos mapas conceituais e utilizamos a estrutura de grafos para identificar as dependências entre os elementos.

4 UMA ABORDAGEM COMPUTACIONAL PARA CONSTRUÇÃO DE MAPAS CONCEITUAIS A PARTIR DE TEXTOS EM LÍNGUA PORTUGUESA DO BRASIL

“[...]
— Que bairro será esse? — perguntou Narizinho.
— Um muito importante — o bairro dos NOMES, ou SUBSTANTIVOS.
— Que emproados! — observou Emília. — Até parecem as Vogais da terra do alfabeto.
— E são de fato as Vogais das palavras. Sem eles seria impossível haver linguagem, porque os Substantivos é que dão nome a todos os seres vivos e a todas as coisas. Por isso se chamam Substantivos, como quem diz que indicam a substância de tudo [...]
[...]
— Agora iremos visitar [...] os VERBOS, uma espécie muito curiosa de palavras. Depois dos Substantivos são os Verbos as palavras mais importantes da língua. Só com um Nome e um Verbo já podem os homens exprimir uma idéia.”

Monteiro Lobato
“Emília no País da Gramática”

Neste capítulo, propomos uma abordagem para construção de mapas conceituais a partir de textos em português do Brasil, com auxílio de recursos computacionais.

Na Seção 4.1 especificamos o problema, definindo-o formalmente.

Na Seção 4.2 apresentamos uma visão geral da abordagem proposta, apresentando o raciocínio e as premissas consideradas na sua concepção.

Na Seção 4.3 detalhamos cada uma das atividades definidas no processo, descrevendo as entradas e as saídas esperadas.

Finalmente, na Seção 4.4, encerramos com as considerações finais deste capítulo.

4.1 Especificação do Problema

A integração de diferentes tecnologias no currículo escolar é uma prática que fomenta o surgimento de maneiras diversas de ensinar e aprender, tanto pela introdução de novas tecnologias quanto pela ressignificação das chamadas tecnologias convencionais, competindo ao professor decidir quais veículos e linguagens privilegiar para levar os estudantes à melhor compreensão dos conteúdos (SILVA, 2005, p. 36).

Nesta pesquisa, consideramos que os mapas conceituais são ferramentas que possibilitam a ressignificação de conteúdos, apoiando, desta forma, práticas educacionais voltadas para a aprendizagem significativa. No processo de criação, avaliação e evolução dos mapas conceituais, o estudante é instigado a se tornar sujeito ativo do próprio aprendizado (NOVAK, 2000), atuando de forma colaborativa com pares e professor para a construção do conhecimento.

Assim, ao longo desta dissertação, investigamos o problema da ressignificação de conteúdos textuais, por meio de mapas conceituais, buscando, em última instância, a construção de conteúdos dinâmicos a partir de conteúdos estáticos. Propomos uma solução amparada na aplicação de recursos computacionais por considerar que a independência do auxílio humano no processo de construção de mapas conceituais contribui para a superação das dificuldades inerentes à construção a partir do “zero” (CHANG et al., 2008), na redução de tempo e esforços despendidos na aquisição de conhecimento (LEE et al., 2009) (TSENG et al. 2007), sobretudo, em domínios extensos (VALERIO; LEAKE, 2006) e dependentes de especialistas (CHANG et al. 2008).

A pesquisa propõe uma abordagem para a construção automática de mapas conceituais a partir de textos. Propusemo-nos a tratar textos em português do Brasil, em conformidade com as recomendações da UNESCO que orientam para a criação e aplicação de recursos de aprendizagem em língua materna (UNESCO, 2009a, p.14). Por conseguinte, o escopo, determinado com mais especificidade, resume-se à **definição de uma abordagem que permita a construção de mapas conceituais**

a partir de textos em Língua Portuguesa do Brasil⁷ por meio do uso de recursos computacionais.

Formalmente, o problema resume-se na necessidade de tomar um documento d , com conteúdo expresso em português do Brasil, e representá-lo por meio de um mapa conceitual mc . O conteúdo do documento d é constituído por um conjunto de sentenças $s_1...s_n$, ou seja, $d = \{s_1, s_2, \dots, s_n\}$. Para cada sentença s_i definida por meio de linguagem natural, em que i designa o índice da sentença no documento d , é necessário identificar proposições $p_i...p_n$ a serem extraídas de cada sentença, com o intuito de possibilitar a construção de um mapa conceitual mc . Uma proposição p_i é definida por um conjunto de três elementos ordenados, $c1_i$, r_i e $c2_i$, nos quais $c1_i$ e $c2_i$ são conceitos e r_i uma relação entre esses conceitos.

A transformação de uma sentença s_i em triplas proposicionais no formato $\{\text{conceito} - \text{relação} - \text{conceito}\}$ requer a identificação prévia dos elementos candidatos a conceitos e relações. Sob o ponto de vista computacional, enfrentamos a questão da descoberta de elementos candidatos a conceitos e relações de mapas conceituais a partir de uma sentença s_i , como um problema de reconhecimento de padrões linguísticos e de entidades. A construção de proposições a partir de conceitos e relações é abordada como um problema de otimização que pode ser tratado por meio da seleção de um arranjo adequado entre elementos com o objetivo de uma tripla.

4.2 Visão Geral da Abordagem

Antes de apresentarmos a abordagem proposta para a construção automática de mapas conceituais a partir de textos em português do Brasil, é necessário esclarecer a intenção subjacente à pesquisa. Em capítulos anteriores, afirmamos que os esforços presentes nesta pesquisa não são meros exercícios computacionais, mas ações que buscam alternativas favoráveis à integração de diferentes tecnologias e mídias a fim de criar oportunidades para a descoberta de maneiras de ensinar, aprender e desenvolver o currículo. Portanto, à luz da Educação, a resignificação

⁷ A Língua Portuguesa é a língua materna de cerca de 200 milhões de pessoas e a sexta mais falada no mundo. As variações do português com maior número de falantes são o português europeu e o português brasileiro que diferem em aspectos fonológicos, lexicais, morfológicos e sintáticos. (ALUÍSIO et al., 2004).

de conteúdos convencionais por meio de mapas conceituais, possibilitando o seu uso como recursos de apoio à aprendizagem, passa pela necessidade de que estes mapas sejam: (i) fiéis ao conteúdo do texto que lhes deu origem e (ii) compreensíveis para leitores humanos.

A abordagem que propomos busca resguardar a fidelidade ao conteúdo do texto original, assegurando a compreensibilidade dos mapas conceituais construídos e por isso, necessita enfrentar os desafios inerentes à manipulação da linguagem natural, além das particularidades da Língua Portuguesa. Então, a adoção de métodos e técnicas que restrinjam a perda semântica no mapeamento texto-mapa conceitual sem comprometer a essência do conteúdo e sem limitar a leitura humana é crucial para o êxito desta abordagem.

Dentre os achados relatados no Capítulo 3 desta dissertação, encontramos números importantes: 10 das 15 abordagens, 66,67%, usam fontes de dados não estruturadas; destas, metade constrói mapas conceituais ditos completos por meio de métodos híbridos e métodos linguísticos. As opções metodológicas para esta pesquisa não poderiam ser realizadas à margem dos dados expostos. Levando-se em consideração os resultados apresentados pelas pesquisas similares desenvolvemos uma predileção natural pelos métodos linguísticos e híbridos para tratar do problema da representação do conteúdo de um documento em Língua Portuguesa por meio de mapas conceituais.

Novak e Gowin (1984) afirmam que, na construção de mapas conceituais, é importante que palavras adequadas sejam escolhidas para rotular conceitos e relações. Considerando que um documento é formado por sentenças, que por sua vez são listas de palavras dispostas segundo regras linguísticas, configura-se um grande desafio a identificação de um único subconjunto de palavras para nomear de forma adequada os elementos centrais dos mapas conceituais.

Com vistas a clarificar o problema, lançamos mão da analogia (NONAKA; TAKEUCHI, 1997, p.70) entre o documento e o brinquedo de encaixe, no formato do famoso sistema LEGO⁸. No brinquedo de encaixe, peças podem ser arranjadas de diversas maneiras, produzindo diferentes módulos. Analogamente, no documento,

⁸ LEGO é a designação comercial de uma família de brinquedos composta por peças que se encaixam (www.lego.com)

as sentenças são módulos constituídos por peças, no caso, por peças linguísticas, tais como: palavras, números e sinais de pontuação. Estas peças linguísticas são organizadas observando-se a sintática, a semântica e a pragmática. Tanto no conteúdo de um documento como em módulos construídos a partir das peças de encaixe, pode-se observar a existência de unidades atômicas de construção. O rearranjo das peças do brinquedo permite novas construções, assim como a recombinação das peças linguísticas admite outras representações, sendo a representação proposicional a peremptória para a construção de mapas conceituais.

A partir da definição da analogia, como exposta anteriormente, propomos uma abordagem em que o conteúdo de um documento deve ser fragmentado até que suas peças, de natureza linguística, tornem-se disponíveis. Posteriormente, as peças linguísticas são reconstruídas na forma de conceitos e relações em mapas conceituais. A expressão síntese da abordagem consiste na dicotomia “desconstruir-reconstruir”. Equacionando o problema, projetamos um modelo de processo composto por um conjunto de atividades que se concatenam por meio de suas entradas e saídas, formando um túnel, de sentido único, linear, como ilustrado na Figura 6.

A abordagem se inicia com a seleção do documento que será “desconstruído”. Primeiramente, o conteúdo do documento é processado com o objetivo de eliminar marcas de formatação para que o mesmo passe a conter somente texto puro. Em seguida, o conteúdo é separado em orações, que por sua vez, são divididas em *tokens*, ou seja, em palavras, números e sinais. A seguir, cada *token* é etiquetado com a etiqueta morfológica (*Part-of-Speech tag*⁹ ou *POS tag*) correspondente, deixando evidente a estrutura de constructos da oração.

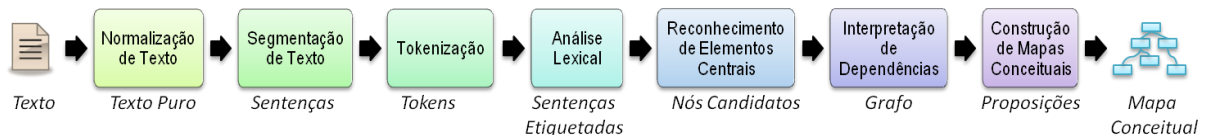


Figura 6. Modelo de Processo para Construção de Mapas Conceituais a partir de Textos

⁹ *Part-of-Speech tag (POS tag)* é um rótulo que identifica uma “classe de palavra”, como por exemplo, verbos, adjetivos, substantivos, advérbios, preposições, etc.

Por meio de um conjunto de padrões linguísticos baseados nas etiquetas morfológicas, grupos linguísticos denominados de *chunks*¹⁰ são reconhecidos em uma atividade que visa à identificação de elementos candidatos a conceitos e relações, com o objetivo de iniciar a reconstrução na forma de um mapa conceitual. Os candidatos a elementos centrais de mapas conceituais, uma vez reconhecidos, passam a ser identificados como nós de grafos¹¹. Por meio de uma atividade que consiste no caminhar pelo grafo, buscam-se arranjos na forma de proposições. Quando uma tripla é reconhecida, os nós são mapeados na forma de conceitos ou de relações. Ao final da atividade, quando todos os nós do grafo já tiverem sido visitados, tem-se o conjunto de proposições que pode ser exibido graficamente como um mapa conceitual.

Na próxima seção, cada uma das atividades será explicada em mais detalhes.

4.3 Atividades

São sete as atividades do processo para a construção de mapas conceituais a partir de textos:

- (i) *Normalização de Texto;*
- (ii) *Segmentação de Texto;*
- (iii) *Tokenização;*
- (iv) *Análise Lexical;*
- (v) *Reconhecimento de Elementos Centrais;*
- (vi) *Interpretação de Dependências; e*
- (vii) *Construção de Mapas Conceituais.*

As atividades são coordenadas de forma sequencial, do primeiro ao último, orquestradas de forma que a cada atividade executada seja produzido um único

¹⁰ *Chunk Linguístico* é uma sequência de palavras que possui um significado quando dispostos em conjunto. Entre essas palavras existe somente uma que é indispensável, exercendo o papel de núcleo e que normalmente é um substantivo ou um verbo.

¹¹ Grafo é uma estrutura matemática composta por nós e arestas.

artefato que será tomado como insumo para a atividade seguinte, do documento original até o mapa conceitual.

4.3.1 Normalização de Texto

A primeira atividade do processo consiste na leitura do texto, denominada de *Normalização de Texto*. A atividade é responsável por eliminar os marcadores de formatação existentes no conteúdo de um documento, tais como *tags*, estilos de fontes, metadados. A eliminação destas marcas permite a manipulação de arquivos em diferentes formatos de texto.

A existência de inúmeros formatos de documentos aliada à velocidade de surgimento de novas extensões e de atualizações de versão que não guardam compatibilidade com as versões anteriores são fatores complicadores para a atividade que depende, fundamentalmente, da definição de filtros apropriados para tratamento das diferentes codificações.

4.3.2 Segmentação de Texto

A manipulação de textos geralmente pressupõe a habilidade de dividi-lo em sentenças individuais (BIRD; KLEIN; LOPER, 2009, p. 112). A segunda atividade, denominada Segmentação de Texto, é justamente o mecanismo que provê a segmentação do conteúdo de texto puro em sentenças. Cada sentença limita um conjunto semântico mínimo para definição de uma proposição. A segmentação de textos é realizada por meio da identificação de caracteres finalizadores de sentenças, principalmente, dos sinais de pontuação. A dificuldade desta atividade está na correta associação dos sinais de pontuação ao fim de sentença, uma vez que tais sinais também podem demarcar abreviações de nomes, separação de dígitos em datas, horas, números de telefones e números ordinais (CIMIANO, 2006).

4.3.3 Tokenização

A terceira atividade é a atividade de identificação de *tokens*, ou seja, de elementos de cada oração, inclusos neste conjunto: palavras, números e sinais de pontuação. Essa etapa, denominada de *Tokenização*, utiliza-se, basicamente, de sinais gráficos

como espaços e algoritmos para o reconhecimento de entidades limítrofes de um *token*.

Segundo Cimiano (2006), o uso dos espaços em branco nem sempre definem de forma adequada as fronteiras das palavras, sobretudo de palavras compostas e de nomes próprios. Por isso, é recomendada a utilização conjunta *tokenizador* juntamente com técnicas de reconhecimento de nomes de entidades.

4.3.4 Análise Lexical

A atividade “*Análise Lexical*” é a quarta atividade deste processo. A atividade consiste em determinar, para cada *token* a etiqueta morfológica correspondente. A etiqueta morfológica, também conhecida como classe morfológica, etiqueta lexical, classe de palavras e POS, identifica um: substantivo, verbo, pronome, preposição, advérbio, conjunção ou artigo (JURAFSKY; MARTIN, 2009, p. 123). Por isso, o *POS Tagger* também é denominado de etiquetador morfológico.

A saída desta atividade é um arquivo contendo *tokens* anotados por suas respectivas etiquetas morfológicas. A dificuldade da atividade está na desambiguação semântica do *token*, pois é bem comum que uma mesma palavra seja classificada em diferentes classes gramaticais, de acordo com o contexto de seu uso.

4.3.5 Reconhecimento de Elementos Centrais

A quinta atividade, “*Reconhecimento de Elementos Centrais*”, tem por objetivo produzir um conjunto de elementos que serão candidatos a conceitos e relações de mapas conceituais. Para isso, recebe como entrada, o conjunto de *tokens* etiquetados morfolologicamente e realiza um processamento utilizando técnicas de *chunking*.

O *chunking* – também conhecido como uma análise sintática parcial ou superficial – aplica técnicas de processamento superficial, tipicamente caracterizados por autômatos finitos e expressões regulares, para manter o agrupamento de palavras, formando constituintes com significados, tipicamente com um núcleo que é modificado por outras palavras da unidade (CIMIANO, 2006). Em português, as

principais sequências de palavras ordenam-se em torno de um nome (Sintagma Nominal), verbo (Sintagma Verbal) ou são precedidas por preposições (Sintagma Preposicional). Esses arranjos são denominados genericamente de “*chunks* linguísticos”.

Segundo Novak e Gowin (1984) é essencial distinguir conceitos de palavras de ligação (relações). Ambos são importantes como unidades de linguagem, mas executam papéis diferentes na transmissão do significado. Enquanto conceitos são rotulados por palavras que representam coisas e eventos, as relações são uma forma de ligar dois conceitos em uma forma proposicional. Não há nenhuma correlação direta entre elementos de mapas conceituais e *chunks* linguísticos. Todavia, de acordo com as orientações de construção de um mapa conceitual, podemos considerar que Sintagmas Nominais (NP), Sintagmas Verbais (VP) e Sintagmas Preposicionais (PP) são candidatos primários ao mapeamento dos elementos centrais de mapas conceituais. O Sintagma Nominal (NP) é o principal candidato ao conceito, o Sintagma Verbal (SV), por sua vez, é um aspirante a se tornar uma relação e o Sintagma Preposicional (PP) pode ser tanto um conceito já associados com a relação, tomando como essência o Sintagma Nominal precedido de preposição, ou uma relação, caso consideremos verbos ou locuções verbais precedidos de preposição.

Cada um dos conceitos reconhecidos na oração é mapeado na forma de nó de grafo a conceito. As relações são mapeadas para nós de grafo candidatos a relações. Há ainda e os *chunks* nominais preposicionados que são nós de grafo candidatos a um conceito já associadas a uma relação.

Os *chunks* identificados são, portanto, mapeados para nós candidatos a conceitos e relações. O mapeamento de um *chunk* para um conceito ou uma relação dependerá, primordialmente, da possibilidade de estabelecer conexões entre estes elementos na forma de proposições.

O *chunking* é uma tarefa importante para o reconhecimento dos candidatos a elementos centrais. Nesta pesquisa, os *chunks* são identificados por um conjunto de padrões linguísticos, cada um deles constituído por expressões regulares. Um

exemplo de padrões linguísticos pode ser visto na Tabela 3 e a relação completa está no Apêndice A.

Tabela 3. Padrões Linguísticos para Reconhecimento de Sintagmas Nominais

#	PADRÕES LINGUÍSTICOS	EXEMPLO
1.	NP: {<N>+}	“PAZ”
2.	NP: {<DA.* DI.* DD.*><NP>}	“A PAZ”
3.	NP: {<NP><QL.*><NP>+}	“SEMANA DA PAZ”
4.	NP: {<DA.* DI.* DD.*><VMP.*>}	“SEMANA PASSADA”
5.	NP: {<FPA><NP><FPT>}	“(PAZ)”
6.	NP: {<NP><NP>+}	“PAZ UNIVERSAL”
7.	NP: {<NP><CC><NP>+}	“AMOR E PAZ”
Legenda da Etiqueta		
NP: Sintagma Nominal N: Substantivo DA: Artigo Definido DI: Artigo Indefinido DD: Demonstrativo	QL: Qualificador VMP: Particípio Passado FPA – Parênteses FPT – Parênteses	.*: Curinga +: Um ou mais <> - Conjunto : Ou

4.3.6 Interpretação de Dependências

A sexta e penúltima atividade é o interpretador de dependências, denominado de “Interpretação de Dependências”. Tomando cada um dos candidatos a elementos de mapas conceituais, um interpretador transforma o candidato a elemento de mapa conceitual em um nó, uma aresta de um grafo, ou um elemento híbrido composto simultaneamente de nó e aresta.

Grafos foram escolhidos como estruturas intermediárias porque sua forma com nós e arcos é similar a mapas conceituais. Além disso, os grafos não têm restrições para lidar com relações binárias. Deste ponto de vista, grafos são mais flexíveis do que mapas conceituais, considerando que no segundo existe o requisito de produção das proposições. Como podemos observar, em grafos, o relacionamento pode também ser representado como um nó da mesma forma de um conceito. A transformação de nós de grafo em elementos de mapas conceituais é um estágio importante na construção do mapa conceitual.

Na atividade, verifica-se o tipo do *chunk*: aqueles com núcleos verbais ou preposicionais são mapeados em arestas e os *chunks* com núcleos nominais são mapeados em nós. À medida que os elementos são transformados em nós ou arestas, o interpretador busca a posição mais adequada no grafo para subsumir o

novo elemento. A subsunção ocorre por meio da identificação da proximidade dos nós, unindo aqueles que possuem maior afinidade, de acordo com regras de aproximação pré-definidas.

4.3.7 Construção de Mapas Conceituais

A última atividade do processo é aquela que consiste na definição de proposições, baseadas em um grafo de entrada, denominada de “*Construção de Mapas Conceituais*”. Para isso, caminha-se no grafo, de forma a encontrar proposições de mapas conceituais. A cada caminho que parte de um conceito, passando por uma relação que leva a outro conceito, é criada uma proposição. Ao fim da atividade, um conjunto de proposições é formado, determinando a representação do conteúdo original por meio de conexões em um mapa conceitual.

4.4 Considerações Finais

O uso de métodos linguísticos associados às técnicas de inteligência artificial para a construção de mapas conceituais a partir de textos não é uma prática inédita. Na realidade, Valério e Leake (2006) também utilizaram Sintagmas Nominais e Verbais para a extração de conceitos e relações. Pérez e Vieira (2004, 2005) propuseram o uso de *chunks*, de natureza sintática, tais como sujeitos, predicados e objetos para mapeá-los em elementos centrais de mapas por meio de um conjunto de regras de composição. Alves e outros (2001) usaram métodos linguísticos combinados com métodos de aprendizagem de máquina para criar predicados binários baseados em recursos léxicos externos. Lau e outros (2008) fizeram uso de conteúdo de fóruns e chats para extrair mapas conceituais por meio da aplicação de padrões linguísticos e métodos estatísticos para computar a relevância de cada conceito. Richardson e Fox (2007) e Richardson, Srinivasane Fox (2008) extraíram proposições de teses e dissertações por meio de ferramentas de manipulação de textos e extração de informações. Zouaq e Nkambou (2008, 2009) extraíram conceitos chaves de textos em linguagem natural através de métodos linguísticos e de aprendizagem de máquina.

As diferenças entre as abordagens mencionadas e a abordagem presente neste trabalho estão relacionadas a seguir:

- Trabalhamos com um único texto em contrapartida às abordagens de Valerio e Leake (2006) e Zouaq e Nkambou (2008, 2009) que recuperam informações de um conjunto de textos;
- Propomos o uso de documentos com conteúdo textual como a entrada para a construção de mapas conceituais, diferentemente de Pérez e Vieira (2008, 2009) que utilizam textos pré-etiquetados;
- Não ignoramos “*stop words*” – na verdade, consideramos que elas são muito importantes para definir a dependência entre conceitos e relacionamentos –, tais como preposições e sinais de pontuação, ao contrário do que é feito por Lau e outros. (2008);
- Objetivamos a produção de mapas conceituais compreensíveis para leitores humanos e o mapa conceitual construído é fim de nossa pesquisa, diferentemente de Zouaq e Nkambou (2008, 2009) que usam os mapas conceituais como linguagem intermediária para a construção de ontologias para manipulação computacional;
- Não restringimos o conjunto de palavras de ligações, ou relações, em oposição à abordagem de Alves, Pereira e Cardoso (2001)

No próximo capítulo, definimos um estudo de caso a partir da abordagem por meio do desenvolvimento de um modelo prototípico.

5 ESTUDO DE CASO

“A resposta certa, não importa nada: o essencial é que as perguntas estejam certas.”

Mário Quintana
“As Indagações”

Ao longo do Capítulo 4, apresentamos as bases teóricas de uma abordagem computacional para a construção de mapas conceituais a partir de textos em português do Brasil. Neste Capítulo, relatamos a condução de experimentos baseados na abordagem proposta.

Iniciamos pela Seção 5.1, em que descrevemos as especificações técnicas do protótipo computacional denominado de *Text2Cmap*. O *Text2Cmap* é composto por um conjunto de módulos que materializam as atividades do processo de construção de mapas conceituais, e, por isso, figura como o principal componente dos experimentos.

Para assegurar o entendimento do funcionamento da ferramenta *Text2Cmap* no contexto da abordagem, detalhamos na Seção 5.2, as saídas intermediárias resultantes do processamento e explicamos como cada uma delas corrobora para a construção do mapa conceitual.

Na Seção 5.3, relatamos a experiência de uso do *Text2Cmap* com uma amostra composta por documentos do corpus linguístico Mac-Morpho. Nesse ensaio, coletamos os dados quantitativos apresentados na Seção 5.4 e analisados na Seção 5.5.

5.1 Protótipo Computacional

Com o intuito de validar a abordagem proposta no Capítulo 4, desenvolvemos um protótipo computacional denominado de *Text2Cmap*.

Implementada em *Python* 2.6.5 [PYTHON, 2010], a versão alfa do *Text2Cmap* é composta por cinco módulos que realizam as atividades do modelo de processo proposto, por meio do mapeamento detalhado na Figura 7.

As atividades de “*Segmentação de Texto*” e “*Tokenização*” são atendidas pelo módulo *Tokenizer*; a atividade de “*Análise Léxica*”, pelo módulo *Tagger*, a atividade “*Reconhecimento de Elementos Centrais*” pelo *Chunker*, a atividade “*Interpretação de Dependências*” pelo *GraphBuilder* e atividade “*Construção de Mapas Conceituais*” pelo módulo *CmapBuilder*. Devido à necessidade de adequar o escopo do protótipo ao cronograma da pesquisa, optamos por suprimir da primeira versão do protótipo o módulo para realização da atividade “*Normalização de Texto*”, que seria responsável pela retirada de marcas de formatação. Entendemos que a exclusão, embora estabeleça a restrição para o processamento de conteúdos que não estejam em texto puro, não impactam diretamente o resultado de nossas pesquisas. Em futuras versões, pretendemos inserir esta funcionalidade.

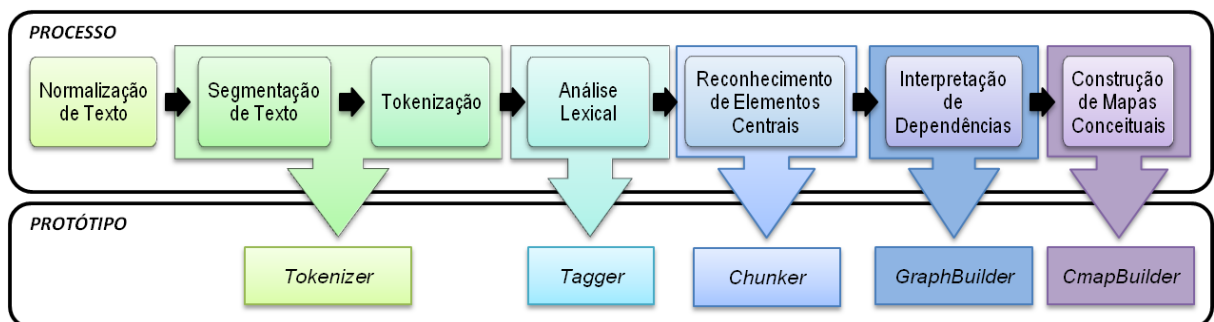


Figura 7. Módulos do Protótipo e Atividades do Processo

5.1.1 Módulo *Tokenizer*

O módulo *Tokenizer* realiza duas atividades de forma sequencial: “*Segmentação de Texto*” e “*Tokenização*”. Primeiramente, o *Tokenizer* realiza a leitura do documento,

em texto puro, segmentando o conteúdo em sentenças. Em seguida, o *Tokenizer* identifica as unidades que compõem o texto como palavras, números, sinais – mais conhecidos como *tokens*.

A definição de sentenças e a identificação de *tokens* são abstrações de um mesmo problema: o problema de segmentação. Enquanto a primeira busca reconhecer os limites da sentença, a segunda procura determinar os limites do *token*. É por considerar este fato que implementamos no *Text2Cmap* uma única solução para endereçar ambas as atividades.

A segmentação de sentenças baseia-se no reconhecimento de marcadores de fim de sentença, tais como pontos finais e marcas de parágrafo. A identificação de *tokens* requer um conjunto de padrões mais refinados, que inclui a definição de expressões regulares para reconhecimento de fronteiras de *tokens* que designam datas, horas, moedas, números, dentre outros. A implementação foi realizada por meio de recursos do módulo de processamento de linguagem natural chamado de *Natural Language Processing Toolkit (NLTK)* (BIRD; KLEIN; LOPER, 2009), uma poderosa biblioteca que, dentre as diversas funções, permite a segmentação e tokenização a partir da definição de padrões linguísticos, conforme exemplificados na Tabela 4, com o intuito de gerar uma lista de *tokens*.

Tabela 4. Exemplos de padrões linguísticos baseados em expressões regulares

Padrões Linguísticos	Expressão Regular
Abreviações. Ex.: “S.A.”	<code>([A-Z]+\.)+</code>
Palavras compostas. Ex.: “beija-flor”	<code>\w+(\w+)*(-\w+)*</code>
Número, moeda, percentual	<code>[A-Z]{0,}\\$?\s?\d+([,.\d+)?\s%??</code>
Sinais	<code>[] [. , ; " ' ? () : - _ `]</code>

5.1.2 Módulo *Tagger*

O módulo *Tagger* foi desenvolvido para associar etiquetas morfológicas para cada *token*, realizando a atividade conhecida como “*POS Tagging*”, ou etiquetagem morfológica. Para a atividade de etiquetagem morfológica, optamos pelo uso do software livre *Freeling* (Atserias, 2006), com as configurações para a Língua Portuguesa de Garcia e Gamallo (2010). O *Freeling* é uma ferramenta de *tagger*, que aplica técnicas estatísticas para a identificação de etiquetas e o reconhecimento

de entidades nomeadas. O *tagger Freeling* é evocado a partir do módulo *Tagger*, recebendo como entrada o arquivo de *tokens* e o arquivo de parâmetros que define o modo de operação para a atividade de *tagging*. Após execução, o módulo *Tagger* produz como saída o conjunto de *tokens* seguidos por suas respectivas etiquetas morfológicas, lemas e probabilidades. Desta saída, geramos outra, mais simplificada, contendo apenas os *tokens* e as etiquetas.

5.1.3 Módulo *Chunker*

O módulo *Chunker* reconhece grupos de *tokens* a partir um conjunto de expressões regulares descritas por meio de cadeias de etiquetas morfológicas. Por meio do processo de *chunking* são identificados os candidatos a elementos de mapas conceituais.

A princípio, o protótipo foi desenvolvido para reconhecer 16 tipos de elementos. Os elementos de núcleo nominal são: Sintagma Nominal (NP), Sintagma Nominal precedido de conjunção coordenativa (CCNP), Sintagma Nominal precedido de pronome relativo (PRNP), Sintagma Nominal precedido de conjunção subordinativa (CSNP), Sintagma Nominal precedido de preposição (SPNP). Os elementos de núcleo verbal são: Sintagma Verbal (SV), Sintagma Verbal precedido de preposição (SPVP), Sintagma Verbal precedido de conjunção coordenativa (CCVP), Sintagma Verbal precedido de pronome relativo (PRVP), Sintagma Verbal precedido de conjunção subordinativa (CSVP). Além destes, há os elementos formados por sinais de pontuação, os advérbios isolados (RP) e os elementos que não são reconhecidos pelos padrões definidos (NA).

A Tabela 5 lista cada um dos elementos reconhecidos neste módulo, seguido de exemplos que auxiliam no entendimento de cada *chunk*.

Cada elemento é reconhecido graças a um conjunto de padrões linguísticos que definem uma gramática para identificação de elementos candidatos. A relação completa dos padrões linguísticos está no Apêndice A.

Tabela 5. Elementos reconhecidos pelo *Chunker*

Elemento	Descrição	Exemplo
NP	Sintagma Nominal	“o secretário”...
CCNP	Sintagma Nominal precedido de conjunção coordenativa	“e o secretário”...
PRNP	Sintagma Nominal precedido de pronome relativo	“que a nova lei”
CSNP	Sintagma Nominal precedido de conjunção subordinativa	[disseram] “que o grupo”
QLNP	Sintagma Nominal precedido de preposição “de”	“do ministro”
SPNP	Sintagma Nominal precedido de preposição	“para a fabricação”
VP	Sintagma Verbal	“foi selecionada”
SPVP	Sintagma Verbal precedido de preposição	“para cobrar”
CCVP	Sintagma Verbal precedido de conjunção coordenativa	“e alterava”
PRVP	Sintagma Verbal precedido de pronome relativo	[ele] “que se aposentou”
CSVP	Sintagma Verbal precedido de conjunção subordinativa	[dizem] “que a Portaria 1.275 revelou”
RP	Advérbio	“assim”
VRG	Vírgula	“ ” ,
PNT	Ponto final, ponto de interrogação, ponto de exclamação	“ ” .
DPT	Dois pontos	“.” .
PTV	Ponto e vírgula	“.” ,

5.1.4 Módulo GraphBuilder

O módulo *GraphBuilder* executa a atividade de construção do grafo que intermedeia a construção do mapa conceitual. Para isso, o módulo toma cada um dos elementos reconhecidos pelo *chunker*, denominados de “candidatos” e os transforma em nós de grafos, os quais são classificados como “nós conceitos”, “nós híbridos” ou “nós link”. O nó conceito é um candidato a conceito de mapas conceituais, o nó link, um candidato a uma relação e o nó híbrido, um misto entre conceitos e relações. A regra de transformação dos elementos para correspondentes em grafos está definida na Tabela 6.

A composição do grafo é realizada por meio de um algoritmo de busca baseado em uma função heurística. O módulo caminha pelo grafo, a partir de um nó folha até encontrar um nó candidato a ser o subsunçor. A escolha do nó folha leva em consideração o peso e a distância do nó a ser subsumido em relação ao nó pai.

Quanto menor a distância, maior a probabilidade de existência de relacionamento entre os dois elementos.

Seguindo regras de compatibilidade, o algoritmo realiza a subsunção quando identificada a possibilidade aproximação sintática entre os dois nós.

Tabela 6. Correspondência de *chunk* para elemento de grafo

<i>Elemento</i>	<i>Significado</i>	<i>Correspondente no Grafo</i>
NP	Sintagma Nominal	Nó conceito
CCNP	Sintagma Nominal precedido de conjunção coordenativa	Nó híbrido
PRNP	Sintagma Nominal precedido de pronome relativo	
CSNP	Sintagma Nominal precedido de conjunção subordinativa	
QLNP	Sintagma Nominal precedido de preposição “de”	
SPNP	Sintagma Nominal precedido de preposição	
VP	Sintagma Verbal	Nó Link
SPVP	Sintagma Verbal precedido de preposição	
CCVP	Sintagma Verbal precedido de conjunção coordenativa	
PRVP	Sintagma Verbal precedido de pronome relativo	
CSVP	Sintagma Verbal precedido de conjunção subordinativa	
RP	Advérbio	Não mapeado
VRG	Vírgula	
PNT	Ponto final, ponto de interrogação, ponto de exclamação	
DPT	Dois pontos	
PTV	Ponto e vírgula	
NA	Não identificado	

A proximidade sintática é calculada por meio de pesos binários que determinam a possibilidade de ligação de dois nós. A relação completa das regras de proximidade sintática está no Apêndice A.

A geração de arquivos de grafos compatíveis com Graphviz (GRAPHVIZ, 2010), no formato “dot”, é feita por meio do módulo GvGen (GVGEN, 2007]e a visualização é feita pela ferramenta “GVEdit” do próprio pacote Graphviz. Uma vez que o arquivo “dot” é gerado, usamos os comandos disponíveis no Graphviz para automaticamente gerar uma imagem do grafo construído.

5.1.5 Módulo CmapBuilder

O módulo *CmapBuilder* realiza a atividade fim do protótipo, ou seja, caminha pelo grafo, buscando a extração de proposições. A proposição é formada sempre que for possível caminhar de um nó conceito a um nó híbrido, ou de um nó conceito, passando por um nó link e chegando a um nó conceito.

No fim do processo, há um conjunto de proposições formadas cuja organização gráfica leva ao mapa conceitual. No protótipo, não há módulo de visualização, por isso, lançamos mão da ferramenta *CmapTools* (NOVAK; CAÑAS, 2006). Pesquisas estão em andamento com o intuito de prover o módulo de visualização de mapas para o *Text2Cmap*.

5.2 Exemplo

Nesta Seção, simulamos a execução do *Text2Cmap* tomando por entrada um documento exemplo. Antes de iniciarmos a simulação, é importante ter em mente que o processamento de um documento qualquer denominado de “[documento].txt” leva à produção de um conjunto de saídas, resultados intermediários do processo, enumeradas na Tabela 7.

Tabela 7. Saídas produzidas pelos módulos de Text2Cmap

Módulo	Saídas
Tokenizer	[documento]_token.txt
Tagger	[documento]_tag.tmp
	[documento]_tag.txt
Chunker	[documento]_chunk.txt
GraphBuilder	[documento]_dot.txt
	[documento].gif
CmapBuilder	[documento]_cmap.txt
	[documento].cmap

É por meio destas saídas que o processo subjacente ao protótipo permite o acompanhamento e a rastreabilidade das ações realizadas. Vejamos como são

produzidas as evidências do processo, por meio do processamento do documento *fc96ma23_0013.txt*, selecionado como exemplo de entrada para o protótipo.

O conteúdo do documento selecionado está ilustrado na Figura 8.

```
O governador Mário Pereira , de o Paraná , e o secretário de a agricultura
José Carlos Tibúrcio aproveitaram o palanque de a Exposição Agropecuária de
Londrina para cobrar de o ministro de a Agricultura , Synval Guazzelli , um
novo modelo de política agrícola , capaz de estimular investimentos e o
crescimento de a produção .
```

Figura 8. Documento de entrada para o protótipo

O protótipo realiza a leitura do documento e em seguida, evoca o módulo *Tokenizer*, que produz o arquivo de saída denominado *fc96ma23_0013_token.txt*, como na Figura 9.

```
O[ENTER] governador[ENTER] Mário[ENTER] Pereira[ENTER] ,[ENTER] de[ENTER] o[ENTER]
Paraná[ENTER] ,[ENTER] e[ENTER] o[ENTER] secretário[ENTER] de[ENTER] a[ENTER]
agricultura[ENTER] José[ENTER] Carlos[ENTER] Tibúrcio[ENTER] aproveitaram[ENTER]
o[ENTER] palanque[ENTER] de[ENTER] a[ENTER] Exposição[ENTER] Agropecuária[ENTER]
de[ENTER] Londrina[ENTER] para[ENTER] cobrar[ENTER] de[ENTER] o[ENTER]
ministro[ENTER] de[ENTER] a[ENTER] Agricultura[ENTER] ,[ENTER] Synval[ENTER]
Guazzelli[ENTER] , [ENTER] um[ENTER] novo[ENTER] modelo[ENTER] de[ENTER]
política[ENTER] agrícola[ENTER], capaz[ENTER] de[ENTER] estimular[ENTER]
investimentos[ENTER] e[ENTER] o[ENTER] crescimento[ENTER] de[ENTER] a[ENTER]
produção[ENTER] .[ENTER]
```

Figura 9. Saída do Tokenizer

Tomando como entrada o arquivo contendo os *tokens*, o protótipo processa, por meio do módulo *Tagger*, cada um dos elementos, produzindo o arquivo intermediário denominado *fc96ma23_0013_tag.tmp*, conforme Figura 10. O arquivo intermediário contém o resultado produzido pelo etiquetador morfológico, neste caso, o *tagger Freeling* para a Língua Portuguesa. No arquivo de saída, é possível perceber que o etiquetador apresenta, para cada *token*, o lema, a etiqueta morfológica correspondente e um número referente à probabilidade estatística da etiqueta morfológica designada.

```
O o DAOMS0 0.964394[ENTER] governador governador NCMS000 0.9375[ENTER]
Mário_Pereira mário_pereira NP00000 1[ENTER] , , Fc 1[ENTER] de de SPS00
1[ENTER] o o DAOMS0 0.964394[ENTER] Paraná paraná NP00000 1[ENTER] , , Fc
1[ENTER] e e CC 0.999818[ENTER] o o DAOMS0 0.964394[ENTER] secretário
secretário NCMS000 1[ENTER] de de SPS00 1[ENTER] a o DAOFS0
0.675226[ENTER] agricultura agricultura NCFS000 1[ENTER]
José_Carlos_Tibúrcio josé_carlos_tibúrcio NP00000 1[ENTER] aproveitaram
aproveitar VMIM3P0 0.5[ENTER] o o DAOMS0 0.964394[ENTER] palanque palanque
NCMS000 0.948425[ENTER] de de SPS00 1[ENTER] a o DAOFS0 0.675226[ENTER]
Exposição_Agropecuária_de_Londrina exposição_agropecuária_de_londrina
NP00000 1[ENTER] para para SPS00 0.999302[ENTER] cobrar cobrar VMN0000
0.3[ENTER] de de SPS00 1[ENTER] o o DAOMS0 0.964394[ENTER] ministro
ministro NCMS000 0.992308[ENTER] de de SPS00 1[ENTER] a o DAOFS0
0.675226[ENTER] Agricultura agricultura NP00000 1[ENTER] , , Fc 1[ENTER]
Synval_Guazzelli synval_guazzelli NP00000 1[ENTER] , , Fc 1[ENTER] um um
DIOMS0 0.869164[ENTER] novo novo AQOMS0 0.993976[ENTER] modelo modelo
NCMS000 0.954545[ENTER] de de SPS00 1[ENTER] política política NCFS000
0.314815[ENTER] agrícola agrícola AQOCS0 1[ENTER] , , Fc 1[ENTER] capaz
capaz AQOCS0 1[ENTER] de de SPS00 1[ENTER] estimular estimular VMN0000
0.314286[ENTER] investimentos investimento NCMP000 1[ENTER] e e CC
0.999818[ENTER] o o DAOMS0 0.964394[ENTER] crescimento crescimento NCMS000
1[ENTER] de de SPS00 1[ENTER] a o DAOFS0 0.675226[ENTER] produção produção
NCFS000 1[ENTER] . . Fp 1[ENTER]
```

Figura 10. Trecho da Saída Intermediária do Tagger

Dos elementos produzidos pelo *Tagger*, aqueles que realmente importam para o protótipo são as etiquetas morfológicas. Por isso, o arquivo temporário é transformado em um arquivo *fc96ma23_0013_tag.txt* que contém somente os dados de interesse à pesquisa, no formato separado por tabulações (Figura 11).

```
O/DAOMS0[TAB] governador/NCMS000[TAB] Mário_Pereira/NP00000[TAB],/Fc[TAB]
de/SPS00[TAB] o/DAOMS0[TAB] Paraná/NP00000[TAB] ,/Fc[TAB] e/CC[TAB] o/DAOMS0[TAB]
secretário/NCMS000[TAB] de/SPS00[TAB] a/DAOFS0[TAB] agricultura/NCFS000[TAB]
José_Carlos_Tibúrcio/NP00000[TAB] aproveitaram/VMIM3P0[TAB] o/DAOMS0[TAB]
palanque/NCMS000[TAB] de/SPS00[TAB] a/DAOFS0[TAB]
Exposição_Agropecuária_de_Londrina/NP00000[TAB] para/SPS00[TAB] cobrar/VMN0000[TAB]
de/SPS00[TAB] o/DAOMS0[TAB] ministro/NCMS000[TAB] de/SPS00[TAB] a/DAOFS0[TAB]
Agricultura/NP00000[TAB] ,/Fc[TAB] Synval_Guazzelli/NP00000[TAB] ,/Fc[TAB]
um/DIOMS0[TAB] novo/AQOMS0[TAB] modelo/NCMS000[TAB] de/SPS00[TAB]
política/NCFS000[TAB] agrícola/AQOCS0[TAB] ,/Fc[TAB] capaz/AQOCS0[TAB] de/SPS00[TAB]
estimular/VMN0000[TAB]
investimentos/NCMP000[TAB] e/CC[TAB] o/DAOMS0[TAB] crescimento/NCMS000[TAB]
de/SPS00[TAB] a/DAOFS0[TAB] produção/NCFS000[TAB] ./Fp[TAB]
```

Figura 11. Saída Definitiva do Tagger

A partir de um conjunto de padrões linguísticos criados com base nas etiquetas morfológicas, o *Chunker* identifica os grupos linguísticos denominados de *chunks* e produz como saída o arquivo *fc96ma23_0013_chunk.txt*, com o conteúdo da Figura 12.

```
(S
  (NP O/DA0MS0 (NP governador/NCMS000 Mário_Pereira/NP00000))
  (VRG ,/FC)
  (QLNP de/QLS00 (NP o/DA0MS0 (NP Paraná/NP00000)))
  (VRG ,/FC)
  (CCNP
    e/CC
    (NP
      (NP o/DA0MS0 (NP secretário/NCMS000))
      de/QLS00
      (NP
        a/DA0FS0
        (NP agricultura/NCFS000 José_Carlos_Tibúrcio/NP00000))))
  (VP aproveitaram/VMIM3P0)
  (NP
    (NP o/DA0MS0 (NP palanque/NCMS000))
    de/QLS00
    (NP a/DA0FS0 (NP Exposição_Agropecuária_de_Londrina/NP00000)))
  (SPVP para/SPS00 (VP cobrar/VMN0000))
  (QLNP
    de/QLS00
    (NP
      (NP o/DA0MS0 (NP ministro/NCMS000))
      de/QLS00
      (NP a/DA0FS0 (NP Agricultura/NP00000))))
  (VRG ./FC)
```

Figura 12. Trecho da Saída do Chunker

A árvore de *chunks* é normalizada, por meio de transformação de *chunks* em elementos de grafo. O resultado desta etapa está na Figura 13 que apresenta o arquivo *fc96ma23_0013_dot.txt*, que contém a estrutura do grafo resultante do processamento.

```
digraph G {
  compound=true;
  node1 [shape="rectangle",label="O governador Mário_Pereira"];
  node2 [shape="rectangle",label="o Paraná"];
  node3 [shape="rectangle",label="o secretário de a agricultura
José_Carlos_Tibúrcio"];
  node4 [shape="none",label="aproveitaram"];
  node5 [shape="rectangle",label="o palanque de a
Exposição_Agropecuária_de_Londrina"];
  node6 [shape="none",label="para cobrar"];
  node7 [shape="rectangle",label="o ministro de a Agricultura"];
  node8 [shape="rectangle",label="Synval_Guazzelli"];
  node9 [shape="rectangle",label="um novo modelo de política agrícola"];
  node10 [shape="none",label="capaz de estimular"];
  node11 [shape="rectangle",label="investimentos e o crescimento de a
produção"];
  node1->node2 [label="de"];
  node1->node3 [label="e"];
  node3->node4;
```

Figura 13. Trecho da Saída do GraphBuilder – estrutura do grafo

O arquivo de estrutura do grafo é, por sua vez, processado para a produção do arquivo com a imagem do grafo, *fc96ma23_0013.gif*. Este passo permite uma avaliação visual do grafo produzido, como na Figura 14.

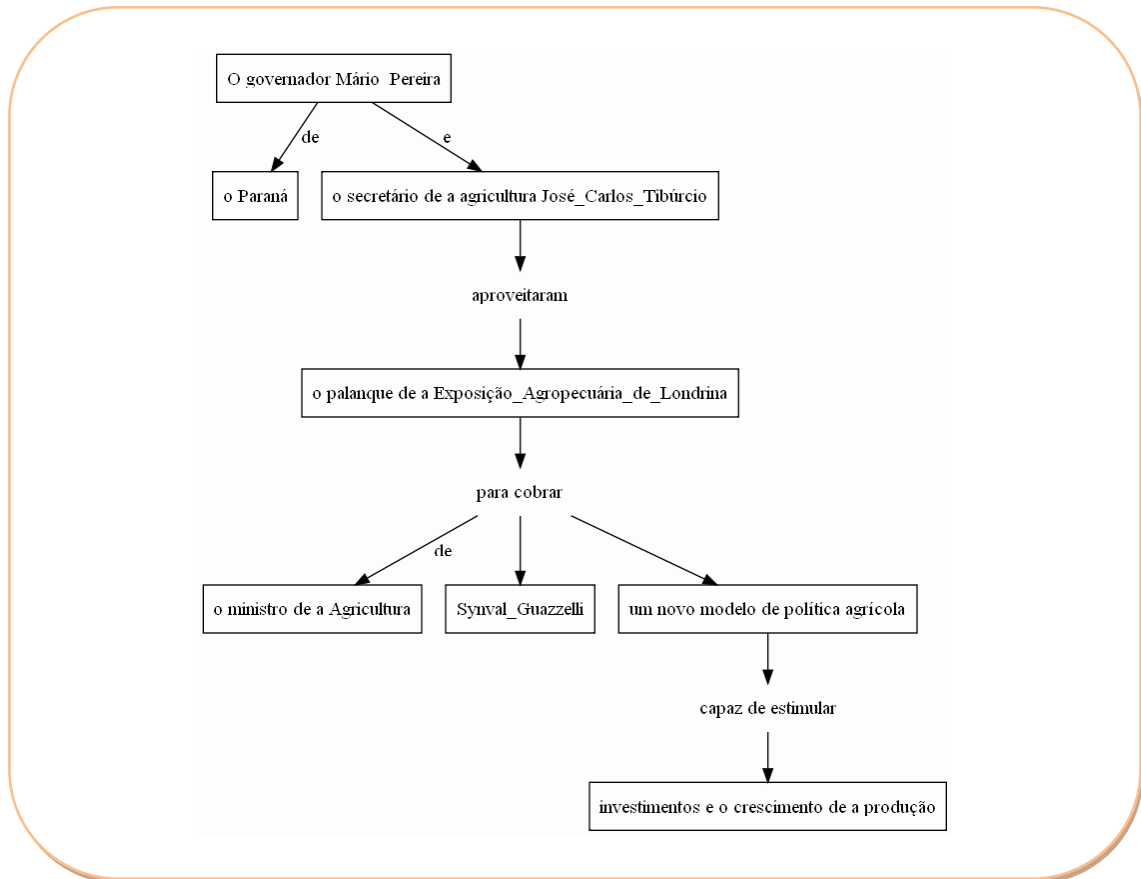


Figura 14. Saída do GraphBuilder – imagem do grafo

A última etapa consiste em caminhar sobre o grafo, identificando as proposições na forma de triplas. O arquivo de proposições é criado com o nome de *fc96ma23_0013_cmap.txt* (Figura 15).

```

O governador Mário_Pereira[TAB] de[TAB] o Paraná
O governador Mário_Pereira[TAB] e[TAB] o secretário de a agricultura
José_Carlos_Tibúrcio
o secretário de a agricultura José_Carlos_Tibúrcio[TAB] aproveitaram[TAB] o
palanque de a Exposição_Agropecuária_de_Londrina
o palanque de a Exposição_Agropecuária_de_Londrina[TAB] para cobrar de[TAB]
o ministro de a Agricultura
o palanque de a Exposição_Agropecuária_de_Londrina[TAB] para cobrar[TAB]
Synval_Guazzelli
o palanque de a Exposição_Agropecuária_de_Londrina[TAB] para cobrar[TAB] um
novo modelo de política agrícola
um novo modelo de política agrícola[TAB] capaz de estimular [TAB]
investimentos e o crescimento de a produção
  
```

Figura 15. Saída do CmapBuilder – Proposições do Mapa Conceitual

Com o auxílio da ferramenta *CmapTools*, as proposições são transformadas em mapas conceituais, cujo arquivo é nomeado de *fc96ma23_0013.cmap*.

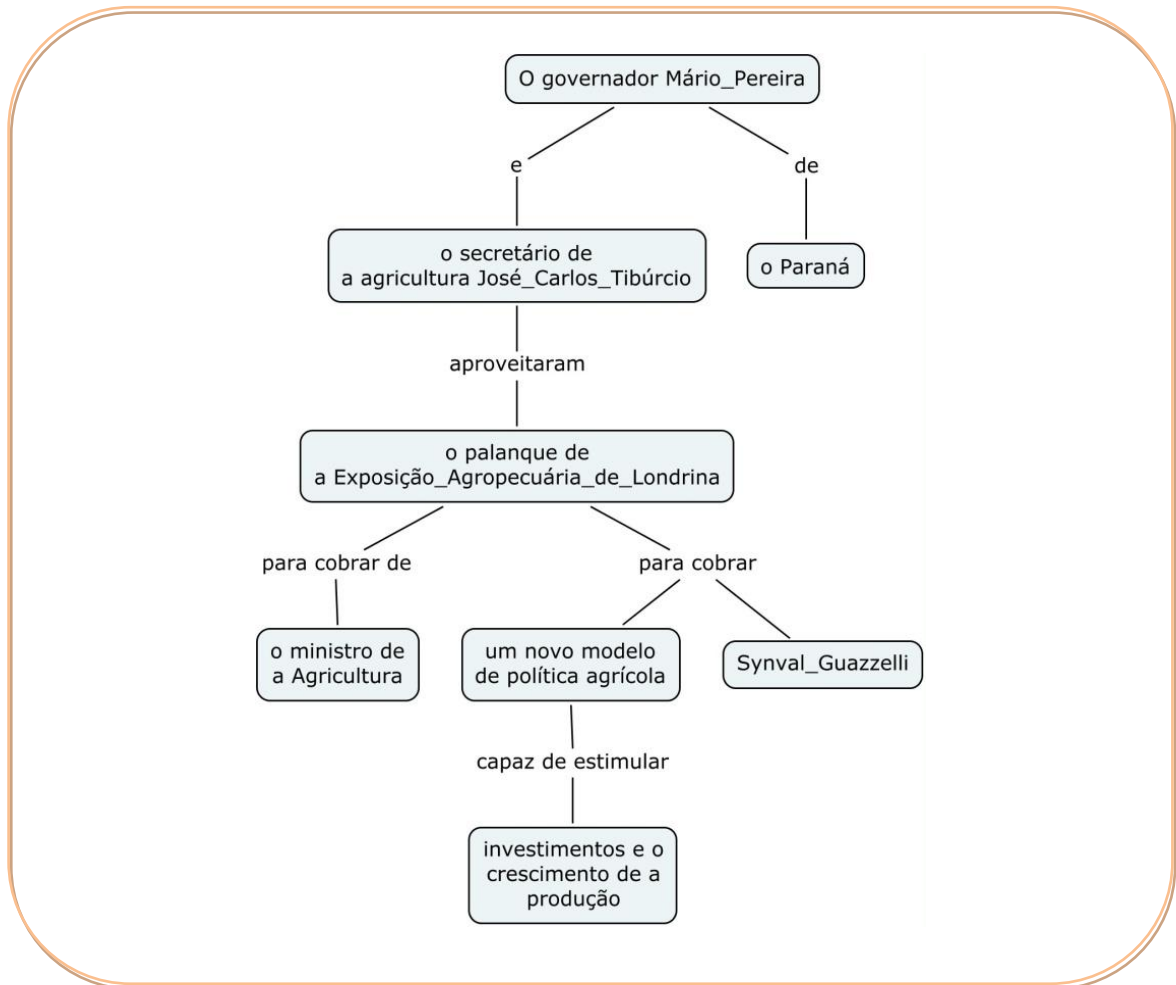


Figura 16. Saída do CmapBuilder – Imagem do Mapa Conceitual

5.3 Experimentação com Corpus

Nesta Seção, apresentamos a experimentação realizada com o corpus Mac-Morpho. Os experimentos foram conduzidos por meio do protótipo computacional, baseado em uma amostra, com o intuito de produzir dados quantitativos que permitissem uma análise do processo.

5.3.1 Mac-Morpho

O Mac-Morpho é um corpus fechado e anotado¹², em português do Brasil, desenvolvido no contexto do projeto Lácio-Web (ALUÍSIO et al., 2004).

Este recurso linguístico foi formado por artigos jornalísticos retirados da Folha de São Paulo, ano 1994, dos cadernos: “esporte” (ES), “dinheiro” (DI), “ciência” (FC), “agronomia” (AG), “informática” (IF), “ilustrada” (IL), “mais!” (MA), “mundo” (MU), “brasil” (BR) e “cotidiano” (CO).

Mac-Morpho contém aproximadamente 1,2 milhões de palavras que foram etiquetadas pelo parser Palavras¹³ e, posteriormente, cada etiqueta foi mapeada para o elemento correspondente no conjunto de etiquetas do Projeto Lácio-Web. A anotação morfossintática foi validada manualmente. O corpus é composto por 109 arquivos em texto puro, que por sua vez traduzem-se em 51.397 sentenças.

5.3.2 Definição da Amostra

Do corpus Mac-Morpho, utilizamos uma amostra, denominada de *A1*, que contém aproximadamente 2% do volume total dos textos, composta por 1000 sentenças selecionadas de forma aleatória dos extratos “agronomia” (AG), “ciência” (FC) e “brasil” (BR). O extrato “agronomia” contribui com 324 sentenças, o extrato “Brasil” com 583 e o extrato “ciência”, com 93 sentenças, que representam, respectivamente, 32,4%, 58,3% e 0,93% da amostra *A1*.

Cada uma das sentenças selecionadas constitui o conteúdo de um único documento, identificado pelo nome do texto que lhe deu origem, seguido de um sequencial numérico. Por exemplo, do extrato de ciências, denominado, “*fc96ma23*”, extraíram-se 93 documentos, os quais foram enumerados de “*fc96ma23_0000*” a “*fc96ma23_0092*”.

¹² “Fechado” no sentido de um corpus concluído e que não sofrerá adição de novos textos; anotado pois apresenta etiquetas morfossintáticas junto a cada token.

¹³ <http://visl.hum.sdu.dk/>

5.3.3 Resultados do Experimento

O *Text2Cmap* versão α processou a amostra *A1* resultando no reconhecimento dos elementos descritos na Tabela 8. O processamento da amostra possibilitou a identificação de 21.917 *tokens*, média de 21,92 *tokens* por documento, com desvio padrão de 11,27. Dos 8.119 *chunks* reconhecidos a partir dos *tokens*, 3.855, que representa aproximadamente 47,48% são organizadas em torno de um nome e 1.916, 23,59%, em torno de um verbo, constituindo uma relação média de 3,86 *chunks* nominais para 1,92 *chunks* verbais em cada documento. Os *chunks* exclusivamente formados por sinais de pontuação e elementos não reconhecidos representam 3.191 *chunks* e, compõem juntamente com os 98 advérbios isolados, um grupo que será descartado para a construção do mapa conceitual, no total de 3,29% do total de *chunks*.

Tabela 8. Quantitativos e Médias dos Elementos Identificados

Módulo	Elementos	CONTAGEM			MÉDIA (POR MIL)			DESVIO PADRÃO		
		Individual	Grupo	Total	Individual	Grupo	Geral			
Tokenizer	<i>Token</i>	21.917	21.917	21.917	21,92	21,92	21,92	11,27		
Chunker	<i>NP</i>	2.260	3.855	8.119	2,26	3,86	8,12	1,38		
	<i>CCNP</i>	64			0,06			0,25		
	<i>PRNP</i>	25			0,03			0,16		
	<i>CSNP</i>	26			0,03			0,16		
	<i>QLNP</i>	203			0,20			0,48		
	<i>SPNP</i>	1.277			1,28			1,07		
	<i>VP</i>	1.320			1,32			0,69		
	<i>SPVP</i>	186	0,19		0,42					
	<i>CCVP</i>	75	1.916		0,08	1,92		0,28		
	<i>PRVP</i>	188	1.916		0,19	1,92		0,46		
	<i>CSVP</i>	147			0,15			0,41		
	<i>RP</i>	98			98			0,10	0,10	0,31
	<i>VRG</i>	1.183			3.191			1,18	3,19	1,28
	<i>PNT</i>	1.039						1,04		0,24
	<i>DPT</i>	27	0,03			0,16				
	<i>PTV</i>	1	0,00			0,03				
	<i>NA</i>	941			0,94			1,46		
GraphBuilder	<i>Nó conceito</i>	2.212	3.673	5.305	2,21	3,67	5,31	1,42		
	<i>Nó híbrido</i>	1.461			1,46			1,22		
	<i>Nó link</i>	1.632			1,63			1,05		
CmapBuilder	<i>Conceito</i>	3.485	6.060	6.060	3,49	6,06	6,06	2,13		
	<i>Relação</i>	2.346			2,35			1,65		

Dos 8.119 *chunks* identificados, o módulo *GraphBuilder* reconheceu 5.305 elementos de grafos, entre nós conceitos e nós link. Os nós conceito representam 69,23% dos elementos reconhecidos, enquanto os nós aresta representam 30,77%. A média encontrada foi de 3,67 nós conceito e de 1,63 de nós aresta por

documento. Os 1.461 nós híbridos podem se transformar em um conceito associado a uma relação. Daí resulta que a soma dos elementos de grafo a serem trabalhados na forma de mapas conceituais totaliza 2.212 referentes a nós conceitos, 1.632 de nós link e 1.461 nós híbridos.

O módulo *CmapBuilder* foi capaz de identificar 3.485 conceitos e 2.346 relações, cerca de 6,06 elementos em um mapa conceitual, contendo em média 3,49 conceitos e 2,35 relações.

A Tabela 9 ilustra que foi possível construir um total de 2.575 proposições, média de 2,57 proposições de cada documento. Do total de *tokens* existentes no documento original, descartamos os sinais de pontuação (vírgula, ponto e vírgula e dois pontos), e encontramos o número de *tokens* considerados válidos.

A relação entre o total de *tokens* válidos existentes nos documentos originais e o número de *tokens* do mapa conceitual construído apresenta uma redução global de 14,35% após a execução do processo. Cada proposição é composta de aproximadamente 6,54 *tokens*.

Tabela 9. Quantitativo de tokens do mapa conceitual comparado ao documento original

<i>Documento</i>			<i>Mapa Conceitual</i>			<i>Redução de tokens no processo (%)</i>
<i>Total de Tokens</i>	<i>Descarte</i>	<i>Tokens Válidos</i>	<i>Tokens</i>	<i>Proposições</i>	<i>Média de tokens por Proposição (%)</i>	
21.917	2.250	19.667	16.845	2.575	6,54	14,35%

A redução global de 14,35% observada a partir dos resultados do processamento não reflete um comportamento uniforme da amostra. Dos 1.000 documentos processados, pode-se encontrar tanto a redução parcial de *tokens* em 44,90% da amostra, quanto a redução total, impedindo a formação de mapas em 12,30%; em 26,20% manteve-se o quantitativo de *tokens* da fonte original e em 16,60% houve um crescimento no número de *tokens*.

Portanto, considerando exclusivamente o resultado quantitativo do processamento da amostra *A1*, é possível definir quatro grupos, distintos, conforme ilustrado na Tabela 10:

- **G1:** Documentos que levaram à formação de mapas com aumento da quantidade de tokens;

- **G2:** Documentos que levaram à formação de mapas conceituais **COM REDUÇÃO PARCIAL** da quantidade de tokens;
- **G3:** Documentos que levaram à formação de mapas conceituais **SEM REDUÇÃO** da quantidade de tokens;
- **G4:** Documentos que não levaram à formação de mapas conceituais, ocasionado pela **REDUÇÃO TOTAL** do quantitativo de tokens.

Tabela 10. Formação de grupos de acordo com a quantidade de tokens do documento original e do mapa conceitual construído

Grupos			Documentos			Mapas Conceituais		Redução Tokens no processo (%)	
ID	Descrição	Ocorrência	Em relação à amostra (%)	Média Chunks Nominais	Média Chunks Verbais	Média. Tokens válidos	Média Tokens		Média proposições
G1	Formação de mapas conceituais COM AUMENTO de quantidade de tokens	166	16,60%	5,06	1,89	24,54	27,68	3,93	-12,82%
G2	Formação de mapas conceituais COM REDUÇÃO PARCIAL de quantidade de tokens	449	44,90%	4,34	2,27	22,56	17,70	2,88	21,52%
G3	Formação de mapas conceituais SEM ALTERAÇÃO de quantidade de tokens	262	26,20%	3,51	1,69	16,92	16,92	2,50	0,00%
G4	Não houve formação de mapas conceituais, COM REDUÇÃO TOTAL de quantidade de tokens	123	12,30%	1,44	1,24	9,46	0,00	0,00	100,00%
AMOSTRA A1		1.000	100,00%	3,86	1,92	19,67	16,85	2,58	14,35%

A redução do número de *tokens* pode ocorrer em dois momentos distintos do processo. Na atividade de construção de grafos, o descarte de *tokens* ocorre quando um *chunk* pode ser transformado em nó conceito ou nó link. Também há casos em que o nó conceito ou nó link, não encontra ponto de subsunção no grafo que intermedeia a construção de mapas conceituais, como na Figura 17. Na imagem, os *chunks* que não encontraram correspondência para a formação de um nó do grafo, acabaram por ser descartados, e, aqueles que foram convertidos a nós, precisaram ainda, passar pela prova de subsunção na estrutura.

A Tabela 11 apresenta uma relação dos *chunks* que não foram reconhecidos durante o processo de *chunking* e por isso, não puderam ser transformados em elementos de grafos.

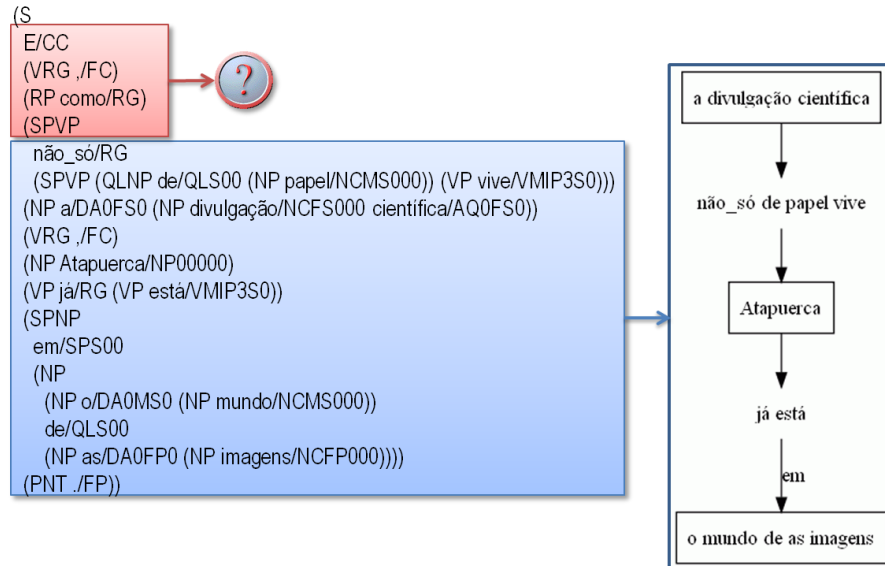


Figura 17. Redução na transformação em grafo

O segundo momento de redução pode ocorrer durante a construção do mapa conceitual, em que, não sendo possível identificar a participação de um elemento do grafo em proposições, ocorre o seu descarte. Normalmente isso ocorre quando nós link não possuem uma valência que permita a criação de proposição, em outras palavras, são nós link ligados a outro nó.

Tabela 11. *Chunks* não reconhecidos

<i>Chunks</i> NÃO Reconhecidos	COM AUMENTO de tokens (G1)	COM REDUÇÃO de tokens (G2)	SEM ALTERAÇÃO de tokens (G3)	SEM mapas conceituais (G4)	TOTAL
Adjetivo Ordinal	0	3	0	0	3
Artigo Definido	22	110	6	8	146
Artigo Indefinido	3	12	1	0	16
Aspas	23	251	21	16	311
Colchetes	0	2	0	0	2
Conjunção Coordenativa	11	76	5	6	98
Conjunção Subordinativa	3	17	1	0	21
Parênteses	22	59	4	2	87
Preposição	19	85	9	9	122
Pronome Demonstrativo	1	6	2	0	9
Pronome Indefinido	2	13	2	2	19
Pronome Interrogativo	1	6	1	0	8
Pronome Pessoal	0	18	1	1	20

Pronome Possessivo	0	1	0	0	1
Pronome Relativo	2	5	0	0	7
Qualificador	10	56	0	4	70
Sublinhado	0	1	0	0	1
TOTAL	119	721	53	48	941

O aumento da quantidade de *tokens* ocorre durante a definição das proposições, quando há mais de um nó conceito ligado a um nó link, conforme ilustrado na Figura 18.

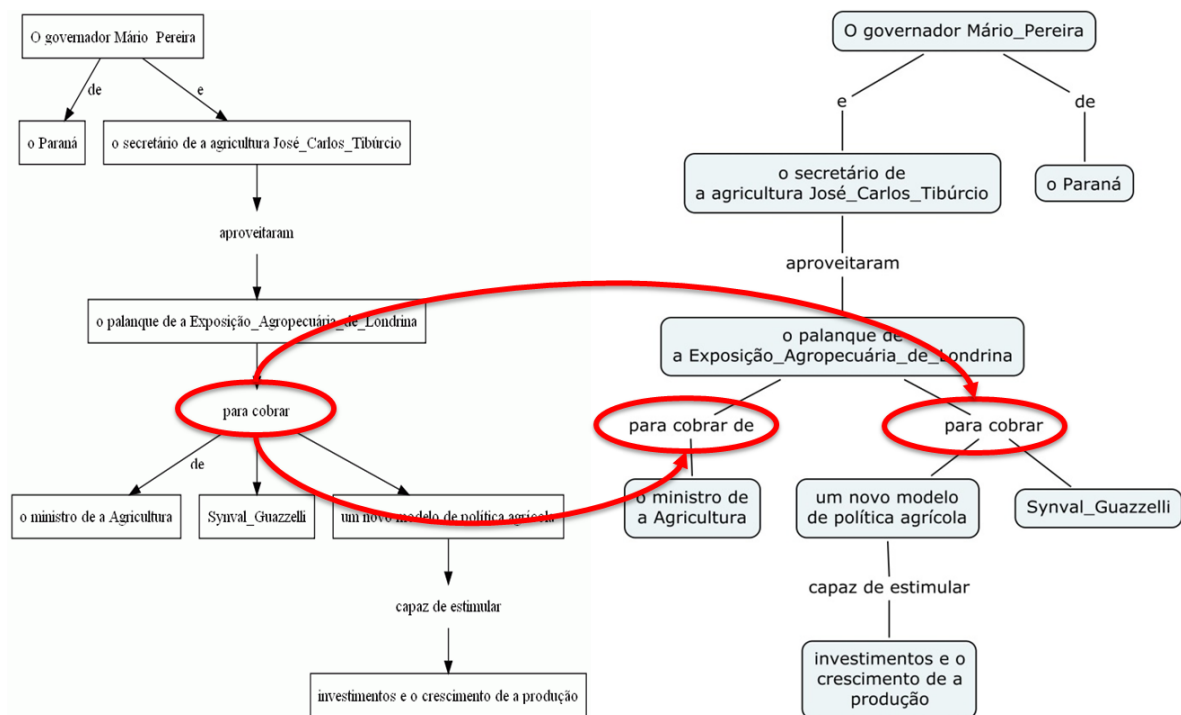


Figura 18 - Aumento do número de tokens

Nesse caso, a elaboração da proposição fará com que um nó link ligado a um nó híbrido aproprie-se da aresta do segundo para compor o rótulo da relação. No exemplo ilustrado, do conector único “cobrar” surge outro rótulo, “cobrar_de”, que denominam relações binárias distintas no mapa conceitual.

É a partir de fenômeno como esse que se torna possível aumentar a quantidade de *tokens* no processo de construção de mapa conceitual a partir de um documento original.

A Tabela 12 fornece detalhes sobre o conteúdo médio dos documentos existentes na amostra processada, exibindo, quantidade de documentos da amostra que se enquadra em cada um dos grupos identificados, a média de *tokens* válidos nesses

documentos e a representatividade de *chunks* na composição dos *chunks* verbais, nominais e outros.

Tabela 12. Detalhamento dos elementos dos grupos

Grupo		Qtde	Média Tokens Válidos	Chunks (Médias)							Geral
G1	Formação de mapas conceituais COM AUMENTO de quantidade de tokens	173	24,32	Nominais	NP	CCNP	PRNP	CSNP	QLNP	SPNP	Média
					2,73	0,05	0,05	0,01	0,34	1,87	5,06
				Verbais	VP	SPVP	CCVP	PRVP	CSVP	-	Média
	1,38	0,19	0,06	0,25	0,01		1,89				
	Outros	RP	VRG	PNT	DPT	PTV	NA	Média			
	0,07	1,82	1,00	0,04	0,00	0,72	3,64				
G2	Formação de mapas conceituais COM REDUÇÃO de quantidade de tokens	442	22,51	Nominais	NP	CCNP	PRNP	CSNP	QLNP	SPNP	Média
					2,55	0,09	0,03	0,05	0,24	1,38	4,34
				Verbais	VP	SPVP	CCVP	PRVP	CSVP	-	Média
	1,47	0,23	0,08	0,20	0,28		2,27				
	Outros	RP	VRG	PNT	DPT	PTV	NA	Média			
	0,14	1,45	1,01	0,03	0,00	1,61	4,24				
G3	Formação de mapas conceituais SEM ALTERAÇÃO de quantidade de tokens	262	16,29	Nominais	NP	CCNP	PRNP	CSNP	QLNP	SPNP	Média
					2,15	0,05	0,01	0,00	0,11	1,19	3,51
				Verbais	VP	SPVP	CCVP	PRVP	CSVP	-	Média
	1,20	0,18	0,08	0,22	0,01		1,69				
	Outros	RP	VRG	PNT	DPT	PTV	NA	Média			
	0,04	0,75	0,99	0,01	0,00	0,22	2,00				
G4	Não houve formação de mapas conceituais	123	9,33	Nominais	NP	CCNP	PRNP	CSNP	QLNP	SPNP	Média
					0,85	0,04	0,02	0,01	0,06	0,46	1,44
				Verbais	VP	SPVP	CCVP	PRVP	CSVP	-	Média
	0,94	0,08	0,02	0,05	0,15		1,24				
	Outros	RP	VRG	PNT	DPT	PTV	NA	Média			
	0,14	0,31	1,30	0,06	0,00	0,43	2,24				

Separámos os *chunks* em nominais e verbais, segundo a classe morfológica do seu núcleo. Além destes, designamos por outros, os *chunks* que são desconsiderados durante o processo.

5.4 Análise dos Resultados Obtidos

Nesta seção, analisamos a abordagem proposta, considerando os dados coletados a partir do processamento da amostra *A1* pelo protótipo *Text2Cmap*.

Ao final do processamento, é natural que nos perguntemos sobre a efetividade da abordagem na construção de mapas conceituais de qualidade. Nesta pesquisa, entendemos como qualidade os aspectos que fazem com que o mapa conceitual seja, além de legível, propício ao entendimento humano.

A avaliação da legibilidade e da compreensibilidade de um mapa conceitual sob a perspectiva humana requer igualmente uma análise estritamente humana. A questão imposta consiste na sujeição desta natureza de análise às interpretações de ordem subjetiva, pois a compreensão não se dá de forma uniforme em um grupo de pessoas: o que é de fácil entendimento para uns pode não o ser para outros. Se por um lado, tratar a subjetividade da análise é um desafio, por outro, a condução de um estudo sobre aspectos associados à compreensibilidade sem a interferência humana requer o desenvolvimento de mecanismos de interpretação semântica que extrapolam, sobremaneira, o escopo desta pesquisa.

Dado os fatos, certos de que o viés quantitativo é preferível àquele suscetível ao discernimento e opinião de cada indivíduo que realiza a análise, optamos por uma redução do conceito de compreensibilidade e legibilidade do mapa, associando-o ao conceito de “fidelidade semântica”. Sendo o documento original portador de aspectos que lhe garantam o entendimento humano, partimos da premissa de que a fidelidade do mapa conceitual construído ao conteúdo manipulado manterá tais características, se forem conservados os seus elementos originais.

Como indicador do grau de fidelidade com a fonte de dados original, definimos o nível de perda de *tokens* ao longo do processo. Embora a perda de *tokens* não seja um indicador direto de perda semântica ou de compreensibilidade, podemos, indiretamente, associar a perda de *tokens* no processo de construção à existência de lacunas semânticas que prejudicariam a compreensão do conteúdo do mapa conceitual.

Desta forma, entendemos que na base do indicador de fidelidade estão estrutura de componentes do documento original comparado aos componentes exibidos pelo mapa conceitual, excluindo-se da contagem os *tokens* referentes aos sinais de pontuação.

Observou-se que documentos com poucos *chunks* verbais e nominais dificultam a criação de proposições, como vistos no grupo G4, que apresenta uma baixa densidade desses elementos e perda total de *tokens* no processo. Neste grupo, com média de tokens reduzidos em comparação aos demais grupos, os *chunks* nominais ocorrem em média, 1,44 para 1,24 verbais. Como a relação mínima para a formação

de proposições é de dois conceitos para uma relação, é requerido ao menos dois *chunks* nominais para um verbal.

Nos grupos G1, G2 e G3 houve a formação de mapas conceituais. Notadamente, o grupo G1 foi o que mais produziu proposições em relação aos outros dois grupos: 3,89 comparados aos 2,88 do grupo G2 e aos 2,40 do grupo G3. Coincidentemente, o grupo G1 é o que apresenta a melhor relação entre *chunks* nominais e verbais: 5,01 para 1,88.

A observação da Tabela 10, em que estão enumerados os *chunks*, é um indicativo para a identificação das causas para a perda de *tokens* durante o processo de construção de mapas conceituais.

Um dos casos que ilustram as causas de perdas de forma significativa está nos artigos definidos e indefinidos não reconhecidos em um documento. A gramática utilizada por esta pesquisa para o processo de *chunking* admite a presença do artigo definido e indefinido como parte de um sintagma nominal. Não sendo possível identificá-lo dessa forma, reconhecemos as falhas mais proeminentes do processo, algumas das quais estão ilustradas na Tabela 13.

Tabela 13. Exemplos de falhas no processo que ocasionam perda de tokens

Natureza do Erro	Descrição	Documento	Saída do Processo
Erro de <i>tagger</i>	Termo “equipe” reconhecido pelo <i>tagger</i> como verbo, ao invés de substantivo	fc96ma23_0083	A/DA0FS0 equipe/VMSP1S0
Erros sintáticos	Artigo definido finalizando a sentença.	ag94ab12_0003	em/SPS00 o/DA0MS0 (PNT ./FP))
Erros de <i>chunker</i>	<i>Chunker</i> incapaz de identificar artigos que precedem sintagmas nominais	ag94ab12_0074	em/SPS00 o/DA0MS0 (NP (NP primeiro/AO0MS0 (NP trimestre/NCMS000)) de/QLS00 (NP 93/Z)) (SPNP (SPNP para/SPS00 (NP (NP (NP US/NP00000) \$/FZ) (NP 90/Z milhões/NCMP000))) (SPNP em/SPS00 (NP (NP o/DA0MS0 (NP período/NCMS000)) de/QLS00 (NP 94/Z))))

As falhas de origem no *tagger* apontam a existência de erros de associação da etiqueta morfológica decorrente de ambiguidades das palavras. A escolha do etiquetador, portanto, passa a ser um elemento central do processo quando implementado computacionalmente.

As perdas no processamento, de origem nos erros sintáticos e de *chunker*, indicam a insuficiência da abordagem no tratamento de muitos fenômenos presentes em um texto escrito em linguagem natural. Nesse sentido, direcionam a pesquisa no sentido da busca pelo aperfeiçoamento dos mecanismos, de modo que a abordagem passe a abarcar um número maior de circunstâncias sintáticas.

Mas, considerando a impossibilidade de tratar os fenômenos linguísticos à exaustão, caberia outra forma de endereçamento do problema, focado muito mais na busca por textos que apresentem estruturas adequadas para construção de mapas conceituais, do que, necessariamente, na possibilidade de tratamento de todas as estruturas possíveis na Língua Portuguesa.

6 CONCLUSÕES E TRABALHOS FUTUROS

“[...] se o trabalho for bem feito, normalmente, depois [...], verificar-se-á a irrupção de um grande frenesim de trabalho [...] E isto é sinal de que [...] vos activou o metabolismo intelectual, que foi uma experiência positiva. E ainda sinal de que são agora vítimas de uma coacção para investigar, um pouco como o Chaplin dos Tempos Modernos, que continuava a apertar parafusos mesmo depois do trabalho: e terão de fazer um esforço para parar.”

Umberto Eco

“Como se faz uma tese”

Queremos concluir, prestando contas sobre os resultados alcançados por esta pesquisa.

Iniciamos retomando as hipóteses levantadas na fase inicial de investigação, enumeradas no capítulo introdutório.

Confirmamos a hipótese que versava sobre a adequação dos métodos e técnicas de processamento de linguagem natural para o endereçamento do problema de construção de mapas conceituais a partir de textos.

Em pesquisa bibliográfica conduzida por esta pesquisa e relatada no Capítulo 3, observamos que de um total de 15 abordagens analisados, cinco das 10 abordagens que lidaram com fontes de dados não estruturadas produziram mapas conceituais ditos “completos” por meio de métodos híbridos e métodos linguísticos. Neste levantamento foi constatado que os métodos estatísticos não foram capazes de produzir mapas conceituais com conceitos e relações nomeadas.

Em nosso estudo de caso, notamos que os métodos e técnicas linguísticas combinadas aos algoritmos de busca permitiram a definição de uma solução viável para a construção de mapas conceituais a partir de textos em português do Brasil. Encontramos na abordagem linguística o esteio para a preservação de características semânticas dos conteúdos de documentos submetidos ao processo

de construção dos mapas conceituais. Mesmo em situações em que houve perda de elementos estruturais do texto, os métodos linguísticos atuaram no sentido de favorecer a identificação do prejuízo semântico e as possíveis causas das ocorrências.

Além do que citamos, some-se o fato de que o uso de recursos linguísticos está cada vez mais acessível ao pesquisador que já pode lançar mão de bibliotecas inteiras para manipulação de linguagem natural e de ferramentas específicas para análise lexical, sem a necessidade de preocupação com minúcias relacionadas com a implementação de teorias linguísticas.

Também confirmamos a hipótese sobre a possibilidade de correlacionar elementos constituintes de textos e os elementos presentes no mapa conceitual.

Apesar de não haver na literatura declaração explícita sobre a correspondência de elementos textuais e os conceitos e relações de um mapa conceitual, em nossa abordagem, consideramos a possibilidade de uso de grupos linguísticos para avaliação da hipótese. Consideramos que os elementos constituintes de um texto devem ser identificados com vistas à formação de conceitos e relações. Para isso, foi necessário quebrar o texto em sentenças e em seguida, em elementos atômicos, denominados de *tokens*. O conjunto de *tokens*, essencialmente formado por palavras, números e sinais, foi encaminhado para um processo de análise lexical, visando posterior reagrupamento na forma de grupos reduzidos de palavras, chamados de *chunks* linguísticos.

Cada um dos *chunks* reconhecidos foi considerado um candidato em potencial a se tornar um elemento de mapa conceitual. Os *chunks* de núcleo nominal foram mapeados como candidatos a conceitos e os de núcleo verbal foram considerados como candidatos a relações. Embora o mapeamento entre os candidatos potenciais e os elementos de mapa conceitual não fosse direto, sendo necessária a definição de uma heurística, o grau de proximidade entre a definição de conceito e relações aos *chunks* nominais e verbais, respectivamente, permitiu a inferência de que há sim, um nível de correlação entre os elementos constituintes de textos e os elementos de mapas conceituais.

A identificação de *chunks* nominais e verbais remete à validação da terceira hipótese, direcionada à adoção de gramática sintagmática para identificação de elementos constituintes. Com efeito, vimos ser possível estruturar os elementos do texto em forma de sintagmas de ordem nominal, verbal e proposicional que podem ser mapeados a elementos de mapas conceituais. Em nosso estudo de caso, entretanto, nos deparamos com situações em que foi necessário o uso de *chunks* que não necessariamente correspondem os tipos presentes na gramática sintagmática. Embora, grande parte do processo de construção seja realizada por meio dos sintagmas, há *chunks* que necessitam ser considerados na manutenção da semântica e que não representam, necessariamente, uma realização sintagmática. Portanto, a hipótese não foi confirmada por completo, apesar de responder por grande parcela do processo.

Por fim, tratamos a hipótese sobre o uso de grafos como estruturas de dados intermediárias entre textos e mapas conceituais. Mais uma vez, o estudo de caso possibilitou a comprovação da hipótese. A extensa literatura sobre métodos caminhamento em grafos, bem como a pesquisa consolidada na área, garantiu êxito na aplicação de grafos na intermediação entre o conteúdo do documento e o mapa conceitual. O grafo possibilita uma visão preliminar das dependências entre elementos que levam à formação de um mapa conceitual. E, pelo caminhamento pelo grafo, comprovamos a viabilidade de reconhecimento de proposições.

Convalidadas as hipóteses, faremos menção ao objetivo principal da pesquisa que se resume na **definição de uma abordagem computacional que permita a construção de mapas conceituais a partir de textos em português do Brasil.**

Sob o ponto de vista da definição da abordagem computacional, consideramos que a pesquisa foi bem sucedida, na medida em que especificou o processo e as atividades de construção, compartilhando o conhecimento que possibilitará a continuidade do trabalho por outros pesquisadores. Na ótica de viabilidade de uso da abordagem, a implementação de um protótipo computacional que realiza o processo foi determinante. Pelo protótipo computacional pudemos avaliar as forças e fraquezas da abordagem, bem como as oportunidades de melhoria. Estes aspectos foram descritos nas considerações finais do Capítulo 5.

De modo geral, encontramos respostas para as principais questões norteadoras da pesquisa, provando ser possível a construção de mapas conceituais a partir de textos em português do Brasil, assegurando indicadores objetivos e quantitativos para a apuração da perda semântica do processo, examinando os métodos e técnicas mais proeminentes nesta área de pesquisa, e definindo um mapeamento entre constituintes da língua portuguesa e os mapas conceituais.

Em relação à conformidade com os preceitos da Teoria da Aprendizagem, consideramos que a estrutura do conteúdo do texto deve comportar, de forma indireta, as premissas de organização do conhecimento, haja vista ser uma boa prática a introdução de conceitos básicos antes dos conceitos mais trabalhados.

Finalmente, assim como fizemos com as pesquisa similares, classificamos a abordagem relatada nesta dissertação, segundo as perspectivas e classes definidas no Capítulo 3. A pesquisa apresentou uma abordagem para a construção de mapas conceituais que pode ser caracterizada pela definição de objetivos educacionais, o uso de fontes de dados não estruturadas e independentes de domínio, a aplicação de recursos híbridos para a manipulação da fonte de dados e a produção de mapas conceituais de forma semiautomática, avaliadas subjetivamente e que apresentam o formato de um grafo completo, como sintetizado na Tabela 14.

Tabela 14. Síntese da Abordagem Proposta

<i>Abordagem</i>		<i>Fonte de Dados</i>			<i>Mapa Construído</i>		
<i>Referência</i>	<i>Objetivo</i>	<i>Domínio</i>	<i>Estrutura</i>	<i>Método</i>	<i>Construção</i>	<i>Leiaute Completa</i>	<i>Avaliação</i>
Kowata, Cury e Boeres (2010)	Educacional	Independente de Domínio	Não estruturada	Método Híbrido	Semi	Grafo Completo	Subjetiva

6.1 Trabalhos Futuros

Pequenas alterações na abordagem poderiam melhorar de imediato alguns dos resultados aqui apresentados. A primeira delas seria o aperfeiçoamento do método utilizado para realizar a segmentação das sentenças e dos *tokens*, definindo padrões linguísticos mais abrangentes que pudessem reconhecer cifras e locais com maior destreza.

Outro ponto de melhoria consiste na alteração da gramática para a identificação dos elementos candidatos a mapas conceituais. Observamos que vários advérbios foram descartados, bem como as conjunções e artigos, simplesmente porque não havia uma expressão regular que previsse a ocorrência destes em certas posições da oração.

Além destes, é necessário avaliar outros etiquetadores morfossintáticos. Nesta pesquisa, escolhemos o *tagger* Freeling de forma arbitrária, após tentativas com outros etiquetadores para a Língua Portuguesa. Portanto, pesquisa que comparasse a efetividade dos *taggers* para o português do Brasil, considerando o contexto da construção de mapas conceituais a partir de textos, seria uma contribuição relevante na continuidade desta pesquisa.

Trabalhos futuros devem considerar a possibilidade de aprofundamento nas questões linguísticas, sobretudo na definição de mecanismos mais robustos no tratamento de anáforas, número, gênero, sentenças interrogativas e imperativas, além de refinar heurísticas para definição dos grafos e identificação das proposições.

Outro aspecto igualmente relevante seria a seleção de textos adequados para a extração de proposições. Para isso, sugere-se a realização de experimentos que permitam a avaliação dos resultados com textos de uma ou outra natureza.

Como dissemos no Capítulo 5, será também necessário definir no protótipo um módulo para atuar na normalização dos documentos, de forma a ampliar o leque das possibilidades de uso da ferramenta.

Aliás, a ferramenta, construída de forma experimental para ser um protótipo carece de inúmeras evoluções para que se torne, de fato, uma ferramenta útil ao usuário final, leigo das questões tecnológicas.

Uma vez que a ferramenta puder ser disponibilizada para o uso em sala de aula, caberia novo estudo de caso para averiguar a relevância dos mapas conceituais produzidos, desta vez, sobre a perspectiva qualitativa.

7 REFERÊNCIAS

- ALUISIO, Sandra., PINHEIRO, Gisele Montilha., MANFRIM, Aline M.P., OLIVEIRA, Leandro H. M. de, GENOVES Jr., Luiz C., TAGNIN, Stella E.O. **The Lácio-Web: Corpora and Tools to advance Brazilian Portuguese Language Investigations and Computational Linguistic Tools.** In: Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), pp. 1779–1782, Lisbon, 2004.
- ALVES, Ana. O.; PEREIRA, Francisco Câmara.; CARDOSO, Amílcar. **Automatic Reading and Learning from Text.** In: Proceedings of the International Symposium on Artificial Intelligence (ISAI'2001), pp. 302-310. Fort Panhala (Kolhapur), India, 2001.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14724: informação e documentação – projeto de pesquisa – apresentação.** Rio de Janeiro, 2005.
- ATKINS, Sue.; CLEAR, Jeremy.; OSTLER, Nicholas. **Corpus Design Criteria.** 1999. Disponível em <<http://www.natcorp.ox.ac.uk/archive/vault/tgaw02.pdf>>. Acesso em 11 set. 2010.
- ATSERIAS, Jordi J.; Casas, Bernardino., Comelles, Elisabet., González, Meritxell., Padró, Lluís., Padró, Muntsa. (2006). FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. In: Proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006), ELRA. Genoa, Italy. Disponível em <http://www.lsi.upc.edu/~nlp/freeling>.
- BAI, Shih-Ming.; CHEN, Shyi-Ming. **Automatically constructing concept maps based on fuzzy rules for adapting learning systems.** In: Expert Systems with Applications , 35 (1), 41-49, 2008.
- BIRD, Steven.; KLEIN, Ewan.; LOPER, Edward. **Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit.** CA: O'Reilly Media, 2009.
- CHANG, Tao-Hsing.; TAM, Hak-Ping.; Lee, Chia-Hoang.; Sung, Yao-Ting. **Automatic Concept Map Constructing using top-specific training corpus.** In: Proceedings of the Asia-Pacific Educational Research Association Board Meeting (APER'A'2008). Singapore, 2008.
- CHEN, Nian-Shing.; KINSHUK; WEI, Chun-Wang.; CHEN, Hong-Jhe. **Mining e-Learning domain concept map from academic articles.** In: Computers & Education , 50 (3), 1009-1021, 2008.
- CIMIANO, Philipp. **Ontology Learning and Population from Text: Algorithms, Evaluation and Applications.** New York: Springer, 2006.
- CLARIANA, Roy. B., Koul, Ravinder. **A Computer-Based Approach for Translating Text into Concept Map-Like Representations.** In: A. J. Cañas, J. D. Novak, F. M. González (Ed.), In: Proceedings First International Conference on Concept Mapping (CMC'04), Volume 1, pp. 125-133. Pamplona, Spain, 2004.

- DALEY, Barbara J.; CONCEIÇÃO, Simone.; MINA, Liliana.; ALTAMAN, Brian A.; BALDOR, Maria.; BROWN, James. **Advancing Concept Map Research: A Review Of 2004 And 2006 Cmc Research. 2008.** In: A. J. Cañas, P. Reiska, M. Åhlberg e J. D. Novak (Org) Proceedings of the Third International Conference on Concept Mapping (CMC'08). Finland, 2008.
- DELORS, Jacques et al. **Educação: Um Tesouro a Descobrir.** Relatório para a UNESCO da Comissão Internacional sobre Educação para o século XXI. 281p. São Paulo: UNESCO, MEC, Cortez Editora, 1999. Disponível em: <<http://www.dominiopublico.gov.br/download/texto/ue000009.pdf>>. Acesso em: 30 maio 2010.
- DIESTEL, Reinhard. **Graph Theory.** Eletronic ed. New York, USA: Springer Verlag / Heidelberg, 2005
- DOWBOR, Ladislau. **Tecnologias do conhecimento: os desafios da educação.** Petrópolis: Vozes, 2001.
- DUTRA, Ítalo Modesto; FAGUNDES, Léa da Cruz; CAÑAS, Alberto J. **Uma proposta de uso dos mapas conceituais para um paradigma construtivista da formação de professores a distância.** In: X WIE - Workshop sobre Informática na Escola. Salvador, 2004.
- ECO, Umberto. **Como se faz uma Tese.** 19ª ed. Editora Perspectiva, 2007.
- GAINES, Brian. R., SHAW, Mildred. L. G. **Using Knowledge Acquisition and Representation Tools to Support Scientific Communities.** In: Proceedings of the twelfth national conference on Artificial intelligence (AAAI'94) , 1, 707-714, 1994.
- GARCIA, M., Gamallo, P. (2010). Using Morphosyntactic Post-processing to Improve POS-tagging Accuracy. In: Proceedings of the International Conference on Computational Processing of the Portuguese Language (PROPOR'2010). Porto Alegre, RS.
- GIL, Antonio Carlos. **Como elaborar projetos de pesquisa.** São Paulo: Atlas, 1991.
- GÓMEZ-PÉREZ, Asunción.; MANZANO-MACHO, David. **An overview of methods and tools for ontology learning from texts.** In: The Knowledge Engineering Review , 19 (3), 187-212. 2005.
- GRAPHVIZ 2.26.3. **Graphviz Project Website.** 2010. Disponível em: <http://www.graphviz.org/Download_windows.php>. Acesso em: 22 mar. 2010.
- GRAUDINA, Vita., GRUNDSPENKIS, Janis. **Concept Map Generation from OWL Ontologies.** In: A. J. Cañas, P. Reiska, M. Åhlberg, J. D. Novak (Ed.), In: Proceedings of the Third International Conference on Concept Mapping (CMC'08). Tallin, Estonia & Helsinki, Finland, 2008.
- GUIZZARDI, Giancarlo. **Ontological Foundations for Structural Conceptual Models.** CTIT PhD Thesis Series, No. 05-74, Telematica Instituut Fundamental Research Series, No. 015 (TI/FRS/015). Enschede, The Netherlands, 2005.
- GVGEN 0.9. **GvGen Project Website.** 2007. Disponível em: <<http://software.inl.fr/trac/wiki/GvGen>>. Acesso em: 22 mar. 2010.

- JURAFSKY, Dan.; MARTIN, James H. **Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition**. 2nd ed. New Jersey: Pearson Education, 2009.
- KOWATA, Juliana H., Cury, Davidson, Boeres, Maria Claudia. S. **Caracterização das Abordagens para Construção de Mapas Conceituais**. In: Proceedings of the XX Brazilian Symposium on Computer in Education (SBIE 2009), pp. 1-10. Florianópolis, Brazil, 2009.
- KUMAZAWA, Terukazu.; SAITO, Osamu.; KOZAKI, Kouji.; MATSUI, Takanori.; MIZOGUCHI, Riichiro. **Toward knowledge structuring of sustainability science based on ontology engineering**. In: Sustainability Science , 99-116, 2009.
- LAU, Raymond Y.K.; CHUNG, Albert Y.K.; SONG, Dawei.; HUANG, Qiang. **Towards Fuzzy Domain Ontology Based Concept Map Generation for E-Learning**. In: Advances in Web Based Learning (ICWL 2007). 4823, pp. 90-101. Springer Berlin/Heidelberg, 2008.
- LEE, Chun-Hsiung., LEE, Gwo-Guang., LEU, Yungho. **Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning**. In: Expert Systems with Applications, 36 (2), 1675-1684, 2009.
- MERCADO, Luís Paulo Leopoldo (Org). **Formação docente e novas tecnologias**. In: Novas tecnologias na educação: reflexões sobre a prática. 210 p. Maceió: EDUFAL, 2002.
- MOREIRA, Marco Antonio. **Aprendizagem Significativa: da visão clássica à visão crítica**. Apresentado na conferência de encerramento do V Encontro Internacional sobre Aprendizagem Significativa. Espanha, 2006. Disponível em: < <http://www.if.ufrgs.br/~moreira/visaoclasicavisaocritica.pdf>>. Acesso em 30 maio 2010.
- NONAKA, Ikujiro.; TAKEUCHI, Hirotaka. **Criação de conhecimento na empresa: como as empresas japonesas geram a dinâmica da inovação**. Rio de Janeiro: Campus, 1997.
- NOVAK, Joseph D. **Aprender, Criar e Utilizar o Conhecimento: Mapas Conceptuais™ como Ferramentas de Facilitação nas Escolas e Empresas**. 1. ed. Lisboa: Plátano Edições Técnicas, 1998.
- NOVAK, Joseph D., CAÑAS, Alberto J. **The Origins of the Concept Mapping Tool and the Continuing Evolution of the Tool**. 2006b. Information Visualization Journal, 5 (3), 175-184.
- NOVAK, Joseph D.; CAÑAS, Alberto. J. **The Theory Underlying Concept Maps and How to Construct and Use Them**. 2006a. Disponível em <<http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf>>. Acesso em: 6 jun. 2009.
- NOVAK, Joseph D.; GOWIN, D. Bob. **Aprender a Aprender**. 1. ed. Lisboa: Plátano Edições Técnicas, 1984.
- ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA EDUCAÇÃO, CIÊNCIA E CULTURA. **Compromisso renovado para a aprendizagem ao longo da vida: América Latina e Caribe**. In: Educação e Aprendizagem para Todos: Olhares dos Cinco Continentes, 109 p. 2009b. Disponível em:

- <<http://unesdoc.unesco.org/images/0018/001840/184075por.pdf>>. Acesso em 30 maio 2010.
- ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA EDUCAÇÃO, CIÊNCIA E CULTURA. **Padrões de Competência em TIC para professores. Marco Político.** 2009. Disponível em: <<http://unesdoc.unesco.org/images/0015/001562/156210por.pdf>>. Acesso em: 30 maio 2010.
- PEREZ, Claudia Camerini Corrêa., VIEIRA, Renata. **Aquisição de Conhecimento a partir de Textos para Construção de Mapas Conceituais.** In: II Workshop de Teses e Dissertações em Inteligência Artificial (WTDIA 2004). São Luís, 2004.
- PÉREZ, Claudia Camerini Corrêa.; VIEIRA, Renata. **Mapas Conceituais: geração e avaliação.** In: Anais do III Workshop em Tecnologia da Informação e da Linguagem Humana (TIL'2005), (pp. 2158-2167). São Leopoldo, RS, 2005.
- PYTHON 2.6.5. **Python Programming Language Official Website.** Disponível em: <<http://www.python.org/download/releases/2.6.5/>>. Acesso em: 22 mar. 2010.
- RICHARDSON, Ryan.; FOX, Edward. A. **Using Concept Maps in NDLTD as a Cross-Language.** In: 10th International Symposium on Electronic Theses and Dissertations (ETD 2007). Uppsala, Sweden, 2007.
- RICHARDSON, W. Ryan., SRINIVASAN, Venkat., Fox, Edward A. **Knowledge discovery in digital libraries of electronic theses and dissertations: an NDLTD case study.** In: International Journal on Digital Libraries , 9 (2), 163-171. 2008.
- RUIZ-PRIMO, Maria Araceli.; SHAVELSON, Richard. J.. **Problems and issues in the use of Concept Maps in science assessment.** Journal of Research in Science Teaching, 33(6), 569 – 600. 1996.
- SAFAYENI, Frank.; DERBENTSEVA, Natalia.; CAÑAS; Alberto J. **Concept Maps: A Theoretical Note on Concepts and the Need for Cyclic Concept Maps.** 2003. Disponível em <<http://cmap.ihmc.us/publications/ResearchPapers/Cyclic%20Concept%20Maps.pdf>>. Acesso em: 20 abr. 2010.
- SILVA, Edna Lúcia da., MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação.** 3. ed. rev. atual. Florianópolis: Laboratório de Ensino a Distância da UFSC, 121 p. 2001. Disponível em: <<http://projetos.inf.ufsc.br/arquivos/Metodologia%20da%20Pesquisa%203a%20edicao.pdf>>. Acesso em 10 out. 2009.
- SILVA, Ezequiel Theodoro da. **Revalorização do livro diante das novas mídias. Veículos e linguagens do mundo contemporâneo: a educação do leitor para as encruzilhadas da mídia.** In: Integração das Tecnologias na Educação. Salto para o Futuro. Almeida, Maria Elizabeth Bianconcini de., Moran, José Manuel. (Org). 204 p. 2005. Disponível em <<http://portal.mec.gov.br/seed/arquivos/pdf/1sf.pdf>>. Acesso em 30 maio 2010.
- TSENG, Shian-Shyong.; SUE, Pei-Chi.; SU, Jun-Ming.; WENG, Jui-Feng.; TSAI, Wen-Nung. **A new approach for constructing the concept map.** In: Computers & Education, 49 (3), 691-707, 2007.

- VALENTE, José Armando. **Pesquisa, comunicação e aprendizagem com o computador. O papel do computador no processo ensino-aprendizagem.** In: Integração das Tecnologias na Educação. Salto para o Futuro. Almeida, Maria Elizabeth Bianconcini de., Moran, José Manuel. (Org). 204 p. 2005. Disponível em <<http://portal.mec.gov.br/seed/arquivos/pdf/1sf.pdf>>. Acesso em 30 maio 2010.
- VALERIO, Alejandro.; LEAKE, David. **Jump-Starting Concept Map Construction with Knowledge Extracted from Documents.** In: A. J. Cañas, J. D. Novak, F. M. González (Ed.), In: Proceedings Second International Conference on Concept Mapping (CMC'06), 1, pp. 296-303. San José, Costa Rica, 2006.
- VILLALON, Jorge J., CALVO, Rafael A. **Concept Map Mining: A definition and a framework for its evaluation.** In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology , 3 (9-12), 357-360, 2008.
- ZOUAQ, Amal.; NKAMBOU, Roger. **Building Domain Ontologies from Text for Educational Purposes.** In: IEEE Transactions on Learning Technologies , Volume 1 (1), p. 49-62, 2008.
- ZOUAQ, Amal.; NKAMBOU, Roger. **Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project.** In: IEEE Transactions on Knowledge and Data Engineering (10.1109/TKDE.2009.25), 2009.

APÊNDICES

APÊNDICE A – DETALHES DA IMPLEMENTAÇÃO DO PROTÓTIPO

Tabela 15. Ambiente de Desenvolvimento

Recurso	Descrição
Windows Vista	Sistema Operacional
Eclipse Java EE IDE for Web Developers (Build id: 20090621-0832)	Ambiente de Desenvolvimento de Software
PyDev for Eclipse (1.6.1.2010080312)	Plugin para Desenvolvimento Python em Eclipse
Python 2.6.2	Biblioteca de recursos Python
Graphviz 2.26.3	Biblioteca de recursos para grafos
NLTK 2.6	Biblioteca de recursos para Processamento de Linguagem Natural
Freeling 2.0	Aplicativo para etiquetagem morfosintática
Mac-Morpho	Corpus Linguístico

Tabela 16. Relação de Padrões para reconhecimento de *tokens*

Padrões Linguísticos	Expressão Regular
Abreviações. Ex.: “S.A.”	<code>([A-Z]+\.)+</code>
Reticências	<code>\.\.\.</code>
Palavras compostas. Ex.: “beija-flor”	<code>\w+(\w+)*(-\w+)*</code>
Número, moeda, percentual	<code>[A-Z]{0,}\\$?\s?\d+([,.\]\d+)?\s?%?</code>
Sinais	<code>[] [.,; "'?() :- _ `]</code>

Tabela 17. Trecho do código para reconhecimento de *chunks*

Chunk	Expressão Regular ¹⁴
NP	<pre> NP: {<N.* AQ.* Z.* PP...N.* PX.* PIO.* PDO.*>+} NP: {<DA.* DI.* DD.*><VMP.*>} NP: {<DA.* DI.* DD.*><QL.*>?<NP>} NP: {<DA.* DI.* DD.*><PRO.*>} NP: {<NP><FZ>} NP: {<AO.*><NP>} NP: {<NP><AO.*>} NP: {<NP>(<QL.*><NP>)+} NP: {<NP><VMP00.*>} NP: {<NP><CC><NP>} NP: {<NP><RG.* RN.*><NP>} NP: {<FPA><NP><FPT>} NP: {<NP><NP>+} </pre>

¹⁴ O TAGSET completo está disponível em: <http://garraf.epsevg.upc.es/freeling/doc/userman/parole-es.html>

Chunk	Expressão Regular¹⁴
	NP: {<RG.* RN.*><NP>}
CCNP	CCNP: {<CC><NP>}
PRNP	PRNP: {<PRO.*><NP>}
CSNP	CSNP: {<CS.*><NP>}
QLNP	QLNP: {<QL.*><NP> QLNP: {<QLNP><QLNP>+}
SPNP	SPNP: {<SPS.*><NP>} SPNP: {<SPS.*><RG.* RN.*>} SPNP: {<SPNP><CC><SPNP>} SPNP: {<SPNP><SPNP>+} SPNP: {<RG.* RN.*><SPNP>}
VP	VP: {<V.*>+} VP: {<NP><QL.*><V.*>} VP: {<PP.....0.*><V.*>} VP: {<V.*><PP.....0.*>} VP: {<RG.* RN.*>+<VP>} VP: {<VP><RG.* RN.*>+} VP: {<VP>(<CC FC><VP>)+} VP: {<VP><VP>+}
SPVP	SPVP: {<SPS.*><VP>} SPVP: {<SPVP.* QL.*>+<VP>} SPVP: {<SPVP.*><DA.*><VP>} SPVP: {<RG.* RN.*><SPVP.*>}
CCVP	CCVP: {<CC><VP>}
RP	RP: {<RG.* RN.*>+} RP: {<FC><RP><FC>}
PRVP	PRVP: {<PRO.*><VP>}
CSVP	CSVP: {<CS.*><VP>}
VRG	VRG: {<FC>}
PNT	PNT: {<FP FIT FAT>}
DPT	DPT: {<FD>}
PTV	PTV: {<FX>}

Tabela 18. Regras de Compatibilidade entre Nós Candidatos

Filho	Condição	Ação
NP	Pai = VP	Anexar nó filho
	Pai = PRVP	Anexar nó filho
	Pai = SPVP	Anexar nó filho
	Pai = SPNP e distância de até um nó	Compor nó pai
VP	Pai = NP e Avô <> PRVP	Anexar nó filho
	Pai = VP e distância zero	Compor nó pai
	Pai = CCNP e distância zero	Anexar nó filho
SPNP	Pai = VP	Anexar nó filho
	Pai = PRVP	Anexar nó filho
	Pai = SPVP	Anexar nó filho
	Pai = NP e distância de até um nó	Anexar nó filho
	Pai = RP e distância de até um nó	Compor nó pai
QLNP	Pai = NP e distância de até um nó	Anexar nó filho

Filho	Condição	Ação
	Pai = VP e distância de até um nó	Anexar nó filho
	Pai = PRVP	Anexar nó filho
	Pai = SPVP	Anexar nó filho
	Pai = SPNP	Anexar nó filho
	Pai = QLPNP e distância de até um nó	Compor nó pai
PRVP	Pai = SPVP	Mudar para VP e não anexar
	Pai <> VRG e Pai <> SPVP	Anexar nó filho
PRNP	Pai = VP	Anexar nó filho, mudando seu tipo para NP
	Pai = SPVP	Anexar nó filho, mudando seu tipo para NP
CCVP	Pai = NP	Anexar nó filho, mudando seu tipo para VP
SPVP	Pai = VP	Não anexar, mudar o tipo para VP
	Pai = PRVP	Não anexar, mudar o tipo para VP
	Pai = NP	Anexar nó filho
	Pai = RP e distância de até um nó	Não anexar
CCNP	Pai = NP	Anexar nó
RP	Pai = VP e distância de até um nó	Não anexar o nó

Tabela 19. Busca de Elemento Subsunçor a partir das folhas

```

def searchlastnode(anchorset, currentnode):
    # procurando nó de início da busca
    lastnode = None

    # identifica o melhor valor encontrado até o momento
    bestprobability = 0

    # o anchorset é uma lista em ordem histórica
    # quanto maior o índice, mais recente o nó foi adicionado
    for i in range(0, len(anchorset)):

        if i == 0:
            lastnode = anchorset[i]
        else:
            weight = 0
            probability = 0
            lastnodecandidate = anchorset[i]
            if lastnodecandidate.type == 'CCNP':
                weight = 0
            elif lastnodecandidate.type == 'VP':
                weight = 50
            elif lastnodecandidate.type == 'SPNP':
                weight = 50
            elif lastnodecandidate.type == 'QLNP':
                weight = 50
            probability = weight + ( i * 10 )
            if bestprobability <= probability:
                bestprobability = probability
                lastnode = anchorset[i]

    return lastnode

```