



**UNIVERSIDADE FEDERAL DO PARÁ**  
**CENTRO DE CIÊNCIAS EXATAS E NATURAIS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

Clay Palmeira da Silva

**Text2MARK: Uma ferramenta de mineração de  
textos para auxílio na representação do  
conhecimento**

Belém

2013

Clay Palmeira da Silva

# Text2MARK: Uma ferramenta de mineração de textos para auxílio na representação do conhecimento

Dissertação de Mestrado apresentada para a obtenção do grau de Mestre em Ciência da Computação no Programa de Pós Graduação em Ciência da Computação do Instituto de Ciências Exatas e Naturais da Universidade Federal do Pará.

Área de Concentração: Sistemas de Computação  
Sub-Área: Sistemas Inteligentes.

Orientador Prof. Dr. Jefferson Magalhães de Moraes

Belém

2013

Silva, Clay Palmeira da

Text2MARK: Uma Ferramenta de mineração de textos para auxílio na representação do conhecimento / (Clay Palmeira da Silva); orientador, Jefferson Magalhães de Morais.. - 2013.

128. il. 28cm.

Dissertação (Mestrado) - Universidade Federal do Pará. Instituto de Ciências Exatas e Naturais. Programa de Pós-Graduação em Ciência da Computação. Belém, 2013.

1. Inteligência Artificial. 2 Mineração de textos. I. Morais, J.M de, orient. II. Universidade Federal do Pará, Instituto de Ciências Exatas e Naturais, Programa de Pós-Graduação em Ciência da Computação. III. Título.

**CDD 22. ed.515.35**

Clay Palmeira da Silva

# Text2MARK: Uma ferramenta de mineração de textos para auxílio na representação do conhecimento

Dissertação de Mestrado apresentada para a obtenção do grau de Mestre em Ciência da Computação do Programa de Pós Graduação em Ciência da Computação do Instituto de Ciências Exatas e Naturais da Universidade Federal do Pará.

Data da aprovação: Belém-Pa. 04 – 10 - 2013

Banca Examinadora

Prof. Dr. Jefferson Magalhães de Moraes

Programa de Pós Graduação em Ciência da Computação - UFPA - Orientador

Prof. Dr. Dionne Cavalcante Monteiro

Programa de Pós Graduação em Ciência da Computação – UFPA – Membro Externo

Prof. Dra. Miriam Lúcia Campos Serra Domingues

Programa de Pós-Graduação em Engenharia Elétrica – UFPA – Membro Externo

Prof. Dra. Marianne Kogut Eliasquevici

Faculdade de Computação - FACOMP– UFPA – Membro Externo

## AGRADECIMENTOS

A Deus, pois toda e qualquer existência, independente de crença evolutiva, requer a presença de um ser supremo, um Grande Arquiteto do Universo. Que nos auxilia durante todo e qualquer obstáculo.

Ao Prof. Jefferson Moraes que aceitou o desafio de concluir essa etapa importante da minha formação mesmo com todos os imprevistos.

À minha amada Família a qual tive que viver afastado por algum tempo, mas que soube compreender minha ausência. É a ela que credito a razão pela qual faço na vida os sacrifícios necessários.

Amigos de PPGCC constituídos ao longo desse tempo juntos. Todos com grande contribuição na vida pessoal, acadêmica e profissional.

Aos professores Eduardo Cerqueira e Raimundo Viégas que em momentos distintos souberam com seus conselhos e ensinamentos me dar o norte necessário nesta trajetória. Aos Professores Eloi Favero e Dionne Monteiro pela compreensão em momento decisivo.

À Professora Miriam Domingues, pelo seu prestativo auxílio e ajuda em momento decisivo deste trabalho sem o qual não o teríamos concluído.

Ao corpo administrativo do PPGCC, em específico as atividades da Secretaria acadêmica na pessoa da querida Telma Mafra, que sempre soube auxiliar e amparar nos momentos necessários. Agradeço também aos acadêmicos que de alguma forma colaboraram para que o curso pudesse ser realizado dentro das instalações da Universidade.

À CAPES pelo apoio financeiro.

“Ainda que eu andasse pelo vale da sombra e da morte, não temeria mal algum, porque tu estás comigo; a tua vara e o teu cajado me consolam”

Salmos 23:4

“E conhecereis a verdade, e a verdade vos libertará”

João 8:32

## RESUMO

Representar conhecimento no início do ano 2000 tornou-se um desafio devido ao grande volume de informações aos quais somos submetidos. Associar técnicas inteligentes tornou-se um método relevante, porém, ainda faltam estudos para tornar o processo mais simples para o usuário. Considerando o atual estado da arte dos etiquetadores de texto da área de processamento de linguagem natural (PLN), os quais apresentam alta acurácia na etiquetagem de sequências de palavras, o objetivo deste trabalho, chamado Text2MARK, é desenvolver uma ferramenta computacional que possa minerar de forma eficiente e sem complexidade textos livres em português, além de abranger a construção de regras gramaticais que foram denominadas predicados. Para isso foram utilizadas técnicas híbridas de mineração de textos associadas a predicados léxicos construídos com base na estrutura da gramática portuguesa. Além disso, foi desenvolvido um método automático de extração de palavras-chave em forma de tuplas para construção de mapas conceituais. As tuplas obtidas serviram como entrada na ferramenta CMapTools para representar conhecimento através de mapas conceituais. A proposta obteve como média 84,2% de tuplas que correspondiam à representação do texto analisado. Com base em trechos de textos extraídos de cinco artigos científicos distintos foram gerados cinco mapas conceituais. Os mapas foram submetidos à avaliação subjetiva de três grupos diferentes de usuários que atribuíram notas de 0 a 10 e obteve como média as notas 8,1, 8,1 e 8,3 respectivamente em termos de similaridade com o trecho do texto analisado. Os resultados comprovam que é possível utilizar tuplas em formatos NOME-VERBO-NOME para representar conhecimento através de sentenças gramaticais para construir mapas conceituais, ainda demonstra na avaliação subjetiva que textos elaborados com poucas contrações ou abreviações são mais fáceis de representar em forma de mapa.

**PALAVRAS-CHAVE:** Mineração de textos, processamento de linguagem natural, mapas conceituais, conhecimento.

## ABSTRACT

Representing knowledge in the beginning of year 2000 has become a challenge due to the large volume of information to which we are submitted. Associate intelligent techniques has become a relevant method, however, studies are still lacking to make the process simpler for the user. Considering the current state of the art part-of-speech taggers text area of natural language processing (NLP), which present high accuracy in the labeling of sequences of words, the aim of work, called Text2MARK, is to develop a computational tool that can mine the efficiently and without complexity free texts in Portuguese, besides covering the construction of grammatical rules that were called predicates. For this hybrid text mining techniques associated with lexical predicates constructed based on the structure of Portuguese grammar were used. Furthermore, we developed an automated method of extracting keywords in the form of tuples to construct concept maps. Tuples obtained served as input CmapTools tool to represent knowledge through concept mapping. The proposal achieved as average 84.2% of tuples that correspond to the representation of the analyzed text. Based on excerpts from texts taken from five separate paper five conceptual maps were generated. The maps underwent subjective evaluation of three different groups of users assign grades 0-10 and obtained as average grades 8.1, 8.1 and 8.3 respectively in terms of similarity to the words of the text analyzed. The results show that it is possible to use tuples NOUN-VERB-NOUN formats to represent knowledge through grammatical sentences to construct concept maps still shows the subjective evaluation that texts produced with few contractions or abbreviations are easier to represent in map form .

**KEYWORDS:** Text mining, natural language process, concept maps, knowledge.

## LISTA DE ILUSTRAÇÕES

Figura 1. O processo de mineração de textos. Stravrianou et al (2007).....	23
Figura 2. Representação da construção de constituintes. Adaptado de Perini (1995, pg44)....	25
Figura 3. Representação da hierarquia dos constituintes identificando elementos da frase. Adaptado de Perini (1995, p-69) .....	28
Figura 4. Sistemas de memória chave do cérebro onde todos os elementos interagem quando estamos aprendendo.....	30
Figura 5. Representação em fluxograma. Adaptado Tavares (2007) .....	33
Figura 6. Mapa conceitual modelo Teia de Aranha. Adaptado Silva et al (2009) .....	33
Figura 7. Mapa conceitual modelo Entrada e Saída .....	34
Figura 8. Mapa hierárquico apresenta de forma simples a compreensão de conceitos. Adaptado de Moreira et al (1982).....	35
Figura 9. Representação hierárquica da associação de conceitos.....	36
Figura 10. Um mapa conceitual para o conteúdo de Eletromagnetismo .....	40
Figura 11. Mapeamento do conteúdo de Eletromagnetismo. Adaptado Moreira (2006).....	40
Figura 12. O site completamente desenvolvido como mapa conceitual. Uma nova abordagem no processo de construção e representação cognitiva. ....	42
Figura 13. Na interface do Text2MARK há uma área para receber o texto de entrada .....	51
Figura 14. Inicialmente chamado de etiquetador, gerava apenas os formatos NVN e VNV ...	54
Figura 15. O diagrama com os requisitos das relações. Adaptado Heuser (2001).....	57
Figura 16. O mapa conceitual gerado no CMapTools com as saídas do Text2MARK. ....	57
Figura 17. (a) Já com o nome de Text2MARK. ....	58
Figura 18. Arquitetura funcional do Text2MARK.....	62
Figura 19. Metodologia de extração do Text2MARK.....	64
Figura 20. (a) Extraindo a forma tradicional NVN e em (b) extraindo o determinante ART ..	66
Figura 21. O diagrama com os requisitos das relações. Adaptado Heuser (2001).....	67
Figura 22. O mapa mais completo gerado no CMapTools com as saídas do Text2MARK. ...	67
Figura 23. Abaixo à esquerda a seleção inicial das tuplas encontradas automaticamente .....	72

Figura 24. O mapa conceitual do texto com base nas tuplas obtidas no Text2MARK.....	74
Figura 25. O Text2MARK permite selecionar uma ou várias tuplas e adicionar ao final .....	77
Figura 26. Círculos em laranja orientam o usuário sobre a ordem de leitura do MC.....	79
Figura 27. Um dos textos mais interessantes. Muita complexidade linguística.....	84
Figura 28. Mapa com grandes ramificações devido sua estrutura textual.....	89
Figura 29. Último teste do Text2MARK com 100% das tuplas válidas .....	94

## LISTA DE TABELAS

Tabela 1. Os principais desafios da mineração de textos .....	22
Tabela 2. Os autores e suas contribuições para o processo de extração de textos.....	52
Tabela 3. As descrições das abreviaturas apresentadas na Tabela 2 .....	53
Tabela 4. Os requisitos e as saídas geradas automaticamente .....	56
Tabela 5. Léxicos e as siglas atribuídas pelo etiquetador. Fonte: TreeTagger .....	60
Tabela 6. Regras de predicados com base nos SN, SV e SP. ....	63
Tabela 7. Os requisitos identificados no texto com as regras e saídas .....	65
Tabela 8. Frequência das categorias de palavras mais relevantes no texto de Geografia .....	70
Tabela 9. Tuplas selecionadas automaticamente no texto de Geografia .....	71
Tabela 10. 50% dos predicados não foram encontrados no texto analisado .....	72
Tabela 11. Tuplas marcadas com COMPLEMENTO dão maior detalhe ao mapa.....	72
Tabela 12. Frequência dos léxicos mais relevantes no texto de EAD.....	75
Tabela 13. Tuplas selecionadas automaticamente no texto de EAD.....	76
Tabela 14. 54% dos predicados não foram encontrados no texto analisado .....	77
Tabela 15. Tuplas complemento são requeridas para dar mais detalhe ao MC.....	77
Tabela 16. Frequência dos léxicos mais relevantes no texto de Linguística .....	80
Tabela 17. Tuplas selecionadas automaticamente no texto de Linguística .....	80
Tabela 18. 50% dos predicados não foram encontrados no texto analisado .....	82
Tabela 19. Menor quantidade de tuplas adicionais e maior utilização das mesmas.....	82
Tabela 20. Frequência dos léxicos mais relevantes no texto de Meio Ambiente .....	85
Tabela 21. Tuplas selecionadas automaticamente no texto de Meio Ambiente.....	86
Tabela 22. 60% dos predicados não foram encontrados no texto analisado .....	86
Tabela 23. Todas tuplas adicionais necessárias na construção do MC .....	87
Tabela 24. Frequência dos léxicos mais relevantes no texto de Pedagogia .....	90
Tabela 25. Tuplas selecionadas automaticamente no texto de Pedagogia .....	91
Tabela 26. 55% dos predicados não foram encontrados no texto analisado .....	92

Tabela 27. Tuplas adicionais quando a média da soma deixa de fora termos importantes .....	92
Tabela 28. A avaliação subjetiva feita pelos alunos de letras da UNIFAP .....	96
Tabela 29. A avaliação subjetiva feita na UFPa .....	96
Tabela 30. Distribuição mais uniforme e consistente entre os alunos do sétimo semestre .....	97
Tabela 31. Dados comparativos consolidados .....	98
Tabela 32. O percentual de tuplas válidas encontradas automaticamente pelo Text2MARK	100

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>14</b>
1.1	Contexto do trabalho .....	14
1.2	Motivação .....	15
1.3	Objetivos .....	15
1.3.1	Geral.....	16
1.3.2	Específicos .....	16
1.4	Metodologia do trabalho .....	16
1.5	Contribuição da dissertação .....	19
1.6	Estrutura da dissertação .....	19
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>21</b>
2.1	Características e definições sobre mineração de textos.....	21
2.2	Uma revisão de Gramática descritiva .....	23
2.3	Mapas Conceituais .....	29
2.3.1	Principais Conceitos .....	29
2.3.2	As representações .....	31
2.3.3	Uso prático de Mapas Conceituais.....	36
2.3.4	Ferramenta escolhida .....	41
2.3.5	Considerações Finais do Capítulo.....	42
<b>3</b>	<b>O ESTADO DA ARTE E DEFINIÇÃO DA PROPOSTA</b> .....	<b>44</b>
3.1	Os trabalhos relacionados com extração automatizada de textos para construção de mapas conceituais .....	46
3.2	As tentativas desenvolvidas.....	48
3.3	Os problemas não resolvidos .....	49
3.4	A simplificação do modelo de extração e mineração .....	51
3.5	Breve introdução ao Text2MARK .....	51
3.6	Considerações finais do Capítulo .....	52
<b>4</b>	<b>A FERRAMENTA DE EXTRAÇÃO DE TEXTOS TEXT2MARK</b> .....	<b>54</b>
4.1	Objetivo da ferramenta .....	55
4.2	Definição do escopo.....	55
4.3	Projeto técnico.....	58
4.4	Arquitetura.....	61
4.5	As regras gramaticais .....	63
4.6	Metodologia de extração .....	64
4.7	A escolha do escopo.....	65
4.8	Considerações finais do Capítulo .....	68
<b>5</b>	<b>AValiação Qualitativa</b> .....	<b>69</b>

<b>5.1 A extração dos textos técnicos.....</b>	<b>69</b>
5.1.1 Geografia .....	70
5.1.2 EAD .....	75
5.1.3 Linguística .....	79
5.1.4 Meio Ambiente .....	85
5.1.5 Pedagogia.....	90
<b>5.2 Os grupos de avaliação dos mapas conceituais .....</b>	<b>95</b>
<b>5.3 Considerações finais do Capítulo .....</b>	<b>99</b>
<b>6 CONCLUSÕES.....</b>	<b>101</b>
6.1 Trabalhos futuros .....	102
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>103</b>
<b>ANEXO A – TRECHO CORRESPONDENTE A SEÇÃO 6.1.1.....</b>	<b>112</b>
<b>ANEXO B – TRECHO CORRESPONDENTE A SEÇÃO 6.1.2.....</b>	<b>117</b>
<b>ANEXO C – TRECHO CORRESPONDENTE A SEÇÃO 6.1.3.....</b>	<b>120</b>
<b>ANEXO D – TRECHO CORRESPONDENTE A SEÇÃO 6.1.4.....</b>	<b>123</b>
<b>ANEXO D – TRECHO CORRESPONDENTE A SEÇÃO 6.1.5.....</b>	<b>126</b>

# 1 INTRODUÇÃO

Inicialmente, serão apresentadas as circunstâncias e dificuldades que nortearam a realização desta pesquisa. Faz-se necessário contextualizar o cenário e características das informações necessárias para compor o conhecimento desde o início do século XXI até os dias de hoje. Após isso, serão apresentados os problemas e as perguntas que nos motivaram a realizar este trabalho. Para auxiliar essa condução, apresentamos os objetivos gerais e específicos e qual metodologia foi utilizada para identificar os problemas e responder as perguntas. Em seguida, será apresentada a contribuição deste trabalho e por fim a estrutura desta dissertação organizada em seus capítulos.

## 1.1 Contexto do trabalho

A partir do ano 2000, escolas de ensino fundamental, médio e superior estão fazendo uso de novas metodologias de ensino para facilitar e dinamizar o processo ensino-aprendizagem. Tais métodos já fazem parte hoje não somente dos círculos pedagógicos, como também das atividades de alguns atores externos como os docentes e discentes de diversas áreas.

A nossa sociedade pode ser definida como uma sociedade de informação, Kowata et al (2012). Com isso, faz-se necessário a inclusão de novos métodos de ensino para acompanhar a evolução de assimilação que a atual geração requer. Porém, há certo impasse de como podemos associar os atuais recursos computacionais, multimídias, instrucionais, além das grandes bases de dados de artigos científicos disponíveis.

Hoje, somos submetidos a um grande volume de informações de naturezas distintas. Tentar representar o conhecimento de forma parcial ou total tem se mostrado um grande desafio. É possível citar algumas técnicas educacionais e inteligentes, como por exemplo, os Mapas Conceituais (MCs) e a Mineração de Textos, como métodos eficientes e dinâmicos para a representação do conhecimento.

O processo de construção do conhecimento por intermédio de MCs na maneira proposta por Novak & Gowin (1984) considera uma estruturação hierárquica dos conceitos que serão apresentados tanto por meio de uma diferenciação progressiva quanto de uma reconciliação integrativa. Tais mapas se estruturam de acordo com a Teoria da Aprendizagem Significativa de David Ausubel (Ausubel, 2003), contribuindo para a construção do conhecimento.

## **1.2 Motivação**

A construção de mapas conceituais na maneira proposta por Novak & Gowin (1984), mesmo considerando uma estruturação hierárquica, diferenciada e progressiva dos conceitos pertinentes ao texto, que irá servir como base para construir o mapa conceitual, não é uma tarefa clara, tampouco simples, justamente, por existirem vários fatores gramaticais que são necessários para compor os conceitos requeridos para a compreensão do texto no mapa conceitual.

Ao identificar o problema, surgiu a pergunta que nos motivou a desenvolver esta pesquisa. Como conseguir identificar quais são as informações necessárias e relevantes em um texto para que possam fazer parte da construção de um mapa conceitual? Tal pergunta se apresenta como um verdadeiro desafio, uma vez que o mapa gerado depende diretamente do grau de domínio que o usuário tiver sobre o assunto em questão. Portanto, é possível obter vários mapas conceituais de um mesmo texto que representem a mesma coisa, com enfoques diferentes, de acordo com grau de conhecimento do assunto de quem o construiu.

Pelo seu centenário e por razões pessoais citamos Alan Turing e sua máquina de estado finito (Hodges, 1997). Desde então, os estudos da Inteligência Artificial vêm se ampliando a cada ano e suas técnicas estão sendo cada vez mais utilizadas em diversas áreas, principalmente na aquisição do conhecimento. Essa característica permite que as pesquisas realizadas nas áreas de mineração de textos e representação do conhecimento possam ser aplicadas por outras áreas da computação para corroborar seus experimentos.

## **1.3 Objetivos**

Compreende-se que para a obtenção dos objetivos deste trabalho, serão necessárias técnicas automatizadas de obtenção de marcação de categorias lexicais e que introduzem ou operam com símbolos complexos permitindo a construção de traços sintáticos especificados

na extração dos conceitos requeridos para a construção dos mapas conceituais. Tais requisitos serão desenvolvidos em Java e embarcados em uma interface chamada Text2MARK.

### **1.3.1 Geral**

Identificar, através do uso de busca automatizada associada com técnicas de mineração de textos, quais são os elementos gramaticais relevantes em um texto de natureza científica ou não. Os elementos extraídos automaticamente devem permitir a obtenção de tuplas gramaticais concisas o suficiente para a formação de sentenças que viabilizem a construção de um mapa conceitual.

### **1.3.2 Específicos**

- i. Realizar processos automatizados de etiquetagem de palavras utilizando o etiquetador TreeTagger (Schmid,1995) associado com as melhorias apresentados em Domingues (2011);
- ii. Extrair os elementos gramaticais necessários para representar em 100% um diagrama entidade-relacionamento de Heuser;
- iii. Extrair a raiz etimológica das palavras do texto e obter o cálculo de sua frequência;
- iv. Identificar quais são os elementos gramaticais mais relevantes em um texto;
- v. Criar regras de predicados que possam representar sentenças em textos;
- vi. Extrair a soma dessas tuplas com base no cálculo da frequência das suas palavras;
- vii. Obter sentenças válidas de categorias lexicais no formato de tuplas para a construção dos mapas conceituais.

## **1.4 Metodologia do trabalho**

Inicialmente, o Text2MARK será aplicado no livro Projeto de Banco de Dados de Carlos Alberto Heuser (2001) o qual apresenta diagramas entidade-relacionamentos que é um tipo de mapa conceitual do texto apresentado. Neste momento, o Text2MARK deve ser capaz de

extrair os elementos gramaticais necessários para construção de um mapa conceitual. Neste momento busca-se representar um texto do livro com acerto de 100% da extração dos elementos relevantes.

Em um segundo momento será utilizado artigos científicos das áreas de: Geografia, Educação a Distância (EAD), Linguística, Meio ambiente e Pedagogia. Destes, serão selecionados, aleatoriamente, seções ou capítulos dos quais serão extraídas as tuplas para a construção de mapas conceituais.

Para o resultado final ser válido é importante compreender características de construção de sentenças da Língua Portuguesa, como por exemplo, contrações e posições dos elementos gramaticais. Além disso, é necessário compreender as complexidades que envolvem o processamento de linguagem natural para que após a realização dos processos de mineração o texto extraído não seja comprometido com ruídos e distorções.

A primeira parte do processo de entrada trata da normalização e segmentação do texto ou tokenização, onde são feitos alguns pré-processamentos para separação das preposições, como por exemplo: i)da em “de a”; ii)das em “de as”; iii) do em “de o”; iv)dos em “de os”. Nas formas maiúsculas e minúsculas, entre outras. Da mesma forma, acentuações na forma de crase e pontuações também são tratadas, assim como os pronomes, espaços, etc. Feito isso, o texto é quebrado em linhas para cada elemento identificado.

O processo seguinte trata da etiquetagem executada com o TreeTagger, esse etiquetador requer um arquivo com parâmetros do português “portugues.par” o qual contém um léxico de palavras extraído de 28 milhões de palavras do Corpus CETENFolha (Linguatca, 2013), necessário para poder etiquetar o texto previamente preparado no formato de uma palavra por linha.

As etapas seguintes são a análise e o refinamento lexical. A análise é feita a partir de comparações e cálculos da acurácia que compara os pares de palavras-etiquetas obtidos com os pares previamente etiquetados correspondentes e identifica os pares iguais (acertos) e os pares diferentes (erros). A acurácia é calculada pela soma de todos os acertos dividida pelo total de pares palavras-etiquetas submetido à etiquetagem. Já o refinamento trata da melhoria obtida no processo de etiquetagem descrito no trabalho de Domingues (2011), que ajusta suas inconsistências de etiquetas aumentando o cálculo de sua acurácia final em valores acima de 98% em textos jornalísticos e científicos.

A questão da importância dos elementos léxicos no texto é tratada através do reconhecimento de elementos centrais, os quais foram obtidos através do cálculo de suas frequências no texto. Com isso, ao identificar os Sintagmas Nominais (SN), Sintagmas Verbais (SV) e os Sintagmas Preposicionais (SP) e da forma como estavam dispostos, partiu-se para entender como esses elementos se relacionavam.

Para fazer a interpretação de dependências dos elementos léxicos SN, SV e SP, várias regras gramaticais foram criadas. Durante os estudos, várias propostas, por exemplo, Dalmolin et al (2009a), Dalmolin et al (2009b), Stravrianou et al (2002), Kowata et al (2009), Kowata et al (2012), Pérez et al (2004) e Pérez et al (2005) apresentaram que a relação NOME-VERBO-NOME se traduz de forma clara e objetiva na construção de conceitos, entretanto, como são vários os elementos que podem compor o NOME (SN) e o VERBO (SV), a escolha de tais elementos não se torna clara para quem vai construir um mapa conceitual.

Também através da frequência, foram identificados dezessete elementos léxicos mais relevantes, que quando associados da forma adequada compõe a tupla NOME-VERBO-NOME de modo a satisfazer a obtenção do conceito requerido. Tais elementos léxicos foram associados de modo a formar várias regras gramaticais, os quais foram chamados de predicados. Levou-se em consideração regras de processamento da linguagem natural e as regras gramaticais da Língua Portuguesa. Assim, os predicados possibilitaram de maneira mais objetiva e explicativa a extração dos textos na forma de sentenças que podem ser usados na construção de mapas conceituais.

Para construir o mapa conceitual, tomou-se como base a sua forma clássica de representação, podendo ser expresso do seguinte modo. O conteúdo do mapa (M) é constituído por um conjunto de sentenças  $st_1... st_n$ , ou seja,  $M=\{st_1, st_2, st_3, ... , st_n\}$ . Para tanto, se faz necessário identificar um conjunto de proposições  $pr_1...pr_n$  que devem ser extraídos. A cada proposição diz-se que é composta por uma tupla de elementos no formato NOME-VERBO-NOME  $\{n_1- v_1- n_2\}$ . As quais serão as saídas mais relevantes de acordo com o texto especificado.

Para a avaliação, serão escolhidos trechos aleatórios de textos de diferentes áreas do conhecimento. Estes textos serão submetidos ao Text2MARK para extração dos elementos mais significativos conforme as regras que foram estipuladas. Após se obter os conjuntos de palavras serão construídos os mapas conceituais que sofrerão uma avaliação subjetiva quanto a sua similaridade com o texto fornecido.

Após isso, os mapas serão submetidos a grupos distintos de avaliadores que apresentam pouca ou nenhuma experiência em mapas conceituais. A avaliação irá obedecer as seguintes etapas:

1 – Fazer a leitura do texto fornecido e marcar as palavras que o avaliador considere como relevante e/ou chave;

2 – Fazer uma leitura do mapa conceitual partindo-se da esquerda para a direita e de cima para baixo buscando estabelecer uma relação de similaridade entre o texto e o mapa;

3 – Fornecer uma nota de 0 (zero) a 10 (dez), sendo zero nenhuma relação e dez muita relação.

## **1.5 Contribuição da dissertação**

Percebeu-se no estado da arte uma dificuldade para extração automatizada de sentenças em textos. Portanto, a contribuição desta dissertação é tratar do problema a partir da sua descrição formal e textual, em Português, de conceitos educacionais de qualquer natureza. A partir disso, será possível identificar de forma mais objetiva, dinâmica e explicativa quais são os conceitos e termos relevantes em um texto que devem ser extraídos. Estes por sua vez devem ser claros o suficiente para a obtenção de conceitos na forma de tuplas e que possam servir de entrada para construir um mapa conceitual que contemple a descrição textual fornecida.

## **1.6 Estrutura da dissertação**

O Capítulo 1, que trata da Introdução, apresentou o contexto no qual o trabalho se insere. Além disso, a motivação com a identificação do problema e a pergunta que norteou esta pesquisa. Foram também apresentados os objetivos gerais e específicos e a metodologia, assim como a contribuição desta dissertação. Em seguida serão descritos a estrutura dos demais Capítulos.

O Capítulo 2 aborda a fundamentação teórica no que tange a área de mineração de textos e uma breve revisão de gramática descritiva. Trata também dos mapas conceituais (MCs) e suas complexidades, assim como as principais ferramentas disponíveis atualmente para construção de MCs, por fim, as considerações finais sobre o Capítulo.

O Capítulo 3 faz uma revisão do estado da arte das extrações automatizadas associadas com mineração de textos para construção de mapas conceituais, assim como aborda a

definição do problema e identifica os problemas não resolvidos. Apresenta uma análise que comprova as arguições que justificam a necessidade da criação do Text2MARK.

O Capítulo 4 trata do método desenvolvido para solucionar o problema identificado no estado da arte, onde foi desenvolvida uma ferramenta chamada de Text2MARK, além de apresentar as definições gerais da ferramenta como: objetivo, definição do escopo, projeto técnico, arquitetura, as regras gramaticais, metodologia de extração e a escolha do escopo final e encerrando as considerações finais do Capítulo.

O Capítulo 5 trata da avaliação qualitativa e subjetiva do resultado alcançado, explicando quais os textos utilizados e os resultados, em tuplas, obtidos através do Text2MARK. Aborda também a inclusão das tuplas na ferramenta CMapTools para gerar um mapa conceitual. Apresenta os resultados obtidos na avaliação feita por três grupos distintos de usuários. Por fim trata das considerações finais do Capítulo.

O Capítulo 6 encerra esta dissertação tratando das Conclusões obtidas assim como quais os trabalhos futuros e os que ainda podem ser feitos na área de mineração de textos e mapas conceituais para aperfeiçoar e auxiliar o processo ensino-aprendizagem.

## **2 FUNDAMENTAÇÃO TEÓRICA**

Existem diversas formas de construir e representar o conhecimento. Tais métodos devem servir para melhorar e auxiliar o processo ensino-aprendizagem sem perder a essência do conhecimento obtido e/ou construído. Dessa forma, neste Capítulo serão apresentadas as técnicas utilizadas para respaldar as problemáticas identificadas no capítulo anterior. Com isso, faz-se necessário conhecer alguns conceitos: mineração de textos; gramática descritiva e mapas conceituais.

### **2.1 Características e definições sobre mineração de textos**

Mineração de texto se refere à descoberta de conhecimento anteriormente desconhecido que pode ser encontrado em coleções de texto. Um exemplo de ferramentas para auxiliar nessa descoberta são os etiquetadores de textos. Como visto no trabalho de Domingues et al (2007), “etiquetar é a tarefa básica de rotular palavras de uma sentença, com etiquetas morfossintáticas, que identificam as palavras como categorias gramaticais, tais como: i) substantivos; ii) verbos; iii) adjetivos; iv) advérbios; v) preposições, etc”. As etiquetas podem conter atributos para cada categoria, por exemplo, o gênero e o número para um substantivo.

A técnica de etiquetagem costuma apresentar o problema da ambiguidade que, em uma sentença, é normalmente resolvida com base no contexto em que se encontra a palavra. Palmeira et al (2012) ressaltam que “as pesquisas atuais sobre etiquetagem podem ser definidas em três grandes grupos: i) de regras; ii) probabilísticas e iii) as híbridas, que faz junção entre regras e probabilísticas”.

Em mineração de textos a associação entre documentos e modelos de base de dados pode ser resumida em alguns pontos: i) população de um esquema de banco de dados com os dados obtidos a partir de documentos da Web; ii) descoberta de informações existentes nos textos e armazenamento em um formato relacional ou XML; iii) integração e consulta de dados de

texto depois de ter sido armazenada em bases de dados; iv) duplicação de um conjunto de dados usando técnicas básicas de mineração de dados, tais como agrupamento (Stravrianou et al, 2007).

A percepção das principais dúvidas quanto se trata de mineração de textos foi bem identificada em Stravrianou et al (2007). A Tabela 1 traduz essa questão, em que ressalta a necessidade de se responder a várias perguntas quando se trata de mineração de textos. Porém, é possível chegar à conclusão de que o método de mineração utilizado pode associar mais de uma resposta sem necessariamente estar errado. Na Figura 1, Stravrianou et al (2007) mostra em 4 etapas os passos necessários na mineração de textos segundo alguns autores Yamasaki et al (2010), Dalmolin et al (2009a), Dalmolin et al (2009b), Dalmolin et al (2008) e Kornilakis et al (2004b).

Tabela 1. Os principais desafios da mineração de textos

#	Problemática	Dúvida gerada
1	Palavras de parada	Devemos levar em conta as palavras de parada?
2	Raízes	Deveríamos reduzir as palavras as suas raízes?
3	Ruídos	O texto deve ser limpo dos ruídos?
4	Palavras sem ambiguidade	Devemos esclarecer o significado das palavras em um texto?
5	Tagging	E quanto a dados de anotação e / ou parte de características da fala?
6	Colocações	Sobre o contexto ou termos técnicos?
7	Gramática / Sintaxe	Devemos fazer uma análise sintática ou gramatical? Ou sobre dependência de dados, problemas anafóricos ou a ambiguidade de escopo?
8	Tokenization	Devemos aplicar token nas palavras ou frases, e se sim, como?
9	Representação de texto	Quais termos são importantes? Palavras ou frases? Substantivos ou adjetivos? Qual o modelo de texto que devemos usar? O que dizer palavra de ordem, o contexto, fundo e conhecimento?
10	Aprendizagem automatizada	Devemos usar categorização? Quais medidas de similaridade devem ser aplicadas?

De acordo com Dalmolin et al (2009a). Além das perguntas identificadas anteriormente, o processo de mineração de textos, requer algumas etapas, como: i) cálculo da frequência; ii) parsing (avaliar e compor regras); iii) reconhecimento de padrões; iv) análise semântica; v) tokenização (palavras-chave); vi) heurísticas. Em outra abordagem do autor, observa-se também a utilização de matrizes, de adjacência ou sobreposição, para formar links bi-direcionais (Dalmolin et al, 2008).

Para associar os resultados da mineração de textos com os MCs, os fatores principais são as entradas, ou seja, os textos que geralmente são em forma de tuplas. Vale ressaltar que a

principal prerrogativa de se utilizar a mineração de textos neste contexto é a de simplificar a obtenção dessas tuplas. Tal observação é tida como prioritária no trabalho de Moreira (1997), uma vez que as ferramentas disponíveis atualmente se apresentam com certo grau de dificuldade em sua usabilidade, necessitando do usuário uma maior compreensão. Essa dificuldade é observada nos trabalhos de Gouli et al (2004), Dalmolin et al (2009a), e Kornilakis et al (2004a).

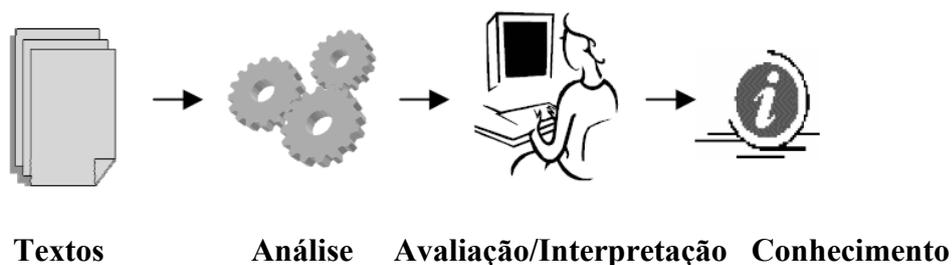


Figura 1. O processo de mineração de textos. Stravrianou et al (2007)

Percebe-se ainda que além da Estatística, vários modelos precisam utilizar técnicas de aprendizado de máquina para dinamizar seus processos e até tomar decisões melhores. Entretanto, durante o processo de mineração, vários problemas surgem, alguns em particular por conta do processamento de linguagem natural (Stravrianou et al, (2007), como sinônimos ou grafia errada, além das contextualizações (Yamazaki et al, 2010). Com isso, decisões precisam se tomadas para não prejudicar o resultado final.

## 2.2 Uma revisão de Gramática descritiva

Um dos princípios do estudo da gramática, segundo Perini (1995, pg 44), é “tomar como base a descrição das línguas. Tal premissa torna-se importante porque a tradição de estudo gramatical frequentemente despreza a fundamentação teórica e metodológica”. O resultado deste lapso é uma descrição cheia de inconsistências e por vezes sem grande conexão com a língua real.

Em sintaxe, observa-se que o método tradicional para a obtenção de dados depende principalmente da análise dos falantes sobre frases isoladas, ou seja, o domínio do conteúdo analisado. Às vezes, pode-se recorrer ao exame de textos, porém, este recurso se torna impraticável para o todo da análise sintática, justamente porque iria requerer a manipulação

de quantidades imensas de material. Dessa forma, para alcançar uma análise válida, construímos exemplos e utilizamos o nosso próprio julgamento.

Segundo Perini (1995, pg 48), “seria desejável elaborar uma metodologia para elucidar dados, interpretar a variação de julgamento e incorporá-las na descrição. Tal recurso, mais voltado para a área da sociolinguística, iria auxiliar a compreensão do conteúdo”. Ainda de acordo com Perini (1995, p37), “uma gramática só pode apresentar um compromisso aceitável entre o retrato fiel da realidade (o que complicaria além do razoável) e a simplificação excessiva (o que lhe tiraria a relevância)”.

Entende-se a partir desta premissa que a descrição de uma língua se apresenta composta basicamente de três elementos: uma descrição formal; uma descrição semântica; e um sistema que faz relação entre a descrição formal e a descrição semântica. Identifica-se como descrição formal a fonologia, morfologia e sintaxe; Jean-Marya et al (2009) afirmam “já o sistema que correlaciona à semântica com descrição formal é chamado de regras de interpretação semântica”.

Na gramática, a fonologia, a morfologia e a sintaxe são todas formadas por regras, assim como a semântica. A diferença principal é que esta fornece as relações entre as construções da língua e seus significados, enquanto que aquelas definem quais são as construções possíveis na língua. De acordo com Perini (1995, p41), em geral, “pode-se dizer que os traços formais da língua sofrem interpretação, ou seja, a cada traço formal corresponde algum traço de significado”. Mas há traços formais que não têm nenhuma interpretação, como por exemplo, certos verbos têm complemento sem preposição, como observado na oração 1, já outros exigem a presença da preposição, como na oração 2.

(1) *Meu filho detesta alface*

(2) *Meu filho gosta **de** alface*

Tal exigência ou não da presença da preposição antes do complemento verbal é um traço puramente formal; nenhuma diferença semântica está acompanhando a diferença entre a presença ou ausência da preposição. Dessa forma, segundo Chomsky (1965, pg 88) e Perini (1995, pg 42) “este é um fenômeno gramatical que só pode ser estudado no plano formal”.

Da perspectiva verbal, faz-se necessário compreender que dentro da própria linguagem nem todos os elementos existentes devem ser levados em consideração para a construção

formal de um conceito. Existem situações literárias que fogem a nossa compreensão normal sobre um determinado assunto, como por exemplo, os poemas, as poesias, onde, em regra geral, sua construção não obedece a todos os conceitos gramaticais.

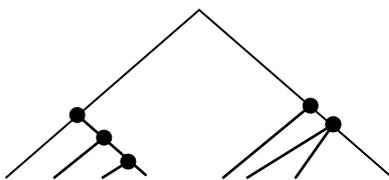
Em Perini (1995, p 43-45), faz uma definição de algumas regras que devem ser levadas em consideração para os linguístas, o que ele define como “gírias de ofício, que também são chamados de ‘fatos sintáticos’”. Dentre os fatos, ressalta-se para o contexto desta dissertação a necessidade da compreensão do que é a representação em posição linear e os dos constituintes.

A posição linear é simplesmente a posição que uma unidade ocupa em relação às outras unidades do enunciado. Muitas vezes faz-se necessário formular propriedades sintáticas de um item, como por exemplo, uma palavra. Já os constituintes são definidos como certos grupos de unidades que fazem parte de sequências maiores, mas que mostram certo grau de coesão entre eles. As orações 3 (a) e (b) e 4 exemplificam a posição linear e os constituintes, respectivamente.

- (3) a. *Todos os motoristas entraram em greve;*  
 b. *Os motoristas entraram todos em greve;*
- (4) *A casa de Lulu é azul e branca*

A representação da estruturação de um constituinte é frequentemente representada por meio de um diagrama em forma de árvore. Na Figura 2, há um exemplo simples dessa representação. Faz-se necessário ter boa compreensão da estruturação das frases em um constituinte porque toda sua análise se baseia neste modelo Chomsky (1965, pg 88). Perini (1995, pg 42) finaliza seu pensamento sobre constituinte ressaltando que:

Os falantes têm muitas vezes intuições bem definidas sobre os constituintes; em outros casos as intuições não são claras, e é preciso lançar mão de outros recursos, como tentar observar o comportamento sintático de uma sequência em outras frases.



*A casa de Lulu é azul e branca*

Figura 2. Representação da construção de constituintes. Adaptado de Perini (1995, pg44)

Durante o processo de construção gramatical, além dos elementos citados, ressaltamos a presença dos advérbios transpostos, os quais podem ocorrer em diversas posições dentro da oração e com isso resultam em grupos de correspondência. Dessa forma, segundo Perini (1995, pg48), “pode-se ressaltar o predeterminante transposto, como os elementos: i) todos; e ii) ambos, que também podem ocorrer em diversas posições”.

Percebeu-se com isso uma relação interessante de que nem todos os elementos que constituem uma oração na forma ativa podem ser classificados como simétricos, ou de correspondência total, pois existirão casos de que uma frase composta de: sujeito + verbo + objeto, não terá sua representação gramatical correta quando for expresso na sua forma simétrica: objeto + verbo + sujeito. A oração 5 (a) e (b) mostra como pode ser obtida e representada uma correspondência total ou simétrica. Já em 5 (c) mostra claramente a problemática do predeterminante transposto, inexistindo a simetria.

- (5)     a. *O gato comeu o rato*  
           b. *O rato, o gato comeu*  
           c. *O rato comeu o gato*

Perini (1995, p-51) define de forma clara como se articulam os componentes da gramática para, conjuntamente, definir quais são as sequências que constituem frases corretas da língua e quais as que não constituem. Dessa forma:

Frases bem construídas têm como resultado que os falantes a aceitem como sendo uma frase legítima do português. Em contrapartida, as frases mal construídas são rejeitadas pelos falantes, sejam elas por razões morfológicas ou sintáticas.

Neste caso, surge o conceito de sintaxe (Perini, 1995, p-62), que “é a parte da gramática que estuda as orações e suas partes – ou seja, a estrutura interna da oração”.

Para a compreensão dos elementos que se articulam para constituir frases, assim como para validar a importância na ordem dos elementos nas frases, tal qual ter o intuito de identificar quais os elementos necessários no processo de extração da mineração de textos, buscou-se fundamentar o estado da arte com os principais autores, porém, a melhor definição encontrada foi em Perini (1995, p 68-69), em que remete a explicação da hierarquia dos constituintes. Esta será a abordagem dos parágrafos seguintes.

A oração se estrutura de maneira hierárquica, isto é, contém constituintes que, por sua vez, contêm outros constituintes. É preciso levar esse fato em conta ao se fazer a análise. Por exemplo, digamos que ao se analisar a oração (6), é possível fazer um primeiro corte,

definindo os grandes constituintes (ou sintagmas) da oração, da seguinte forma, como na oração (7)

(6) *Meus vizinhos arranjaram um cachorro horrivelmente barulhento*

(7) *[Meus vizinhos] – [arranjaram] – [um cachorro horrivelmente barulhento]*

Esses são os constituintes imediatos da oração; cada um deles terá uma função especial, tais funções se denominam: i)sujeito; ii)predicado; e iii)objeto direto, respectivamente. Porém, é possível identificar também que alguns desses constituintes têm, por sua vez, uma estrutura interna sintática. Dessa forma, [meus vizinhos] se divide em [meus] e [vizinhos], e cada uma dessas palavras tem sua função sintática dentro do sintagma [meus vizinhos]. Perini (1995, p-69) chama a essas funções de “possessivo” e “núcleo do sintagma nominal” (NSN), respectivamente.

O constituinte [arranjaram] é sintaticamente simples, por ser formado de uma única palavra, o que impede de ser analisado em termos sintáticos, apenas morfológicos. Já o constituinte [um cachorro horrivelmente barulhento], por sua vez, é complexo, e precisa ser analisado. Começa com um “determinante” [um], um “núcleo do sintagma nominal” [cachorro] e um “modificador” [horrivelmente barulhento]. Os três termos são constituintes imediatos do SN [um cachorro horrivelmente barulhento]. Por fim, o constituinte que funciona como modificador desse sintagma, [horrivelmente barulhento], ainda pode ser analisado sintaticamente, [horrivelmente] é o “intensificador” e [barulhento] é o “núcleo do sintagma adjetivo”. (Perini, 1995, p.70).

A análise da hierarquia na estruturação da oração demonstra que o termo [meus vizinhos] é o sujeito da oração; mas o termo [meus] não é o possessivo da oração; é o possessivo do SN [meus vizinhos]; [meus] só é parte da oração indiretamente, porque faz parte de um sintagma que por sua vez faz parte da oração. Da mesma forma, [horrivelmente] não é intensificador da oração, nem do SN [um cachorro horrivelmente barulhento]; é o intensificador do sintagma adjetivo [horrivelmente barulhento]. Esse sintagma adjetivo faz parte do SN [um cachorro horrivelmente barulhento], e esse SN faz parte da oração. A Figura 3 trata da representação gráfica da oração.

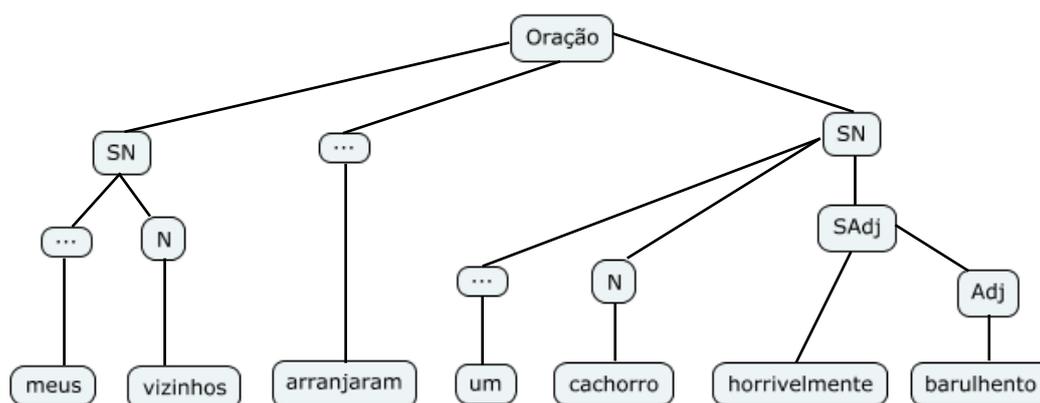


Figura 3. Representação da hierarquia dos constituintes identificando elementos da frase. Adaptado de Perini (1995, p-69)

Uma vez identificados os elementos que compõem os sintagmas nominais e a forma como eles desempenham funções diferentes, razão pela qual não podem receber a mesma definição. Faz-se necessário compreender o elemento que liga esses sintagmas, no caso da oração (6) o verbo [arranjaram]. É possível associar outras funções ao referido constituinte, porém, percebe-se que a função mais clara desempenhada por ele é de “núcleo do predicado” (NdP).

Dessa forma, Perini (1995, p.71) parte do seguinte postulado: “o verbo desempenha na oração unicamente a função de núcleo do predicado; essa é a única função que um verbo pode desempenhar, e somente um verbo pode ser núcleo do predicado”. Em outras palavras, o verbo é sempre o NdP da oração; e o NdP da oração é sempre um verbo. Admite-se ainda que “o NdP faça parte do constituinte chamado “predicado” (Pred), então, o constituinte de nível oracional, estritamente falando, é o Pred, que contém o NdP”. Ressalta-se que exceto nos casos onde houver dois verbos em uma única oração, o verbo em forma finita (conjugado) é chamado de verbo auxiliar (VAUX), e o verbo em forma não-finita (particípio, gerúndio ou infinitivo) é o NdP.

Dos elementos gramaticais que foram pesquisados para esta dissertação, cinco se apresentam com um grande número de regras, são eles: i) sujeito; ii) objeto direto; iii) complemento do predicado; iv) predicativo e v) atributo.

A construção restrita em tais elementos se dá pelo fato de cada um deles desempenhar um número muito grande de sintagmas. Como é ressaltado por Perini (1995, p-85) seria impossível listar todos os sujeitos em português. Justamente por não incluir apenas palavras individuais, o que iria gerar um número ilimitado de sintagmas.

O léxico torna-se responsável por definir quais as regras os falantes da língua aceitam ou não para a construção dos enunciados. É caracterizado como uma longa lista onde a informação idiossincrática da língua é armazenada. Guarda um conjunto de características fonológicas, morfológicas, sintáticas e semânticas de uma palavra.

## **2.3 Mapas Conceituais**

De uma maneira ampla, mapas conceituais são apenas diagramas que indicam relações entre conceitos. Mais especificamente, podem ser interpretados como diagramas hierárquicos que procuram refletir a organização conceitual de um corpo de conhecimento ou de parte dele. Ou seja, sua existência deriva da estrutura conceitual de um conhecimento. (Moreira, 2006, p.9).

Usualmente os mapas conceituais podem ser confundidos com organogramas, porém, diferente destes, os mapas têm como objetivo tornar evidentes os significados atribuídos a conceitos e evidenciar quais são as relações existentes entre os mesmos. A aplicabilidade dos mapas conceituais independe da área de conhecimento, do curso, disciplina ou qualquer forma textual, exceto a poética, pois seu objetivo é representar na forma gráfica um conceito ou definição.

### **2.3.1 Principais Conceitos**

De uma maneira mais específica podemos definir com Moreira (2006), que os mapas conceituais devem ser entendidos “como diagramas bidimensionais que procuram mostrar relações hierárquicas entre conceitos de um corpo de conhecimento e que derivam sua existência da própria estrutura conceitual desse corpo de conhecimento”.

Questões acerca da utilização dos mapas conceituais são pontuais. Sabe-se que é possível construir um mapa de diversas formas de acordo com o domínio do especialista. Com isso, é necessário entender que não existe apenas um mapa certo. Como diz Moreira (2006, p.10) “O ponto importante é que um mapa conceitual deve ser sempre visto como "um mapa conceitual", não como "o mapa conceitual" de um determinado conjunto de conceitos”.

De fato, a presente pesquisa também aponta nesta direção, uma vez que o mapa conceitual sempre depende da profundidade de conhecimento do assunto de quem o está elaborando. Neste aspecto, ressalta-se que realizar comparações entre mapas conceituais elaborados por dois ou mais especialistas de um determinado assunto pode apresentar resultados diferentes.

Isso se dá pelo fato de que cada especialista possui uma percepção particular sobre o tema, onde irá buscar dar ênfase ou maior detalhe a um determinado traço do contexto. Porém, não significa que um mapa mais ou menos detalhado esteja errado, pois o mesmo apenas apresenta uma perspectiva diferente sobre o tema.

Em princípio, os mapas conceituais podem ter uma, duas ou mais dimensões. Os mapas lineares, de organização vertical, dão uma visão macro do conceito. Já os bidimensionais, dispostos de forma vertical e horizontal, apresentam maiores relações entre seus conceitos, dando maior detalhamento sobre o tema.

Geralmente os mapas conceituais apresentam uma disposição hierárquica com os conceitos mais abrangentes ou mais inclusivos situados no topo e os conceitos mais específicos ou secundários distribuindo-se em direção à base, que assume então uma configuração piramidal. Grillo et. al, (2008) afirma que “os mapas conceituais não têm um formato pré-estabelecido, pois dependem muito da estrutura conceitual do conteúdo”.

Ao se tomar como base a obra de Ausubel “Psicologia Educacional” (Ausubel et al, 1980), o autor traduz seu pensamento de uma forma que justifica a utilização da teoria de mapas conceituais no processo de ensino-aprendizagem. “O armazenamento da informação no cérebro humano é altamente organizado, formando uma hierarquia conceitual na qual os elementos específicos do conhecimento são ligados (e assimilados) a conceitos mais gerais e inclusivos” (Ausubel, 1995, p.46). A Figura 4 traduz a organização definida por Ausubel (2003)

Em outro momento Ausubel (1980, p.53) ressalta. “Cada disciplina possui uma estrutura de conceitos hierarquicamente organizados: conceitos mais gerais e inclusivos situam-se no topo da estrutura e incluem conceitos cada vez menos inclusivos e mais diferenciados”.

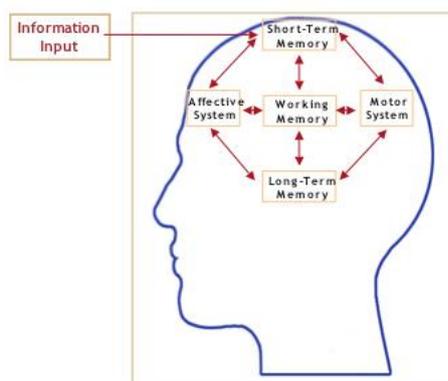


Figura 4. Sistemas de memória chave do cérebro onde todos os elementos interagem quando estamos aprendendo. Fonte:<http://cmap.ihmc.us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps.htm>

### 2.3.2 As representações

Mapas conceituais permitem a representação semi estruturada do conhecimento de um domínio de aplicação, proporcionando uma melhora significativa aos problemas decorrentes do uso da linguagem natural (Palmeira et al, 2012). Com isso, a formação dos elementos que resultam na obtenção de um conceito, como de um objeto de uso comum a todos, por exemplo, uma mesa, não se faz necessário maiores explicações de suas características e funcionalidades. Entretanto, existem particularidades na maneira em que cada indivíduo vê uma mesa, que reflete a forma idiossincrática que formamos o conceito.

De forma particular, a definição do conceito de mesa vem com o passar dos tempos, pois cada um de nós formou um conceito próprio do que seria este objeto. Tais características idiossincráticas vêm sendo definidas com as formas e situações que todos visualizamos e vivenciamos desde que nascemos e começamos a trilhar nossas próprias experiências.

Conforme Silva et al (2009), dois pressupostos se diferenciam no armazenamento das estruturas cognitivas do ser humano:

1.É menos difícil para os seres humanos apreenderem os aspectos diferenciados de um todo, anteriormente apreendido e mais inclusivo, do que formular o todo inclusivo a partir das partes diferenciadas anteriormente aprendidas;

2.A organização que o indivíduo faz do conteúdo de uma determinada disciplina no próprio intelecto consiste numa estrutura hierárquica, onde as idéias mais inclusivas ocupam uma posição no vértice da estrutura e subsumem, progressivamente, as proposições, conceitos e dados factuais menos inclusivos e mais diferenciados.” (Ausubel, 2003: 166)

Ainda conforme o estudo de Silva et al (2009), embora os mapas conceituais possam transmitir informações factuais tão bem quanto os textos, “esses organizadores gráficos são mais efetivos que os textos para ajudar os leitores a construir inferências complexas e integrar as informações que eles fornecem” (Cimiano et al ,2005); (Baader et al, 2003); (Arlotta et al, 2003). Eles também têm o potencial de melhorar a acessibilidade e usabilidade materiais durante uma pesquisa na medida em que apresentam marcas visuais-espaciais que podem guiar uma seleção ou categorização.

Há uma grande variedade de tipos de mapas disponíveis que foram imaginados e construídos pelas mais diversas razões. Alguns são preferidos pela facilidade de elaboração

(tipo aranha), pela clareza que explicita processos (tipo fluxograma), pela ênfase no produto que descreve, ou pela hierarquia conceitual que apresenta. Existe a comprovação empírica sobre a eficiência dessas buscas de conteúdo, onde os interessados localizam mais informações quando elas são apresentadas em formas de mapas ao invés de textos (Tao et al, 2009) e Sarawagi (2008).

Outro modelo, o tipo fluxograma, organiza a informação de uma maneira linear. Ele é utilizado para mostrar passo a passo determinado procedimento, e normalmente inclui um ponto inicial e outro ponto final. Um fluxograma é normalmente usado para melhorar o desempenho de um procedimento (AL-Kamha, 2007).

A aplicabilidade dos mapas conceituais no desenvolvimento de software, em específico na área da análise de requisitos, (Rosson et al, 2007); (Sayão et al, 2003) e Robinson (2004) apresenta relevante importância justamente por essa etapa ser de grande complexidade para o usuário final, o qual não está acostumado com as técnicas da Engenharia de Software. Ressalta-se a dificuldade que muitos usuários tiveram na utilização das ferramentas apresentadas pelos autores para criação dos mapas conceituais da análise de requisitos.

Existem outros processos de construção de software (Rolim, 2006), (MPS.BR, 2011), como a elicitação, (Cohn, 2011); (Silva, 2011); (T.H. et al, 1991) que fazem uso de mapas conceituais (Palmeira et al, 2012). Por exemplo, quando se deseja otimizar um determinado processo, a utilização do mapa tipo fluxograma mostrado na, Figura 5, é a representação mais adequada. Esse tipo de mapa deixa claro quais são as confluências e as possíveis opções a serem escolhidas. Por sua facilidade de compreensão, esse modelo ainda é utilizado quando se deseja construir um algoritmo eficiente para determinada função. (Sommerville, 2011)

As principais vantagens observadas são a sua facilidade de leitura e as suas informações que estão organizadas de uma maneira lógica e sequencial. Em contraponto sua desvantagem está na ausência de pensamento crítico, que normalmente é incompleto na exposição do tema. Ele é construído para explicitar um processo, sem a preocupação de explicar determinado tema; na sua gênese não pretende facilitar a compreensão do processo, mas aperfeiçoar a sua execução.

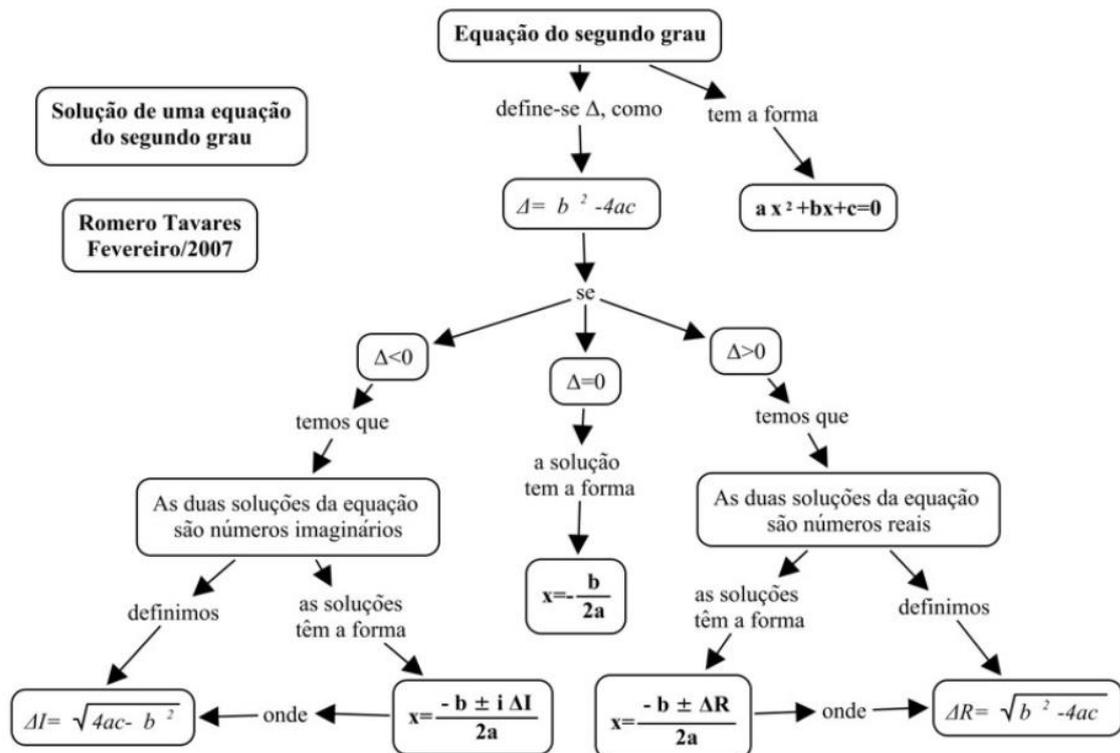


Figura 5. Representação em fluxograma. Adaptado Tavares (2007)

No entanto, o único tipo de mapa que explicitamente utiliza uma teoria cognitiva em sua elaboração é o mapa hierárquico do tipo proposto por Novak & Gowin (1984). Dentre as formas citadas de mapas conceituais é preciso observar alguns detalhes e características dos tipos disponíveis. Como por exemplo, o modelo Teia de Aranha, mostrado na Figura 6, que é organizado colocando-se o conceito central (ou gerador) no meio do mapa.

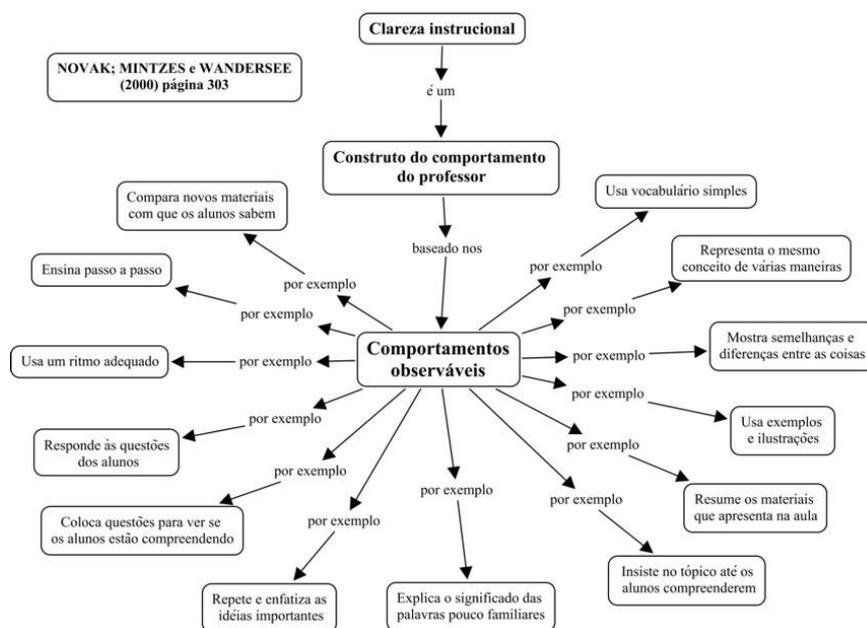


Figura 6. Mapa conceitual modelo Teia de Aranha. Adaptado Silva et al (2009)

Neste modelo os demais conceitos vão se irradiando na medida em que nos afastamos do centro. Sua vantagem se apresenta como a facilidade de estruturação, pois todas as informações estão unificadas em torno de um ou vários temas centrais. O foco principal é a irradiação das relações conceituais, sem preocupação com as relações hierárquicas, ou transversais (Wang et al, 2006). Em contrapartida, a desvantagem do modelo está justamente na dificuldade em mostrar as relações entre os conceitos, e desse modo permitir a percepção de uma integração entre as informações (Michelson et al, 2007).

Outro formato também utilizado na construção dos mapas conceituais é o tipo Sistema, chamado de entrada e saída. O qual organiza a informação num formato que é semelhante ao fluxograma, mas com o acréscimo da imposição das possibilidades “entrada” e “saída”.

Assim como os modelos já apresentados, este também possui vantagens e desvantagens. A principal vantagem é justamente poder mostrar as várias relações entre os conceitos. Sua desvantagem é que algumas vezes é difícil de ler devido ao grande número de relações entre os conceitos. É adequado para explicar processos que impliquem em entrada e saída. A Figura 7 mostra sua representação.

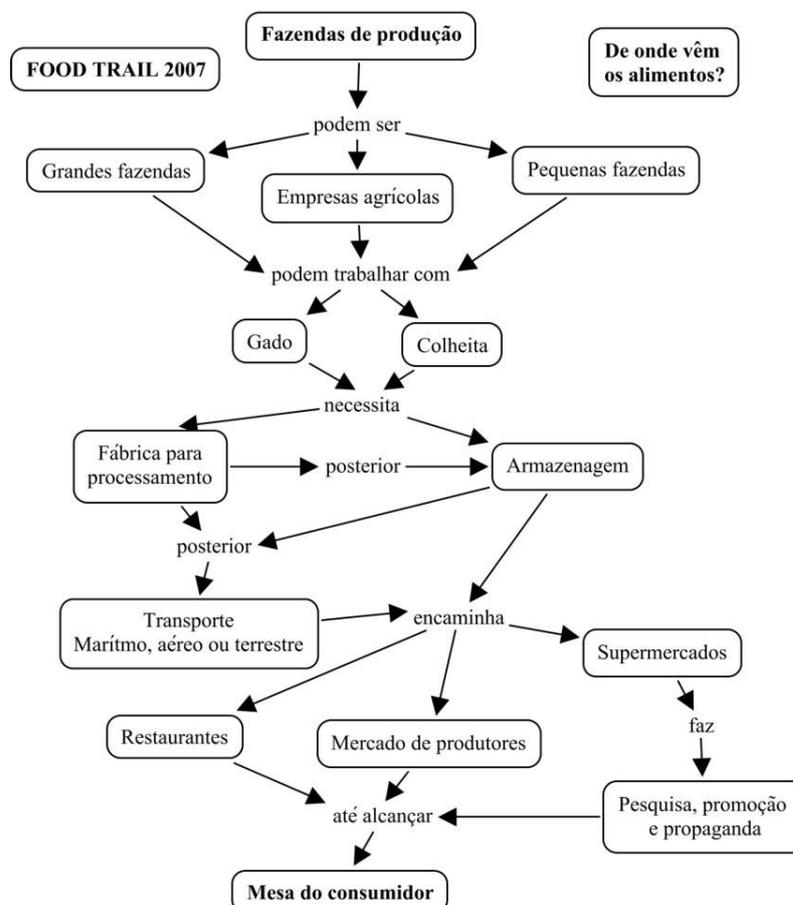


Figura 7. Mapa conceitual modelo Entrada e Saída

O modelo que está consoante com a “Teoria Ausubeliana” é o hierárquico (Figura 8) em que a informação é apresentada numa ordem descendente de importância. A informação mais importante (inclusiva) é colocada na parte superior. Um mapa hierárquico é usado para nos dizer algo sobre um procedimento. Sua principal vantagem é que os conceitos mais inclusivos estão explícitos; os conceitos auxiliares e menos inclusivos estão inter-relacionados. Estrutura o conhecimento de maneira mais adequada a compreensão humana, considerando em posição de destaque os conceitos mais inclusivos. (Silva et al, 2009), (Novak et al, 1998).

Sua desvantagem é justamente por ser o mais difícil de externar e construir, visto que expõe a estrutura cognitiva do autor sobre o assunto. Silva et al (2009) ressaltam ainda que “a clareza do autor sobre o tema fica evidente quando da sua construção. A sua construção sempre representa um desafio, visto que explicita (principalmente para si) a profundidade do conhecimento do autor sobre o tema do mapa”.

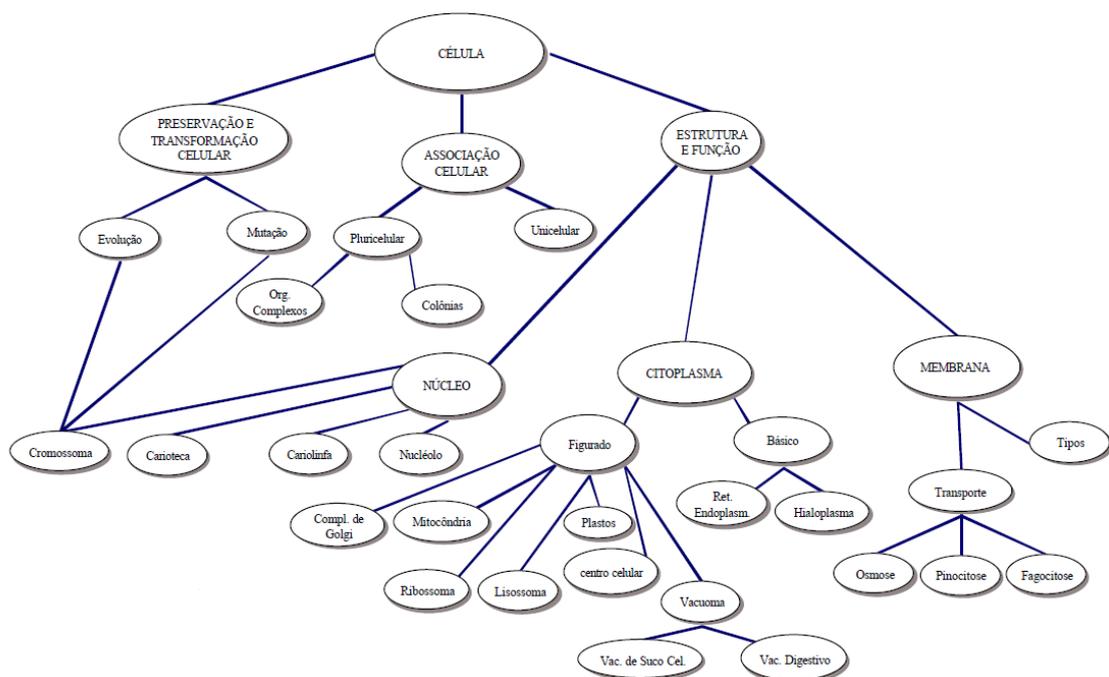


Figura 8. Mapa hierárquico apresenta de forma simples a compreensão de conceitos. Adaptado de Moreira et al (1982)

Existem opiniões diversas com relação aos mapas e suas utilidades, porém, cabe a cada um se ater na aplicabilidade do modelo e como ele pode ser útil na formação do conhecimento, relacionado sempre a sua percepção de auxílio do aprendiz ou do especialista.

O que nos remete a entender que o processo de ensino-aprendizagem requer evoluir e agregar técnicas atuais.

### 2.3.3 Uso prático de Mapas Conceituais

O uso prático dos mapas conceituais vai além da forma tradicional de representar conhecimento. Compreender e entender essa aplicabilidade são alguns dos objetivos que se buscam neste trabalho. A importância de técnicas atuais de auxílio no processo de ensino-aprendizagem é vital para acompanhar a evolução e maturidade da atual geração que dispõe de vários recursos tecnológicos e várias fontes de informação simultaneamente.

#### 2.3.3.1 Mapas conceituais como instrumentos didáticos

Moreira (2006, p.16) afirma que os mapas conceituais podem “mostrar as relações hierárquicas entre os conceitos que estão sendo ensinados em uma aula, em uma unidade de estudo ou em um curso inteiro. Eles explicitam relações de subordinação e superordenação que possivelmente afetarão a aprendizagem desses conceitos”. Tais representações das estruturas conceituais são mostradas de forma concisas que por sua vez, provavelmente, facilitarão sua aprendizagem. A Figura 9 traduz de forma simples e eficiente como ocorre essa organização.

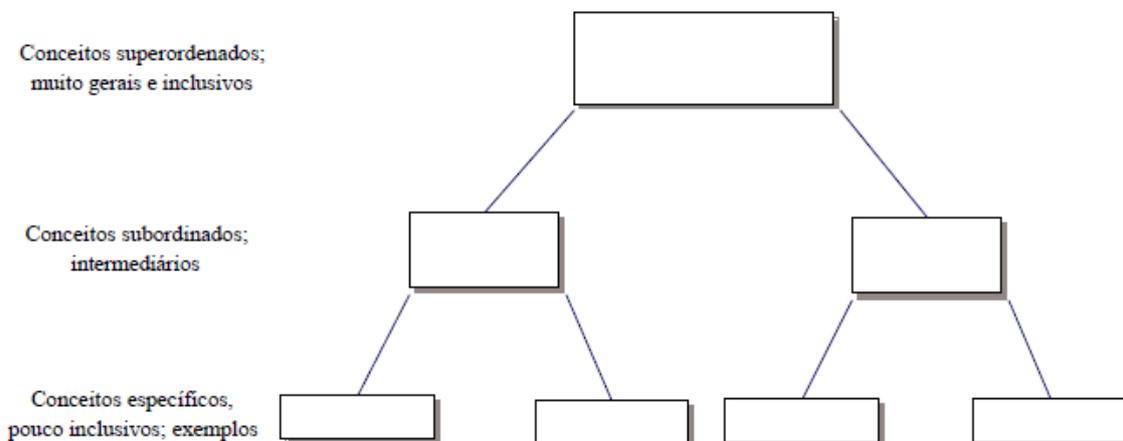


Figura 9. Representação hierárquica da associação de conceitos

No processo ensino-aprendizagem é visto como uma forma de estabelecer uma relação entre os conceitos tradicionais de ensino e a permitir uma nova perspectiva do conteúdo que se ministra. Essa metodologia aplicada ao ensino deve ser pautada de forma a dar suporte para que os alunos se sintam à vontade para receber os conceitos passados em mapas e respeitando o tempo necessário de amadurecimento do conhecimento.

Ressalta-se que a utilização dos mapas conceituais não dispensa a explicação e familiarização com o mesmo. Uma vez todo e qualquer novo conhecimento precisa ser introduzido pelo professor de forma a deixar que a interação do mapa pelo aluno seja feita naturalmente e que este chegue as suas próprias conclusões baseadas no tema que se estuda.

Para Ausubel (1978, 1980), “A instrução deve ser planejada não somente para promover a diferenciação progressiva, mas também para explorar, explicitamente, relações entre proposições e conceitos, evidenciar semelhanças e diferenças significativas”. Dessa forma a construção do conhecimento cognitivo por parte do aluno fica mais forte uma vez que ele próprio estabelece os vínculos e relações que acha importante naquele contexto em específico. Porém, sem deixar de compreender a essência do que está sendo ministrado.

Novak (1977, 1981) diz, “para conseguir a reconciliação integrativa de maneira mais eficiente, a instrução deve ser organizada de tal forma que se ‘baixe e suba’ nas hierarquias conceituais à medida que a nova informação é apresentada”. Isso significa que, embora no enfoque “Ausubeliano” se deva começar com os conceitos mais gerais, é necessário mostrar logo como os conceitos subordinados estão relacionados com eles e, então, voltar, por meio de exemplos, a novos significados para os conceitos de ordem mais elevada na hierarquia. Em outras palavras, de acordo com Moreira et al (1982) “se deve ‘baixar e subir’ no mapa, explorando, explicitamente, as relações de subordinação e superordenação entre os conceitos”.

Segundo Moreira (2006, p.17), no processo ensino-aprendizagem, “o uso de mapas conceituais feitos pelo professor apresenta vantagens e desvantagens. Entre as possíveis vantagens, pode-se mencionar (Moreira, 1979; Moreira e Buchweitz, 1993)”:

1. Enfatizar a estrutura conceitual de uma disciplina e o papel dos sistemas conceituais em seu desenvolvimento;
2. Mostrar que os conceitos de certa disciplina diferem quanto ao grau de inclusividade e generalidade e apresentar esses conceitos em uma ordem hierárquica de inclusividade que facilite sua aprendizagem e retenção;

3. Proporcionar uma visão integrada do assunto e uma espécie de "listagem conceitual" daquilo que foi abordado nos materiais instrucionais.

Dentre as possíveis desvantagens são citadas de acordo com os autores:

1. Se o mapa não tem significado para os alunos, eles podem encará-lo como algo mais a ser memorizado;
2. Os mapas podem ser muito complexos ou confusos e dificultar a aprendizagem e retenção, ao invés de facilitá-las;
3. A habilidade dos alunos em construir suas próprias hierarquias conceituais pode ficar inibida em função de já receberem prontas as estruturas propostas pelo professor (segundo sua própria percepção e preferência).

### 2.3.3.2 *Mapas conceituais como instrumentos de avaliação.*

Outra possibilidade de uso dos mapas conceituais está na avaliação da aprendizagem. Avaliação não com o objetivo de testar conhecimento e dar uma nota ao aluno, a fim de classificá-lo de alguma maneira, mas no sentido de obter informações sobre o tipo de estrutura que o aluno vê para um dado conjunto de conceitos. Para isso, pode-se solicitar ao aluno que construa o mapa ou este pode ser obtido indiretamente através de suas respostas a testes [...] construídos a partir de entrevistas (Moreira & Novak, 1987). (Moreira, 2006, p.19)

Portanto, o uso de mapas conceituais como instrumentos de avaliação implica uma postura que, para muitos, difere da usual. Palmeira et al (2012) ressalta que na avaliação “por meio de mapas conceituais a principal idéia é a de avaliar o que o aluno sabe em termos conceituais, isto é, como ele estrutura, hierarquiza, diferencia, relaciona, discrimina, e integra conceitos de uma determinada unidade de estudo, tópico, disciplina, etc”.

Moreira (2006) define que “a estrutura cognitiva de um indivíduo, em certa área de conhecimento, como o conteúdo e organização conceitual de suas idéias nessa área, mapas conceituais podem ser usados como instrumentos para representar a estrutura cognitiva do aprendiz”.

Ao utilizar mapas para fazer avaliações é necessário compreender o domínio com que o aluno estabelece as relações entre os conceitos existentes. Como já foi observado, assimilar um determinado conhecimento depende do quanto o aluno domina o tema. Dessa forma uma visão resumida do mapa de um determinado aluno não significa dizer que está incompleto se comparado a um mapa mais detalhado de outro aluno.

A utilização dos mapas conceituais como avaliação geralmente trás problemas ao aluno justamente por que este ainda não tem a percepção completa do assunto que está aprendendo. Com isso alguns cuidados e detalhes são importantes. Araújo et al (2002) propõem um ambiente para avaliação do processo ensino-aprendizagem baseado em mapas conceituais os quais servem para gerar facilidades tanto para o professor quanto para o aluno. Seu objetivo é integrar as tarefas de construção e avaliação dos mapas, facilitando o acompanhamento e evolução do aluno e de uma turma como um todo, oferecendo inclusive um *feedback* ao aluno.

### 2.3.3.3 *Mapas conceituais como recurso para análise de conteúdo*

Percebe-se na construção dos mapas que, segundo Moreira (2006), os “conceitos abrangentes, integradores, podem servir de base para o planejamento curricular de um determinado curso, enquanto conceitos mais específicos, pouco inclusivos, podem orientar a seleção de materiais e atividades instrucionais específicos”.

Apesar de apresentar o conteúdo como resultado ressalta-se que a ausência de uma padronização neste tipo de abordagem apresenta certa dificuldade de entendimento, ainda mais quando o mapa for muito resumido ou elaborado por quem não detém o domínio do assunto, como visto em Hartfield (1991). Dessa forma é importante ressaltar o domínio do especialista sobre um assunto para alcançar o resultado esperado pelo usuário aprendiz, que é sua análise de forma adequada.

Stewart et al. (1979) ressaltam que mapas conceituais podem “ser ferramentas importantes para focalizar a atenção do planejador de currículo para o ensino de conceitos e para distinção entre conteúdo curricular e conteúdo instrumental”. Ou seja, entre o conteúdo que se espera que seja aprendido e aquele que servirá de veículo para a aprendizagem.

Em Moreira (2006), “um bom planejamento de currículo implica uma cuidadosa análise de quais são os conceitos centrais para o entendimento da disciplina, ou parte da disciplina, que está sendo considerada”. Destaca Moreira (2006) que os “mapas conceituais podem ser extremamente úteis nessa tarefa”. A Figura 10 mostra o mapeamento conceitual do conteúdo de eletromagnetismo que serviu de guia para um curso introdutório nesse assunto Moreira (1977). Na coluna central estão os conceitos mais abrangentes e que são comuns aos fenômenos elétricos (lado esquerdo do mapa) e aos magnéticos (lado direito).

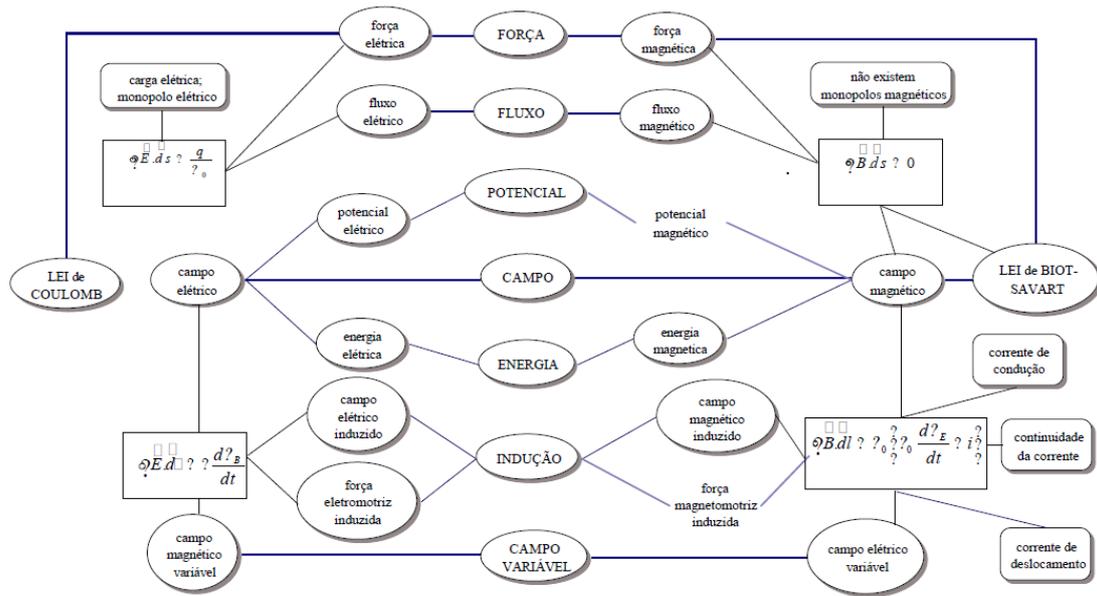


Figura 10. Um mapa conceitual para o conteúdo de Eletromagnetismo (Moreira, 1977, 1979, 1983; Moreira e Buchweltz, 1993). Neste mapa explora-se a simetria entre conceitos e leis existentes neste campo conceitual da Física

Mostrar que existem formas distintas de representar conhecimentos relevantes dentro de um contexto sem perder sua essência é visto com clareza nas Figuras 10 e 11 adaptadas de Moreira (2006). Isto apenas reforça o que já foi enaltecido de que não existe apenas um mapa conceitual para um determinado conteúdo. Tais percepções endossam o entendimento deste trabalho acerca dessa prerrogativa.

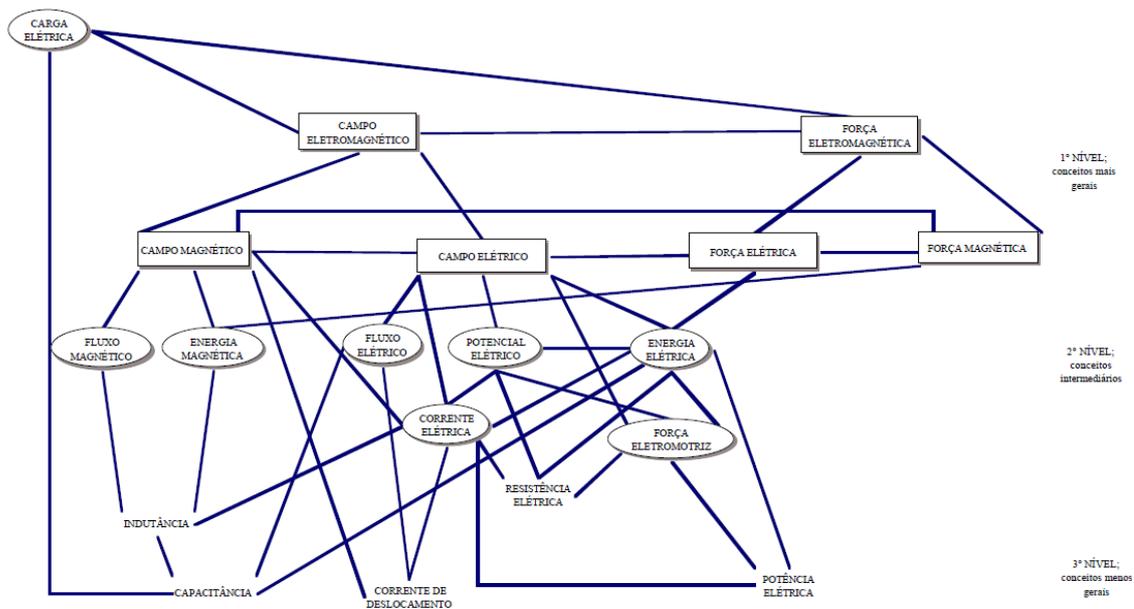


Figura 11. Mapeamento do conteúdo de Eletromagnetismo. Adaptado Moreira (2006)

Os exemplos que buscamos apresentar para contextualizar os usos dos mapas conceituais classificam-se em uma categoria maior de compreensão no processo de ensino-aprendizagem. Porém, não devemos limitar a abordagem apenas pelo conhecimento sobre o tema. Moreira (2006) ressalta que “Novak e Gowin (1984) apresentam vários exemplos de mapas conceituais construídos por crianças de escola primária”.

As relações significativas existentes entre conceitos que nem sempre são externadas em palavras são percebidas de formas distintas pelos níveis de instrução e idade que cada aluno tem ao se submeter ao mapa conceitual. Dessa forma, é possível estabelecer novas formas de conceitos e percepções através de um conjunto de novos significados que se encaixam em um novo sistema de referência proposicional.

### **2.3.4 Ferramenta escolhida**

Busca-se durante uma pesquisa científica apresentar soluções e/ou métodos que possam contribuir de forma significativa para a consolidação e engrandecimento do estado da arte. Dessa forma, optou-se então por escolher a ferramenta CMapTools, justamente por estar alinhada com o pensamento deste trabalho. Porém, faz-se necessário uma rápida introdução da mesma, visto que o propósito do trabalho é utilizar a ferramenta como um fim e não como solução ao problema.

#### *2.3.4.1 CMapTools*

O CMapTools<sup>1</sup> é uma ferramenta distribuída gratuitamente pelo Institute for Human and Machine Cognition – IHMC (CmapTools, 2013), que a disponibiliza em conjunto com outras ferramentas com o objetivo de proporcionar ambientes colaborativos e prover aos estudantes meios de colaborar em nível de conhecimento, permitindo que os usuários construam mapas conceituais e dividam o conhecimento expresso em seus mapas com outros estudantes. (Cabral et. al, 2003)

A ferramenta CmapTools, além de apresentar uma estratégia cognitiva para representação do conhecimento através dos mapas conceituais apresenta recursos para formatação dos

---

<sup>1</sup> <http://cmap.ihmc.us/>

mapas, ou seja, adicionar recursos ao mapas como: sons, imagens, vídeos, textos e até mesmo outros mapas para detalhar melhor os conceitos. (Cabral et. al, 2003)

A própria apresentação da proposta da ferramenta no seu *site*, além de ser auto-explicativa agrega uma série de conhecimentos que auxiliam os que buscam utilizá-la de forma a representar conhecimento. Apresenta ainda um extenso trabalho acerca de mapas conceituais e as formas disponíveis de construção do mesmo. A Figura 12 apresenta a *homepage* do projeto.

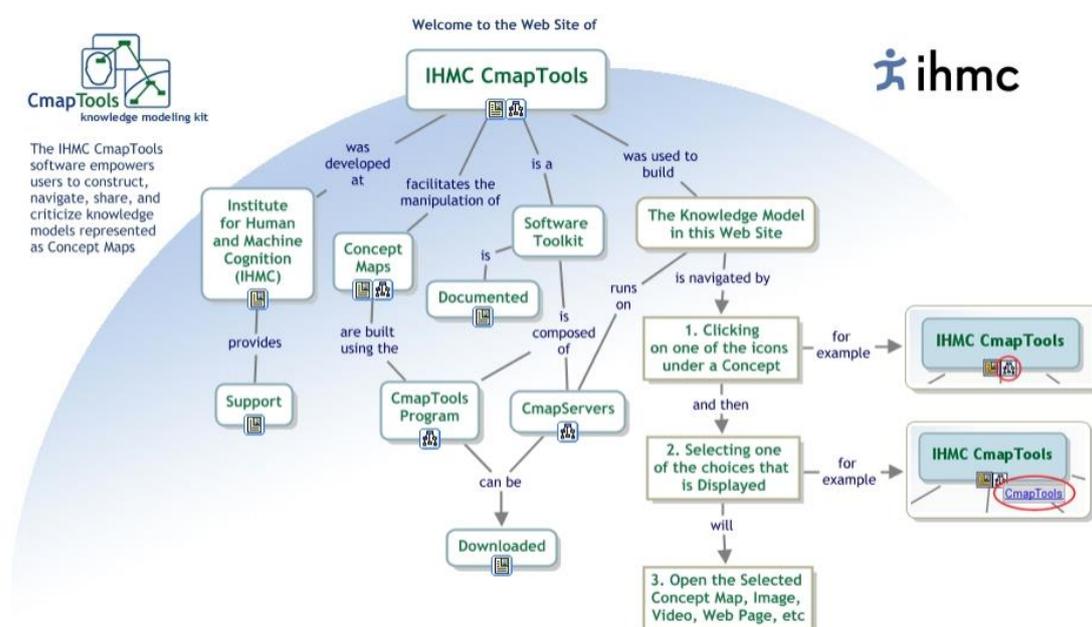


Figura 12. O site completamente desenvolvido como mapa conceitual. Uma nova abordagem no processo de construção e representação cognitiva.

### 2.3.5 Considerações Finais do Capítulo

Percebeu-se que em mineração de textos existem várias características que precisam ser levadas em consideração antes de continuar com o processo de extração automática, (Park et al, 2010). Dentre as problemáticas, um fator predominante ainda fica sem resposta. Como resolver questões de contextualização, grafia errada entre outras sem prejudicar o contexto que se pretende extrair? Tais dificuldades ainda estão sendo deixadas para decisões do usuário final, porém, com a análise devida, tais problemáticas poderiam ser embutidas em regras gramaticais que atendessem os requisitos ou sua maior parte.

Acerca das regras gramaticais, observou-se que uma frase aceitável pelos falantes, precisa passar por várias etapas de avaliação e construção. Pois, só serão bem formadas as frases que

passarem com sucesso por todos os testes, ou seja, as que forem construídas sem desobediência as regras fonológicas, sintáticas, morfológicas e semânticas. Porém, as frases aceitáveis ou não dependem também do conhecimento do linguista sobre a estrutura da língua, dessa forma, o domínio de aceitação de uma frase depende além da estrutura, do conhecimento do tema estudado.

Conclui-se sobre a revisão gramatical que este se caracteriza como sendo o principal problema para a obtenção de um processo de extração mais exato. Pois, o processo de mineração, apesar de complexo, requer algumas etapas básicas, e dentro dessas etapas precisam ser respondidas várias perguntas gramaticais para que a eficiência do método seja alcançada. Dessa forma, percebeu-se que não é a mineração em si o maior desafio e sim a compreensão da forma mais adequada da interpretação da gramática, uma vez que esta se apresenta complexa e com muitas regras que não podem ser generalizadas.

Acerca dos mapas conceituais, conclui-se que aqueles que possuem mais dimensões permitiriam uma representação ainda melhor das relações entre seus conceitos, além de possibilitar a inclusão de outros fatores que afetam a estrutura conceitual de um texto. Entretanto, as representações bidimensionais são mais simples e mais familiares. Portanto, mapas conceituais com mais de duas dimensões teriam utilidade limitada para auxílio na representação do conhecimento das estruturas conceituais estudadas.

### 3 O ESTADO DA ARTE E DEFINIÇÃO DA PROPOSTA

O vínculo entre os métodos e técnicas computacionais aplicados na construção automática e semi-automática 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 (Kowata et al, 2012).

Os métodos que utilizam fontes de dados não estruturadas guardam um estreito relacionamento com o campo do Processamento de Linguagem Natural (Stravrianou et al, 2007) 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 (Manzano-Macho et al, 2005).

A principal técnica utilizada na descoberta de conhecimento de textos é a mineração de textos. Desta forma faz-se necessário conhecer quais são os principais métodos empregados nesta técnica. As classes para os métodos de manipulação são qualificadas em classe de Métodos Linguísticos, classe de Métodos Estatísticos, classe de Métodos de Aprendizado de Máquina e, na impossibilidade de uma classificação única, há a classe dos Métodos Híbridos (Zouaq et al, 2009), (Zouaq et al, 2008).

A classe dos Métodos Linguísticos é caracterizada pelo uso de técnicas de linguística computacional (Manzano-Macho et al, 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 et al, 2009). Neste grupo estão

abordagens de Pérez et al (2004), Pérez et al (2005), Richardson et al (2007), Richardson et al (2008) e Kowata et al (2009)

Na classe dos Métodos Estatísticos, pode-se observar a utilização de técnicas baseadas em indicadores quantitativos (Manzano-Macho et al, 2005). De forma geral, tais técnicas produzem informações que permitem a análise da frequência de um termo e concorrências 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 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. (Kowata et al, 2012)

A perspectiva que pode se destacar nesta classe é a imprevisibilidade dos resultados e a inadequação na representação semântica das sentenças (Zouaq et al, 2009). Neste grupo estão abordagens de Dalmolin et al (2009a), Gaines et al (1994) e Clariana, et al (2004).

Na classe dos Métodos de Aprendizado de Máquina utilizam-se técnicas para a extração de elementos do texto. São técnicas que aplicam vários algoritmos para apoiar a descoberta de conceitos e suas relações (Manzano-Macho et al, 2005) e são, normalmente, utilizadas em conjunto com métodos estatísticos. Exemplos de métodos de aprendizagem de máquina podem ser encontrados nos algoritmos utilizados na descoberta de elementos concorrentes e que podem caracterizar regras de associação, na identificação de palavras-chaves e taxonomias entre elementos (Zouaq et al, 2009), (Kowata, 2012).

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), Valério et al (2006), Zouaq et al (2009), Zouaq et al (2008), Palmeira et al (2012), Kowata et al (2012) e Dalmolin et al (2009b).

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. (Kowata, 2012)

Baseado nessas técnicas foram identificadas 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 et al (2008), Kumazaua et al (2009), Araújo et al (2002), Bodgen (1977), Dalmolin et al (2009a), Dalmolin et al (2009b) Grilo et al (2008) e Moreira (1979) podem ser enquadradas na classe dos métodos de Mapeamento de Elementos porque elas tratam do mapeamento de conceitos, propriedades e associações existentes em ontologias de domínio para elementos de mapas conceituais.

As abordagens de Bai et al (2008), Chen et al (2008), Lee et al (2009), Tseng et al (2007), Jean-Marya et al (2009) e Moreira (2006) 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.

### **3.1 Os trabalhos relacionados com extração automatizada de textos para construção de mapas conceituais**

A contextualização da aplicabilidade dos mapas conceituais já foi delineada nos capítulos anteriores. Ressalta-se, portanto, que o uso dos mesmos requer vários fatores e vários métodos de implementação e que os objetivos são distintos como citado: como instrumentos didáticos; como instrumentos de avaliação; como recurso para análise de conteúdo.

É possível ressaltar a importância que os mapas vêm adquirindo no âmbito do processo ensino-aprendizagem para a representação do conhecimento, principalmente quando associados às novas tecnologias e os meios disponibilizados por elas (Tavares, 2007). Observa-se ainda neste âmbito que várias propostas tentam criar MCs de forma manual, semi-automática e automática, com uso ou não de técnicas inteligentes. Entretanto, no âmbito desta proposta apenas foram analisados os trabalhos que abordam métodos semi-automáticos e automáticos associados com técnicas de Inteligência Artificial.

Partir do cálculo da frequência dos elementos conectores das frases para a obtenção de tuplas de conceitos, foi a abordagem tratada por Dalmolin et al (2009a) na qual apresenta um método de extração que resulta em: i) uma lista de conceitos; ii) uma lista de conectores; iii) uma matriz de sobreposição. Sendo a matriz a representação do primeiro e do segundo item. Este faz uso também de elementos da mineração de textos, como: i) frequência; ii) parsing;

iii) reconhecimento de padrões; iv) análise semântica; v) tokenização; vi) heurísticas, etc. Entretanto, requer que o usuário informe, pela sua percepção, um valor de peso para cada entrada.

Em outra abordagem Dalmolin et al (2009b) apresenta um método de extração automático de textos a partir de fontes externas de entrada, como: imagem; texto; *link*. Nesse modelo cada palavra com espaço é considerada um termo. A partir dessa premissa um conceito pode ser formado com 1, 2 ou 3 termos. Todas as entradas recebem um peso atribuído pelo usuário. Ressalta o desenvolvimento de um algoritmo de extração de textos baseados no modelo híbrido.

Pode ser visto em Kowata et al (2009) uma abordagem das características para construção (semi) automática de mapas conceituais. Estas foram agrupadas em perspectivas relacionadas aos objetivos e à fonte de dados utilizada pelas pesquisas. Ressalta que extração léxica de textos requer dados externos, bases, etc. e que não são tão eficientes por conta disso. Por sua vez trata do foco de mapas conceituais a partir de textos direcionados, não genéricos. Com isso, há necessidade prévia de relacionar os assuntos.

Outra abordagem apresentada por Kowata et al (2012) apresenta um detalhamento melhor sobre o tratamento das particularidades inerentes à manipulação da linguagem natural. Além disso, busca aprimorar métodos e técnicas que restrinjam a perda semântica no mapeamento texto-mapa sem comprometer a essência do conteúdo e sem limitar a leitura humana. Apresenta seus resultados através de uma ferramenta Text2Cmap, o qual é composto por cinco módulos, cada um com regras específicas, que buscam realizar as atividades de mineração propostas.

Uma preocupação relativa à extração de textos é o Processamento de Linguagem Natural (PLN) e relacionada ao método de como as palavras são tratadas para uma extração automática. Uma abordagem interessante é vista no trabalho de Stravrianou et al (2007) que aborda a correção das palavras e a semântica. Cita as técnicas disponíveis de mineração de textos associadas a alguns modelos de base de dados, como o WordNet<sup>2</sup> e trabalha com as palavras através de um vetor de distância para encontrar similaridade.

---

<sup>2</sup> É uma base de dados eletrônica de elementos léxicos, a qual é disposta semanticamente e contém verbos, adjetivos e advérbios

Observa-se que extrações automáticas podem se basear na análise léxico-morfológica do PLN que é observado em Pérez et al (2005), cujo protótipo da ferramenta de extração de textos faz uso do parser PALAVRAS<sup>3</sup>, que realiza as etapas de tokenização, processamento léxico-morfológico e da análise sintática. Na extração do texto obtém tuplas na forma sujeito-verbo-objeto e atribui ao verbo a função de estabelecer a relação entre conceitos. Apesar de fazer uso de ferramentas prontas, obtém resultados satisfatórios.

### 3.2 As tentativas desenvolvidas

Os modelos até então desenvolvidos que envolvem a construção de mapas conceituais a partir de fontes distintas, envolvem uma escolha da fonte de dados utilizada, assim como definem uma variedade de requisitos, tais como: i) o público-alvo; ii) o nível de legibilidade; iii) a riqueza das proposições, entre outros. (Kowata et al, 2012).

Com base no que já foi explanado sobre a aplicabilidade dos mapas conceituais, é possível criar algumas grandes áreas de conhecimento para melhor representar essas associações. As áreas seriam: classe educacional; classe analítica e classe computacional.

No que tange à classe educacional Kowata et al (2012), afirma que seu principal objetivo é o uso de mapas conceituais como recursos complementares no processo de aprendizagem. Dos estudos que se enquadram nesta classe, é possível citar aqueles que:

1. Propõem construção de mapas conceituais de forma interativa (Alves et al, 2001) e (Clariana et al, 2004);
2. Propõem que professores acompanhem a compreensão de alunos em tópicos específicos apresentados em ambientes de aprendizagem tradicionais (Graudina et al, 2008), Araújo et al (2002), Bodgen (1977), Dalmolin et al (2009a), Dalmolin et al (2009b), Grilo et al (2008) e Moreira (1979) ou em ambientes virtuais de aprendizagem (Lau et al, 2008) e (Cimiano et al, 2005); e

---

<sup>3</sup> É parte de um grupo de parsers do projeto VISL (Visual Interactive Syntax Learning), do Institute of Language and Communication da University of Southern Denmark. O parser recebe como entrada o conjunto de sentenças de um corpus, e gera a análise sintática das sentenças.

3. Resumem e compartilham conteúdo digital (Richardson et al, 2007), Richardson et al (2008), Arlotta et al (2003), Dalmolin et al (2008), Gouli et al (2004), Rosson et al (2007) e Tavares (2007).

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 et al, 2008), (Chang et al, 2008), (Chen et al, 2008), (Lee et al, 2009), (Tseng et al, 2007), (Jean-Marya et al, 2009) e Moreira (2006)
- Exibem informações para os especialistas de domínio Gaines et al (1994), Kumazaua et al (2009), Pérez et al (2004), Pérez et al (2005), Kim et al (2007), Stewart et al (1979) e
- Propõem o uso de mapas conceituais como ferramentas para acesso, busca e análise de dados em bibliotecas digitais (Valério et al, 2006), (Michelson et al, 2007), Silva et al, 2009)

No contexto computacional, por meio de aplicações de software, existem pesquisas que fazem uso dos mapas conceituais para a aquisição de conhecimento. Nesta classe estão as pesquisas que utilizam os mapas conceituais como representações intermediárias entre o texto e as ontologias de domínio (Park et al, 2010), (Zouaq et al, 2009), (Jean-Marya et al, 2009), (Simper, 2009), (Zouaq et al, 2008), (Kim et al, 2007) e (Buitelaar et al, 2004).

### **3.3 Os problemas não resolvidos**

Percebe-se em Dalmolin et al (2009a) a tentativa de outra abordagem para a extração automática, porém a mesma apresentou alto custo computacional, requerendo alto processamento, assim como não se percebeu a saída automática pelo seu método extrator como proposto.

Já em Dalmolin et al (2009b) são necessárias várias análises do usuário para definir a qualidade do mapa conceitual. Percebeu-se que o método é relativamente complexo e não é claro, assim como seu trabalho anterior, além de requerer muito ajuste por parte do usuário.

Da mesma forma não foi possível identificar a saída do texto extraído de forma automática como proposto.

Em Kowata et al (2009) não houve a apresentação da ferramenta especificada no trabalho, tampouco foram apresentadas quais seriam as características centrais no desenvolvimento da proposta ou da técnica de extração de textos. Apresentou apenas uma intenção de fazer um mapa conceitual de forma semi-automática.

Em sua outra pesquisa Kowata et al (2012) já apresenta mais claramente quais são as etapas de pré-processamento escolhidas. Apesar de apresentar um trabalho mais elaborado, foram identificadas várias problemáticas em sua proposta. Inicialmente propõe o uso de regras de proximidade para identificar elementos candidatos para compor seu mapa, sejam os núcleos verbais ou os núcleos nominais, porém, não deixa claro como foi feita a composição destas regras.

Os *chunks* escolhidos pelo autor apresentam falhas na escolha de elementos textuais quando a fonte contém poucos sintagmas verbais ou nominais. Apresenta também a limitação quanto ao uso do texto de entrada, por perceber dificuldade na obtenção de certos elementos textuais. Além disso, apresenta em nosso entendimento, uma falha quando afirma não ter encontrado na literatura declaração explícita sobre a correspondência de elementos textuais e os conceitos e relações de um mapa conceitual.

O trabalho de Stravrianou et al (2007) ressalta que é importante a mineração de textos associar fontes distintas de entrada. Apresenta uma relação de ontologias e extração através de vetor de distância. Apesar de ter um forte embasamento, não apresentou um método próprio de extração, apenas uma profunda análise do que já se tentou fazer no estado da arte de mineração de textos e os métodos utilizados.

Apesar de mais embasado linguisticamente, o protótipo de extração automatizada de léxicos no trabalho de Pérez et al (2005), apresenta uma abordagem prática sobre a construção de regras gramaticais para extração de elementos verbais e nominais de um texto. Porém, restringe sua composição em apenas três elementos textuais, onde classifica como argumento1 e argumento2 ligados por elemento verbal que chamou de relação. Apesar de estarem claros os elementos textuais escolhidos, não foi possível reproduzir os mesmos resultados apresentados pelo autor.

### 3.4 A simplificação do modelo de extração e mineração

Percebeu-se durante a pesquisa que as propostas de extração semi ou automatizada apresentaram grande complexidade, sendo esse um dos principais fatores ressaltados pelos usuários destas propostas e ressaltados pelos seus autores. O modelo de extração para ser compatível com a assimilação de quem está construindo o mapa conceitual deve partir do princípio de que o usuário pode ter ou não domínio de um problema.

Com o objetivo de facilitar o processo de extração para o usuário e ao mesmo tempo identificar em um texto quais são os seus principais elementos verbais e nominais. Essa proposta, independente do grau de conhecimento do usuário sobre o texto que se quer analisar, busca atender as necessidades identificadas pelos usuários no capítulo anterior. Uma das versões iniciais do Text2MARK pode ser visto na Figura 13.

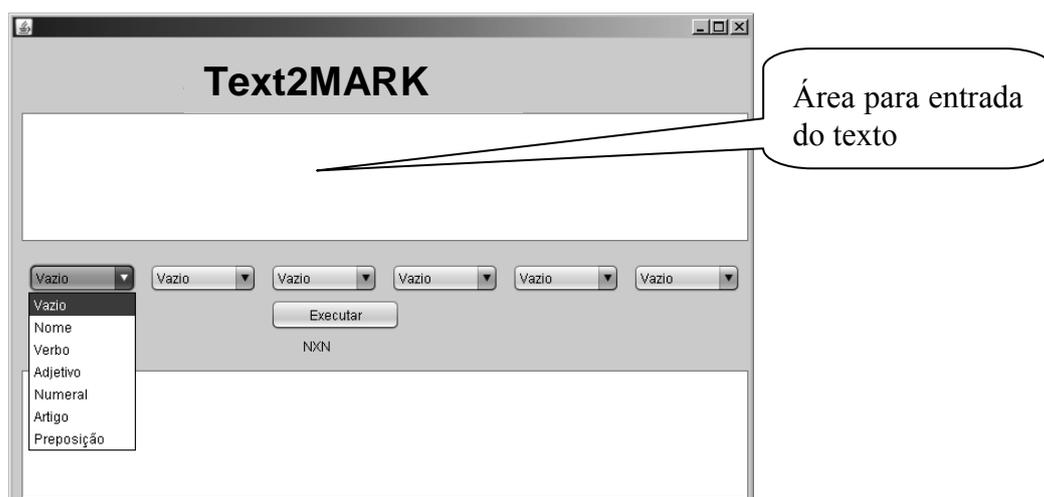


Figura 13. Na interface do Text2MARK há uma área para receber o texto de entrada

### 3.5 Breve introdução ao Text2MARK

A problemática central desta dissertação é apresentar uma ferramenta de mineração de textos para auxílio na representação do conhecimento. Desta forma, buscaram-se através de uma profunda análise gramatical identificar quais eram os elementos centrais e que poderiam representar um texto em poucas palavras.

A ferramenta Text2MARK é o meio pelo qual se buscou a solução dos problemas encontrados. Observou-se no estado da arte da mineração de textos que são requeridas várias

etapas, sendo a etiquetagem das palavras um ponto de partida central em algumas delas. Então, partindo do trabalho de Domingues (2011), que elevou o estado da arte da acurácia dos etiquetadores para acima de 98%, associamos mais etapas de mineração em conjunto com regras linguísticas para determinar uma proposta válida, plausível e diferente do que se tem feito até então.

### 3.6 Considerações finais do Capítulo

Como apresentado, as propostas atuais que têm relação direta com o que se propõe neste trabalho não buscam de forma sistemática realizar as principais etapas tanto de mineração de textos quanto da análise gramatical com a finalidade de compor mapas conceituais para representar conhecimento. Em sua maioria não mostram os métodos de como obter suas saídas após o processamento ou possuem falhas na interpretação deixando o texto sem a devida compreensão ou não é possível reproduzir seus resultados.

É necessário observar algumas características gerais das propostas. A Tabela 2 faz uma comparação com os métodos da última década e apresenta as melhorias e principais características buscadas no estado da arte. Para facilitar a apresentação os autores foram colocados na Tabela 3.

Tabela 2. Os autores e suas contribuições para o processo de extração de textos

Característica	Text2MARK	01	02	03	04	05	06	07	08	09	10	11
Constrói mapas conceituais a partir de fontes externas	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
Aborda classe educacional	✓	✓	✓	✓	✗	✗	✓	✗	✓	✓	✗	✗
Permite construir mapa de forma interativa	✓	✓	✓	✓	✗	✗	✓	✗	✗	✓	✗	✓
Permite acompanhar em ambiente de aprendizagem tradicional	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗	✓	✗
Resume e compartilha conteúdo digital	✓	✗	✗	✗	✗	✗	✓	✓	✓	✗	✗	✗
Baixo custo computacional	✓	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
Processamento totalmente local	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Facilita o uso para o usuário	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
Ferramenta computacional	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
Usa regras gramaticais	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗
Não limita o texto de entrada	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
Tem método próprio de extração	✓	✓	✓	✓	✗	✓	✓	✓	✗	✗	✗	✗

Tabela 3. As descrições das abreviaturas apresentadas na Tabela 2

#		#	Autor
01	Kowata 2012	02	Dalmolin 2009a
03	Dalmolin 2009b	04	Kowata 2009.
05	Graudina 2008	06	Dalmolin 2008.
07	Richardson 2008	08	Stravrianou 2007
09	Clariana 2004	10	Araújo 2002
11	Alves 2001		

Neste capítulo, foram analisadas algumas perspectivas dos principais autores sobre extração de textos de forma semi ou automática. Apesar das pesquisas sólidas acerca do assunto, a pergunta central que nos propomos no início desta dissertação ainda não foi respondida de forma adequada. A extração de forma automatizada dos conceitos mais relevantes de um texto em forma de tuplas para construção de mapas conceituais e que este processo seja feita de forma mais clara, simples e eficiente para o usuário.

No próximo Capítulo será apresentada a ferramenta Text2MARK que usa em seu método: execução das etapas necessárias na mineração de textos; mantém a aplicação local e com baixo desempenho computacional; baixa complexidade para o usuário; apresenta as regras gramaticais para extração dos léxicos verbais e nominais e mostra como se obtém as saídas do processamento.

## 4 A FERRAMENTA DE EXTRAÇÃO DE TEXTOS TEXT2MARK

Após analisar o estado da arte e a gramática descritiva, percebeu-se certa complexidade para obtenção dos elementos léxicos necessários que permitam de forma clara sua extração para construir as sentenças que possibilitem sua representação em um mapa conceitual. Dessa forma, foi necessário desenvolver uma ferramenta em Java que tivesse todos seus recursos trabalhando localmente, com baixo custo computacional, e que deixasse embarcada a complexidade do método de extração. A Figura 14 apresenta a interface inicial do Text2MARK, ainda chamado de Etiketador.

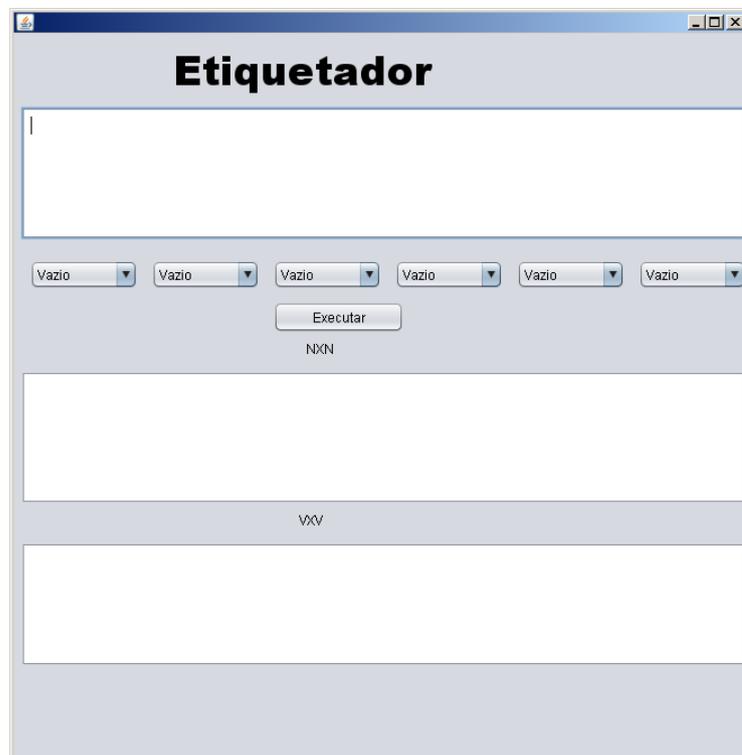


Figura 14. Inicialmente chamado de etiquetador, gerava apenas os formatos NVN e VNV

## 4.1 Objetivo da ferramenta

Esta proposta se baseia no método híbrido de extração automatizada, dessa forma, a ferramenta auxilia na simplificação de várias etapas requeridas no processo de extração. Entre os objetivos principais da ferramenta podemos destacar os seguintes:

- i. Realizar processos automatizados de etiquetagem de palavras utilizando o etiquetador TreeTagger (Schmid,1995) associado com as melhorias apresentados em Domingues (2011);
- ii. Extrair os elementos gramaticais necessários para representar em 100% um diagrama entidade-relacionamento de Heuser;
- iii. Extrair a raiz etimológica das palavras do texto e obter o cálculo de sua frequência;
- iv. Identificar e fazer o cálculo da frequência dos elementos léxicos mais relevantes em um texto;
- v. Criar regras de predicados que possam representar sentenças em textos;
- vi. Extrair a soma dessas tuplas com base no cálculo da frequência das suas palavras;
- vii. Obter sentenças de categorias lexicais no formato de tuplas para a construção dos mapas conceituais;

## 4.2 Definição do escopo

Inicialmente, o Text2MARK tratava apenas a extração de tuplas no formato NOME-VERBO-NOME (NVN) e VERBO-NOME-VERBO (VNV), com o intuito de estabelecer um parâmetro para avaliar e avançar na pesquisa. O experimento tomou como base o texto extraído do livro Projeto de Banco de Dados de Carlos Alberto Heuser (Heuser, 2001, pg 69) Exercício 3.6: Estudo de caso – locadora de vídeos.

Como uma avaliação primária do trabalho em andamento, buscou-se analisar o mapa conceitual gerado pelo CMapTools a partir da saída do Text2MARK, segundo a completude e corretude dos requisitos do estudo de caso citado em Heuser (2001, p.69). Ressalta Sayão et al (2003) que a completude verifica se estão presentes todos os requisitos necessários à

especificação do sistema e Silva (2011) entende a completude como a quantidade de requisitos corretamente identificados.

Já a corretude em Sayão et al (2003) ressalta que há corretude se os requisitos descreverem as funcionalidades de maneira correta e Silva (2011) diz que a corretude é a quantidade de requisitos corretamente descritos.

A Tabela 4 mostra os requisitos observados em Heuser (2001) e as respectivas saídas obtidos pelo Text2MARK. Inicialmente o Text2MARK tratava apenas NOMES e VERBOS, com isso os adjetivos e determinantes presentes no requisito 2 tiram o sentido da frase extraída. Entretanto, quando se analisa os textos extraídos sob o princípio da completude, em relação aos dez requisitos elencados, as saídas geradas pelo Text2MARK mostram-se aptas e com um percentual de acerto promissor, acima de 90%, pois apenas um dos 10 requisitos, o requisito 2, não foi capturado adequadamente.

Tabela 4. Os requisitos e as saídas geradas automaticamente

#	Requisito do texto	Saída gerada pelo Text2MARK
1	“..Cada fita possui um número..”	Fita – possui – número
2	“.. cada filme recebe um identificador próprio. Para cada fita..seu título e sua categoria”	Filme – recebe – fita
3	“.. Para cada filme há pelo menos uma fita..”	Filme – há – fita
4	“.. e cada fita contém somente um filme.”	Fita – contém – filme
5	“... filmes estralados pelo seu ator predileto..”	Filmes – estrelado – ator
6	“..informação dos atores que estrelam em cada filme..”	Atores – estrelam – filmes
7	“..clientes cadastrados podem alugar fitas.”	Clientes – alugar – fitas
8	“.. locadora possui muitos clientes cadastrados..”	Locadora – possui – clientes
9	“.. cada cliente recebe um número de associado..”	Cliente – recebe – número
10	“.. um cliente pode ter várias fitas..”	Clientes – ter – fitas

Observaram-se limitações da cardinalidade (determinante) das relações entre os elementos léxicos. Os requisitos 3 e 4 apresentam apenas a relação existente entre filme e fita, porém não consegue expressar a obrigatoriedade (cardinalidade mínima igual a um) de que “para cada filme há pelo menos uma fita” ou expressar cardinalidade máxima, “cada fita contém somente um filme”, nesse caso a cardinalidade máxima é igual a um. Essa limitação implica em deficiência da corretude dos requisitos. As Figuras 15 e 16 representam o diagrama de Heuser e o mapa gerado respectivamente.

De acordo com a resolução feita por Heuser as entidades do problema seriam: i) locadora; ii)filme; iii)fitas; iv)cliente; v)categoria e vi)ator. Essas foram analisadas pelo Text2MARK como sendo o elemento léxico NOME. Em seguida os relacionamentos identificados são: i) entre FILME e FITA; ii) EMPRÉSTIMO(ALUGAR) entre FITA e CLIENTE; iii) entre

FILME e CATEGORIA e iv) ESTRELA entre ATOR e FILME. O extrator identificou com sucesso as entidades e as relações, com apenas um erro, pois não identificou a entidade CATEGORIA (requisito 2) e a relação existente com este termo, reforçando a validação do método e sua acurácia.

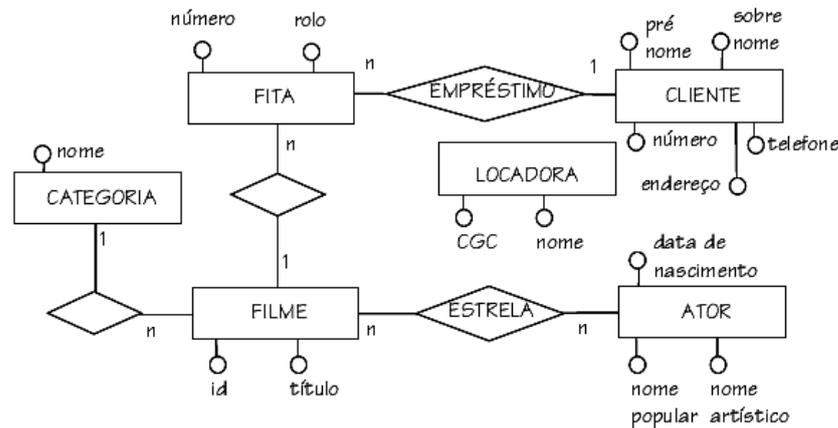


Figura 15. O diagrama com os requisitos das relações. Adaptado Heuser (2001)

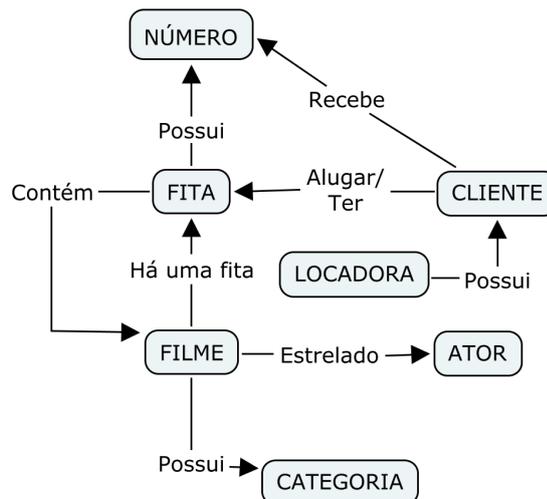


Figura 16. O mapa conceitual gerado no CMapTools com as saídas do Text2MARK.

Apesar do grau de acerto dos elementos extraídos perceberam-se várias limitações na ferramenta para alcançar a solução proposta. Ressalta-se que textos mais complexos não seriam analisados de forma correta, pois não eram tratados outros elementos léxicos. Assim como, a tupla VERBO-NOME-VERBO (VNV), apresentou um número de saídas muito limitado, abaixo de 10% do total das saídas NOME-VERBO-NOME (NVN), dessa forma, e com base na gramática descritiva decidiu-se remover as tuplas neste formato VNV. No total foram desenvolvidas trinta e oito versões do Text2MARK para chegar ao resultado esperado. A Figura 17 mostra as versões 2 (a), 3 (b) e 4 (c) do Text2MARK.

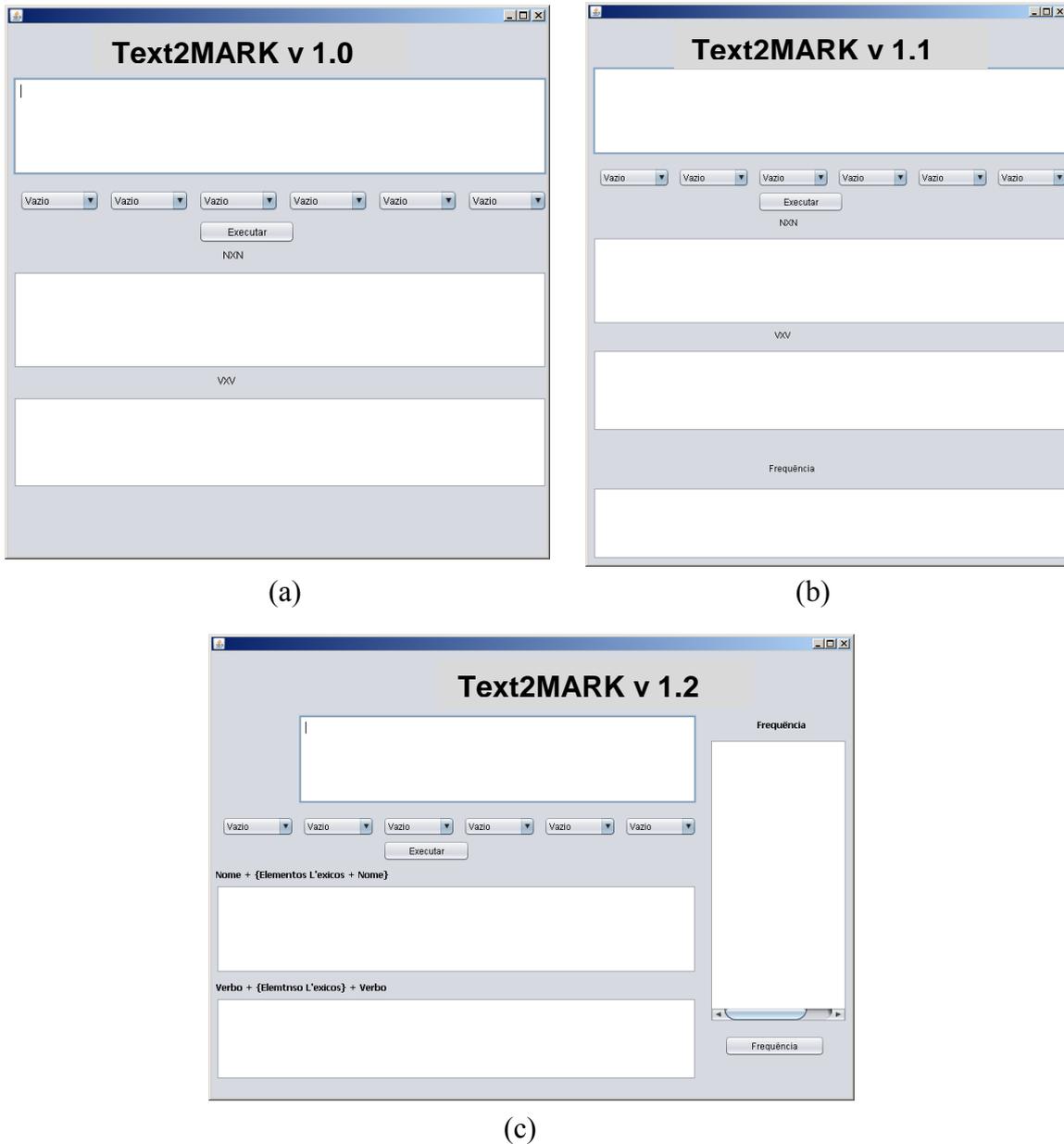


Figura 17. (a) Já com o nome de Text2MARK e com listas suspensas para compor predicados. Em (b) já é possível saber o cálculo da frequência das tuplas NVN extraídas. Em (c) é agregado na interface à extração da raiz etimológica das palavras, ainda sem calcular sua frequência.

### 4.3 Projeto técnico

Buscou-se nessa proposta não modificar o método de entrada do texto livre, com isso nossa definição do processo inicia com a seleção do texto léxico em português brasileiro que contém a definição na qual se deseja obter as tuplas. Na etiquetagem, foram empregados alguns programas desenvolvidos em Java, que a partir da versão 1.19 estão embarcados na

interface. Todas as etapas rodam localmente com baixo custo computacional e tempo de resposta.

A primeira parte do processo de entrada trata da normalização e segmentação do texto ou tokenização onde são feitos alguns pré-processamentos para separação das preposições, como por exemplo: i)da em “de a”; ii)das em “de as”; iii) do em “de o”; iv)dos em “de os”. Nas formas maiúsculas e minúsculas, entre outras. Da mesma forma, acentuações na forma de crase e pontuações também são tratadas, assim como os pronomes, espaços, etc. Feito isso, o texto é quebrado em linhas para cada elemento identificado.

O processo seguinte trata da etiquetagem executada com o TreeTagger e o trabalho de Domingues (2011), já citado. Faz-se necessário para poder etiquetar o texto previamente preparado no formato de uma palavra por linha.

As etapas seguintes são a análise e o refinamento lexical. A análise é feita a partir de comparações e cálculo da acurácia que compara os pares de palavras-etiquetas obtidos com os pares previamente etiquetados correspondentes e identifica os pares iguais (acertos) e os pares diferentes (erros). A acurácia é calculada pela soma de todos os acertos dividida pelo total de pares palavras-etiquetas submetido à etiquetagem. Já o refinamento trata da melhoria obtida na etiquetagem através do trabalho Domingues (2011), que ajusta suas inconsistências de etiquetas aumentando sua acurácia em valores acima de 98% % em textos jornalísticos e científicos.

A problemática da importância dos elementos léxicos no texto é tratada por meio do reconhecimento de elementos centrais, os quais foram obtidos através do cálculo de suas frequências presentes no texto. Com isso, ao identificar os Sintagmas Nominiais (SN), Sintagmas Verbais (SV) e os Sintagmas Preposicionais (SP) e da forma como estavam dispostos, partiu-se para entender como esses elementos se relacionavam.

Pelo cálculo da frequência, também foram identificados dezessete elementos léxicos mais relevantes, que quando associados da forma adequada compõe a tupla NOME-VERBO-NOME de modo a satisfazer a obtenção do conceito requerido. Tais elementos léxicos foram associados de modo a formar várias regras gramaticais, as quais foram chamados de predicados, levaram-se em consideração as regras do processamento da linguagem natural e as nuances da Língua Portuguesa. Desta forma, os predicados possibilitaram de maneira mais objetiva e explicativa a extração dos textos na forma de sentenças que podem ser usados na construção de mapas conceituais.

Entre as otimizações realizadas, é possível citar a remoção das etapas de pré-processamento 1 e 2 modo manual por meio do prompt de comando. Porém, buscou-se não prejudicar o desempenho da ferramenta em termos computacionais, aumentando seu custo, com isso foi possível manter o mesmo desempenho com o ganho visual da interface e dos recursos adicionais propostos nesta pesquisa. Ressalta-se ainda que a extração foi refinada e ruídos como duplicidade dos substantivos presentes no início e final da tupla foram removidos.

A identificação de outras categorias gramaticais e suas etiquetas usadas pelo Text2MARK pode ser vista como outro grande avanço, dessa forma foi possível montar novas tuplas compondo novos predicados que resultam em resultados mais precisos. Essas etiquetas fazem parte do conjunto de etiquetas do Corpus Mac-Morpho que é descrito por Aluísio et al. (2003) e contém 78 etiquetas: 22 regulares, 18 sinais de pontuação e 38 regulares combinadas com complementares. Os elementos gramaticais com suas abreviações ou classes morfológicas podem ser vistos na Tabela 5.

Tabela 5. Léxicos e as siglas atribuídas pelo etiquetador. Fonte: TreeTagger

#	Categoria lexical	Classe Morfológica
1	Adjetivo	ADJ
2	Advérbio	ADV
3	Artigo	ART
4	Conjunção Coordenativa	KC
5	Nome	N
6	Nome Próprio	NPROP
7	Numeral	NUM
8	Preposição	PREP
9	Preposição com Contração	PREP +
10	Pronome Adjetivo	PROADJ
11	Pronome Denotativo	PDEN
12	Pronome Pessoal	PROPESS
13	Pronome Relativo	PRO-KS-REL
14	Pronome Substantivo	PROSUB
15	Verbo	V
16	Verbo Auxiliar	VAUX
17	Verbo No Particípio Passado	PCP

A cada etapa de desenvolvimento outros recursos foram sendo adicionados na ferramenta para que o processo de extração e por consequência a construção do mapa conceitual fosse a mais abrangente possível. Entre estes recursos podemos destacar: i) a extração das raízes das palavras para determinar o cálculo de sua frequência com maior precisão; ii) a soma das frequências das tuplas; iii) a média da soma das frequências das tuplas; iv) a frequência dos

léxicos; v) a inclusão dos predicados. Com esses elementos associados com os já citados, a construção do mapa conceitual torna-se mais simples.

Para construir o mapa conceitual, tomou-se como base a sua forma clássica de representação, podendo ser expresso da seguinte forma. O conteúdo do mapa (M) é constituído por um conjunto de sentenças  $st_1 \dots st_n$ , ou seja,  $M = \{st_1, st_2, st_3, \dots, st_n\}$ . Para tanto, se faz necessário identificar um conjunto de proposições  $pr_1 \dots pr_n$  que devem ser extraídos. A cada proposição diz-se que é composta por uma tupla de elementos no formato NOME-VERBO-NOME  $\{n_1 - v_1 - n_2\}$ . As quais serão as saídas mais relevantes de acordo com o texto especificado.

#### 4.4 Arquitetura

A Arquitetura do Text2MARK, Figura 18, parte inicialmente do etiquetador desenvolvido em Domingues (2011), porém, ressalta-se que no referido trabalho, as etapas de 1 e 2 são executadas no prompt de comando e as Etapa 3 e 4 são executadas por meio de uma aplicação em Java, com suas regras embarcadas.

Após o processo de etiquetagem e o aumento de acurácia acontecerem nas etapas de 1 a 4, o Text2MARK realiza as demais funções de mineração de textos associadas a regras de predicados embasados na gramática descritiva. O resultado final do processamento do texto determina uma tupla a qual tem em sua identificação além do léxico que a compõem a soma de sua tupla.

São especificados alguns arquivos temporários que guardam informações previamente tratadas que ainda precisam de refinamento. Percebeu-se ainda que antes do resultado final a extração das tuplas precisava de refinamentos. Isso se dá pela complexidade que a gramática da língua portuguesa requer; algumas das tuplas importantes para a construção do mapa estavam sendo excluída do resultado final, justamente por não ocorrerem várias vezes no texto, logo sua frequência é menor, porém, seu significado é relevante para a compreensão textual da entrada.

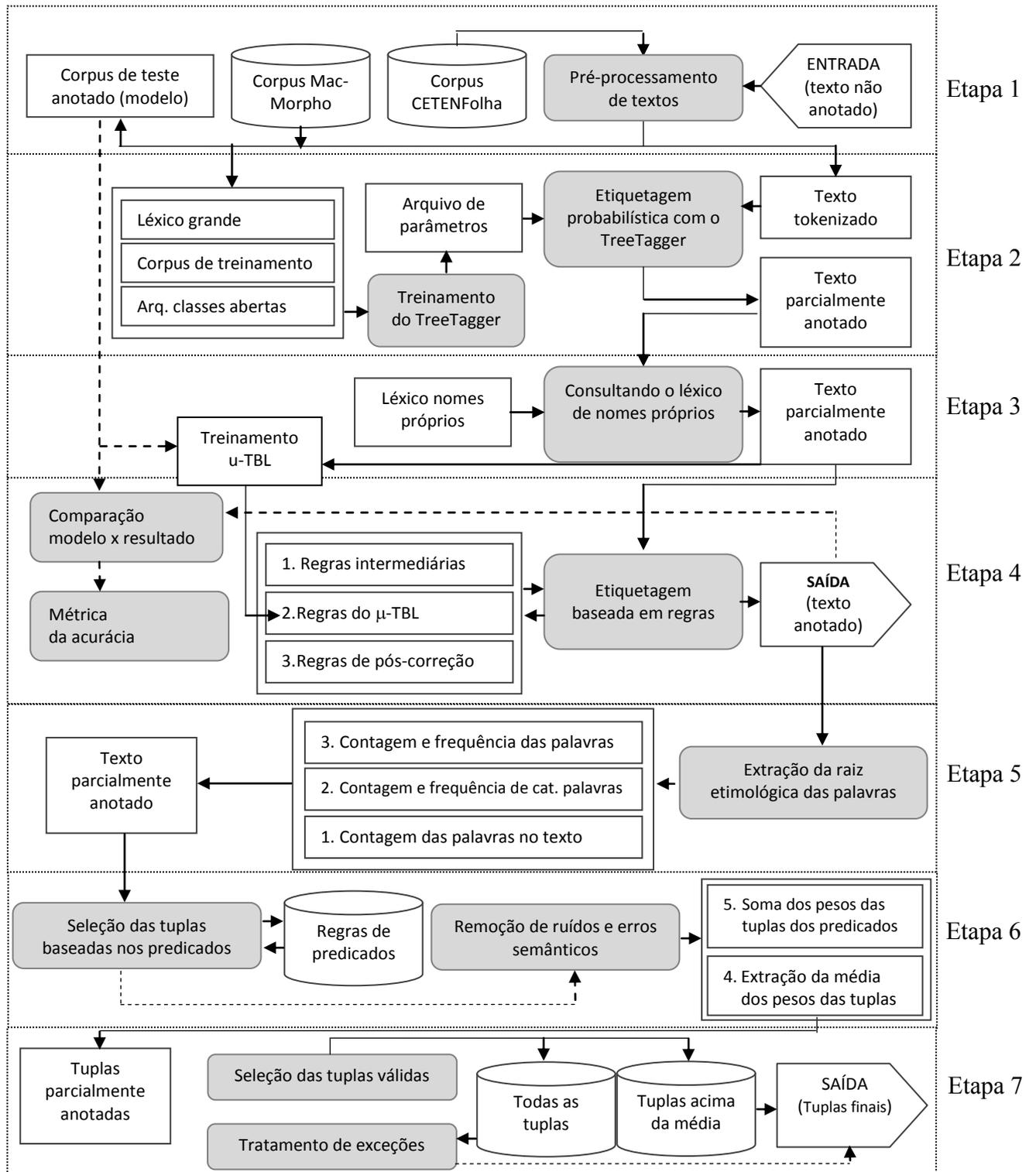


Figura 18. Arquitetura funcional do Text2MARK. As Etapas de 1 a 4 provenientes de Domingues (2011) as quais foram embarcadas e as etapas de 5 a 7 que descrevem o Text2MARK

## 4.5 As regras gramaticais

A composição dos predicados obedeceu a várias regras gramaticais que são necessárias para a compreensão do texto em língua portuguesa, conforme a Tabela 6. Essas regras foram estabelecidas levando-se em consideração a construção dos Sintagmas Nominais (SN), Sintagmas Verbais (SV) e os Sintagmas Preposicionais (SP). Conforme foi visto no capítulo 2 item 2.2, a composição dos sintagmas requer várias regras de posicionamento das categorias gramaticais antes e depois do elemento de ligação das sentenças, geralmente o VERBO.

Inicialmente, a composição dos predicados levou um grande tempo, justamente por que sua composição partiu do elemento básico NOME-VERBO-NOME. Essa estrutura, apesar de atender a vários requisitos gramaticais, deixa de fora outros elementos, logo, prejudicando a compreensão do texto. Com o devido auxílio da gramática, o elemento NOME foi desmembrado em vários outros que também podem ser definidos como NOME, mas que recebem outra identificação do etiquetador de acordo com seu SN.

Após desmembrar o código do formato NVN, com o auxílio do Text2MARK, outros elementos léxicos foram analisados e estudados, por essa razão é que as versões iniciais da interface apresentam seis listas suspensas para escolha de um dos dezessete elementos léxicos apresentados no item anterior. Ressalta-se que existe um número maior de elementos léxicos, porém, no processo de desconstruir e reconstruir as subsentenças de um texto, os dezessete elementos escolhidos foram os que apresentaram maior relevância gramatical dentro dos sintagmas.

Durante o processo de extração todos os predicados são aplicados no texto analisado, porém, de acordo com a sua composição nem todos os predicados são encontrados nos sintagmas, dessa forma, as regras propostas estabelecem generalidades textuais de SN, SV e SP, pois são extraídas tuplas na forma proposta. As discussões sobre a validade e avaliação dos predicados serão discutidas e avaliadas no Capítulo 5.

Tabela 6. Regras de predicados com base nos SN, SV e SP.

#	Regra	#	Regra	#	Regra
1	N-ADJ-V-N-KC-N	9	N-PROADJ-N-VAUX-PCP-N	17	N-VAUX-V-N
2	N-ADV-V-N	10	N-PROADJ-N-VAUX-PCP-V-N	18	N-VAUX-V-PROADJ-N
3	N-ADV-V-V-N	11	N-V-ADJ-V-PROADJ-N	19	N-VAUX-V-V-N
4	N-N-N	12	N-V-ART-ADJ-ADJ-N	20	N-V-N
5	N-PCP-VAUX-V-N	13	N-V-ART-ADJ-N	21	N-KC-PROADJ-N
6	N-PREP-N	14	N-V-ART-N	22	PROADJ-N-V-ART-ADJ-ADJ
7	N-PREP-NUM-N	15	N-V-ART-N-PREP-N		
8	N-PREP-N-V-NUM-N	16	N-VAUX-VAUX-PCP-N		

## 4.6 Metodologia de extração

Em vez de representar textualmente como se dá o método de extração da ferramenta, apresentamos na Figura 19 um mapa conceitual de como se comporta o Text2MARK, desde a entrada do texto livre até seu resultado que geram tuplas no formato NVN para construção de mapas conceituais.

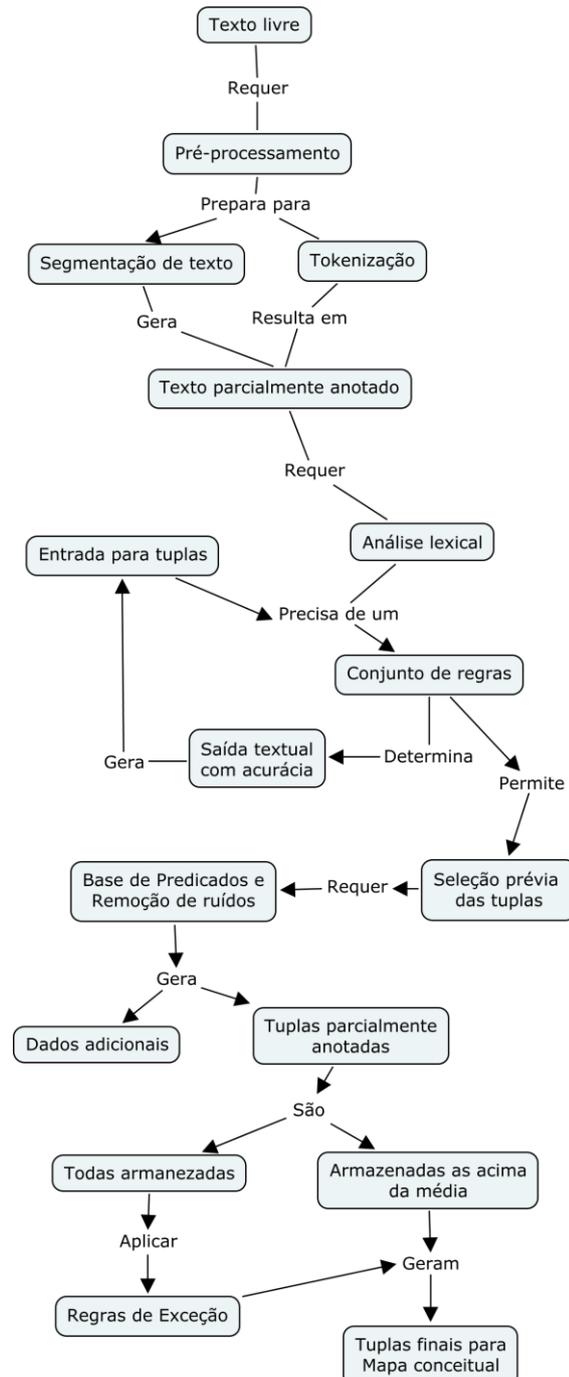


Figura 19. Metodologia de extração do Text2MARK

## 4.7 A escolha do escopo

Em outro momento da pesquisa, mas continuando com o mesmo parâmetro inicial, de avaliar a estrutura apresentada por Heuser, foram feitas novas extrações com a ferramenta, porém com a correção das limitações previamente identificadas no Text2MARK.

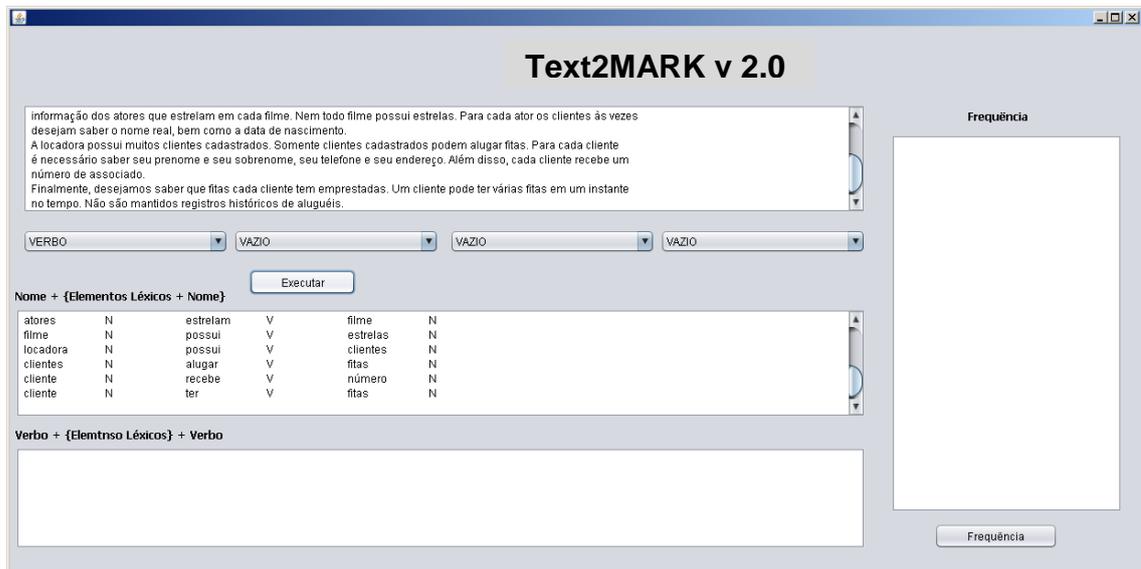
Após identificar previamente os requisitos textuais do estudo de caso, foi possível observar quais são as categorias de palavras que os compõem. Dessa forma foi possível utilizar os predicados estabelecidos nas regras gramaticais e realizar a extração das tuplas, guardando seus resultados para construir o mapa conceitual no CMapTools. Os predicados extraídos foram: i)NVAN; ii)NVN; iii)N-KC-PROADJ-N; iv)N-VAUX-V-N. Para melhorar a visualização, apresentamos na Tabela 7 os requisitos do texto, algumas das regras propostas e a respectiva saída gerada pelo Text2MARK.

Tabela 7. Os requisitos identificados no texto com as regras e saídas

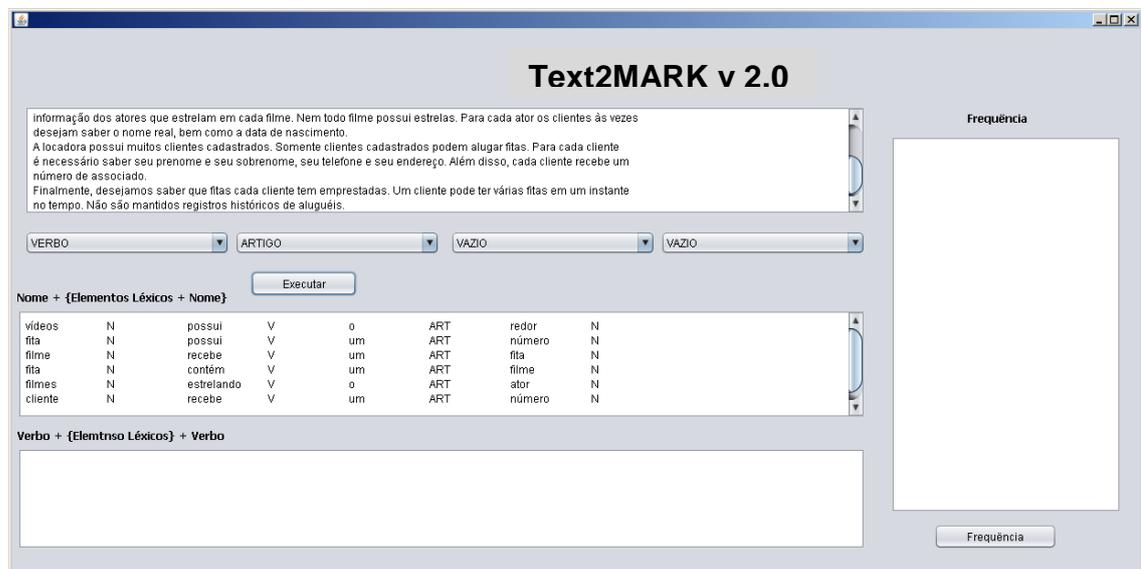
#	Requisito do texto	Regra	Saídas do Text2MARK
1	“...de vídeos possui ao redor de 2000 fitas...”	N-V-A-N	Vídeos – possui – o – fitas
2	“...Cada fita possui um número...”	N-V-A-N	Fita – possui – um – número
3	“... filme, é necessário saber seu título e sua categoria...”	N-KC-PROADJ-N	Título – e – sua – categoria
4	“...cada filme recebe um identificador próprio...”	N-V-A-N	Filme – recebe – um – identificador
5	“.. Para cada filme há pelo menos uma fita...”	N-V-A-N	Filme – há – uma – fita
6	“...e cada fita contém somente um filme...”	N-V-A-N	Fita – contém – um – filme
7	“... encontrar os filmes estrelados pelo seu ator predileto...”	N-V-A-N	Filmes – estrelados – o – ator
8	“...informação dos atores que estrelam em cada filme...”	N-V-N	Atores – estrelam – filmes
9	“... locadora possui muitos clientes cadastrados...”	N-V-N	Locadora – possui – clientes
10	“...clientes cadastrados podem alugar fitas...”	N-VAUX-V-N	Clientes – podem - alugar – fitas
11	“...saber seu prenome e seu sobrenome...”	N-KC-PROADJ-N	Prenome – e – seu – sobrenome
12	“...seu telefone e seu endereço...”	N-KC-PROADJ-N	Telefone – e – seu – endereço
13	“... cada cliente recebe um número de associado...”	N-V-A-N	Cliente – recebe – um – número
14	“... um cliente pode ter várias fitas...”	N-VAUX-V-N	Clientes – pode – ter – fitas

Ainda foram observadas limitações da especificação de contexto das relações entre os elementos léxicos. Como exemplos, os requisitos 1 e 7, pois com o processo de mineração há separação do sintagma nominal. No requisito 1 a palavra AO se desmembra em A (Preposição com contração) + O (Artigo) e no requisito 7 a palavra PELO se desmembra em POR

(Preposição com contração) + O (Artigo). Entretanto, pelo ponto de vista gramatical esse problema não é considerado como falha, uma vez que trata na essência a composição do elemento. A Figura 20 mostra duas etapas da extração.



(a)



(b)

Figura 20. (a) Extraindo a forma tradicional NVN e em (b) extraindo o determinante ARTIGO

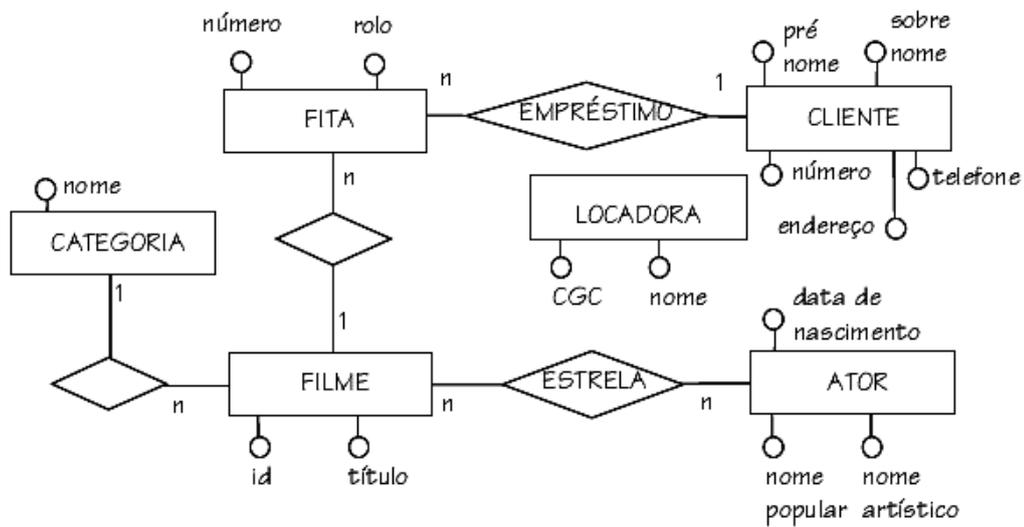


Figura 21. O diagrama com os requisitos das relações. Adaptado Heuser (2001)

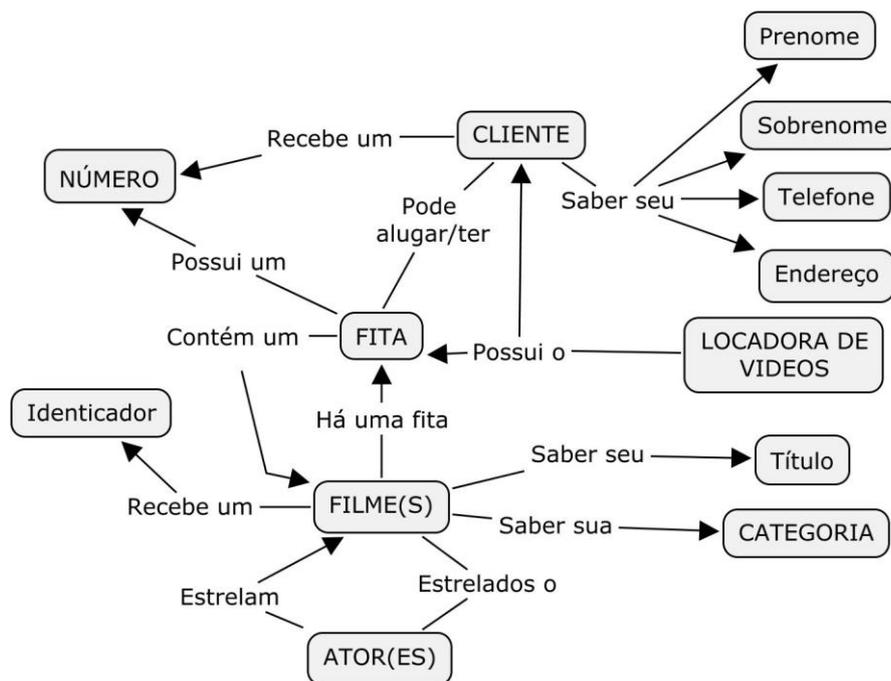


Figura 22. O mapa mais completo gerado no CMapTools com as saídas do Text2MARK.

Diferente do escopo inicial da pesquisa, com a versão contendo mais elementos léxicos e com a composição dos predicados, foi possível construir um mapa conceitual mais completo. Analisando pelo ponto de vista da completude e corretude, o resultado alcançado nesta etapa também satisfaz estas duas condições. Dessa forma, nosso estudo passa para um segundo momento, em que outros elementos textuais serão analisados e extraídos seus principais

léxicos gramaticais para construir mapas conceituais que representem seu contexto, como se vê na comparação entre as Figuras 21 e 22.

#### **4.8 Considerações finais do Capítulo**

Ao comparar o digrama apresentado por Heuser (2001) e posteriormente as duas versões do mapa conceitual gerado com as saídas do Text2MARK, percebeu-se a evolução da proposta e o quanto ela pôde representar o referido diagrama. A primeira versão do mapa apresentou falhas ao deixar de identificar relações existentes, assim como não foi capaz de ressaltar quais seriam os possíveis atributos das entidades. Mesmo assim, a segunda versão do mapa em comparação com o modelo tradicional de Heuser, teve uma taxa de acerto de 100%, justamente por identificar todas as entidades e as relações existentes entre elas.

A escolha e utilização do método híbrido para a construção da interface se mostraram válidas para a extração em textos previamente preparados, como os de Heuser. Sendo este nosso ponto de partida e tendo alcançado todos os seis objetivos inicialmente elencados para a ferramenta, o trabalho segue rumo à extração de textos livres para a construção de mapas conceituais de forma simples e automatizada.

Apesar do processo de extração automatizada requerer vários detalhes, os quais foram compreendidos e incluídos neste trabalho durante a pesquisa, percebeu-se que apenas as regras gramaticais não eram suficientes para representar o conhecimento de forma ampla com precisão. Foram utilizados os princípios da completude e da corretude para determinar o quanto a pesquisa estaria madura o suficiente para avançar e avaliar textos diferentes, que teriam estruturas gramaticais mais complexas e novos elementos semânticos.

## **5 AVALIAÇÃO QUALITATIVA**

Após ter submetido o Text2MARK nos textos de Banco de Dados de Heuser (2001) e ter obtido 100% de representação analisado pela completude e corretude, seria necessário submeter o Text2MARK a outros tipos de textos, com formas de escrita e elaboração diferentes. Dessa forma, para submeter à ferramenta a uma avaliação imparcial foi necessário buscar outras fontes de textos. Neste contexto, com o auxílio da Profa. Dra. Marianne Kogut Eliasquevici da Faculdade de Computação (FACOMP) da Universidade Federal do Pará a qual nos forneceu artigos científicos de ciências distintas para avaliação do Text2MARK. Em nenhum momento houve de nossa parte interferência ou preferência na escolha dos mesmos. As áreas científicas dos artigos são: Geografia; EAD; Linguística; Meio ambiente e Pedagogia.

### **5.1 A extração dos textos técnicos**

Nesta etapa os textos fornecidos serão submetidos à extração automatizada pelo Text2MARK. Todos os ruídos de pontuação e sinalização identificados na obtenção da frequência dos léxicos, que se dá com base no texto, foram removidos. Com isso o resultado que se espera nesta etapa é obter as tuplas e selecionar o escopo de acordo com a perspectiva do usuário.

Percebeu-se durante a extração que a quantidade de tuplas extraídas é em média de 23% a 30% do total das palavras do texto. Logo, é possível afirmar que textos com mais de mil palavras vão gerar em torno de duzentas e trinta a trezentas tuplas, o que vai tornar a construção do mapa conceitual uma tarefa complicada até para o especialista. Dessa forma, um último filtro se faz necessário, que é a escolha do escopo, ou seja, após a extração das tuplas o usuário escolhe quantas tuplas de cada predicado serão utilizadas para a construção do Mapa Conceitual.

### 5.1.1 Geografia

O primeiro artigo escolhido é da autora Patrícia Silva Gomes; Gomes (2011), da área de Ciências Humanas, da subárea Geografia (ANEXO A). A Seção do periódico escolhida, de modo aleatório, foi a terceira, com o título: “a reestruturação produtiva e os novos papéis assumidos pelo planejamento e regulação urbanos”. O texto selecionado tem 1.519 palavras.

Tabela 8. Frequência das categorias de palavras mais relevantes no texto de Geografia

#	Frequência dos léxicos	#	Frequência dos léxicos
1	N(468)	13	KS(16)
2	ART(265)	14	NUM(16)
3	ADJ(188)	15	PROPESS(14)
4	PREP(171)	16	PROSUB(11)
5	V(105)	17	N TEL(7)
6	NPROP(77)	18	PDEN(7)
7	KC(71)	19	N AP(5)
8	PCP(50)	20	N DAT(4)
9	PROADJ(49)	21	PRO-KS(2)
10	ADV(30)	22	KC (1)
11	VAUX(20)	23	V +(1)
12	PRO-KS-REL(18)	24	

#### 5.1.1.1 Os resultados das extrações pelo Text2MARK

O texto analisado apresentou um total de 1.794 palavras, sendo 575 diferentes entre si. A soma total das tuplas obtidas teve como média o peso de 8,71. Apenas as tuplas que ficaram acima da média foram inicialmente selecionadas. O total destas foi de 427 saídas. Após aplicar o filtro com cinco tuplas de cada predicado, o resultado final foram 42 tuplas, sendo estas escolhidas na ordem decrescente da soma do seu peso, com isso as palavras com maior frequência foram selecionadas, conforme a Tabela 8.

Ressalta-se que nem todos os predicados tiveram resultados, como pode ser visto na Tabela 9. Devido ao processo automatizado, algumas saídas não representaram o contexto escolhido, sendo estas desprezadas na construção do Mapa Conceitual. Desta forma, o tratamento de exceção se faz necessário. Na interface do Text2MARK, Figura 23 é possível ver todas as tuplas encontradas inicialmente. A interface permite que sejam selecionadas várias tuplas que representam o texto para serem adicionadas no escopo final. Neste texto vinte e quatro tuplas foram selecionadas para compor o MC.

No que tange ao tratamento de exceção, essa etapa é feita de forma manual ao final da seleção dos predicados realizada automaticamente pelo Text2MARK. Essa etapa requer uma

leitura prévia do texto e os predicados escolhidos podem ser adicionados no resultado final sem prejudicar em nada o texto selecionado automaticamente.

Tabela 9. Tuplas selecionadas automaticamente no texto de Geografia

#	Regra	Saídas do Text2MARK
1	N-V-N	Planejamento(N)-assume(V)-cidade(N)(34) <b>(USADO)</b>
2	N-V-N	Globalização(N)-apóia(V)-cidades(N)(25) <b>(USADO)</b>
3	N-V-N	cidades(N)-referir-(V)-pontos(N)(23)
4	N-V-N	1991(N)-identifica(V)-cidades(N)(22)
5	N-V-N	Consumo(N)-tendo(V)-cidades(N)(21) <b>(USADO)</b>
6	N-N-N	Cidade(N)-cidade(N)-mercadoria(N)(35)
7	N-N-N	Centros(N)-comando(N)-controle(N)(31) <b>(USADO)</b>
8	N-N-N	Economias(N)-brasileira(N)-cidades(N)(30) <b>(USADO)</b>
9	N-PREP-N	Cidade(N)-como(PREP)-todo(N)(44) <b>(USADO)</b>
10	N-PREP-N	1991(N)-como(PREP)-cidades(N)(42)
11	N-PREP-N	cidades(N)-como(PREP)-articuladoras(N)(41) <b>(USADO)</b>
12	N-PREP-N	Trabalho(N)-com(PREP)-emergência(N)(33) <b>(USADO)</b>
13	N-KC-PROADJ-N	Cidade(N)-contudo(KC)-outra(PROADJ)-ponta(N)(24)
14	N-KC-PROADJ-N	Espaços(N)-e(KC)-diversas(PROADJ)-escalas(N)(12)
15	N-KC-PROADJ-N	posição(N)-e(KC)-sua(PROADJ)-ação(N)(9)
16	N-KC-PROADJ-N	todo(N)-e(KC)-algumas(PROADJ)-partes(N)(8) <b>(USADO)</b>
17	N-KC-PROADJ-N	internet(N)-entanto(KC)-essas(PROADJ)-condições(N)(7)
18	N-V-ART-N	Planejamento(N)-assume(V)-a(ART)-cidade(N)(34) <b>(USADO)</b>
19	N-V-ART-N	planejamento(N)-adotar(V)-uma(ART)-formulação(N)(28) <b>(USADO)</b>
20	N-V-ART-N	reestruturação(N)-resulta(V)-uma(ART)-mudança(N)(26) <b>(USADO)</b>
21	N-V-ART-N	Globalização(N)-apóia(V)-as(ART)-cidades(N)(25) <b>(USADO)</b>
22	N-V-ART-N	cidades(N)-referir-(V)-os(ART)-pontos(N)(23)
23	N-VAUX-V-N	planejamento(N)-passa(VAUX)-adotar(V)-formulação(N)(26) <b>(USADO)</b>
24	N-VAUX-V-N	lugar(N)-passou(VAUX)-ocorrer(V)-mudanças(N)(22)
25	N-VAUX-V-N	Capital(N)-passa(VAUX)-assumir(V)-poder(N)(18) <b>(USADO)</b>
26	N-VAUX-V-N	entornos(N)-passam(VAUX)-receber(V)-indústrias(N)(18) <b>(USADO)</b>
27	N-VAUX-V-N	indústria(N)-começaram(VAUX)-ser(V)-lugar(N)(15)
28	N-ADV-V-N	Planejamento(N)-não(ADV)-assume(V)-cidade(N)(37) <b>(USADO)</b>
29	N-ADV-V-N	1991(N)-primeiro(ADV)-cunhou(V)-termo(N)(11)
30	N-PREP-NUM-N	palavras(N)-de(PREP)-2004(NUM)-planejamento(N)(18)
31	N-PREP-NUM-N	CGPs(N)-para(PREP)-trinta(NUM)-anos(N)(13)
32	N-PREP-NUM-N	Ação”(N)-conforme(PREP)-2003(NUM)-substituição(N)(10)
33	N-VAUX-VAUX-PCP-N	planejamento(N)-passa(VAUX)-ser(VAUX)-desafiado(PCP)-direita(N)(30) <b>(USADO)</b>
34	N-VAUX-VAUX-PCP-N	décadas(N)-passam(VAUX)-ser(VAUX)-substituídas(PCP)-políticas(N)(22)
35	N-V-ART-ADJ-N	modo(N)-coloca(V)-o(ART)-tradicional(ADJ)-planejamento(N)(18) <b>(USADO)</b>
36	N-V-ART-N-PREP-N	Globalização(N)-apóia(V)-as(ART)-cidades(N)-como(PREP)-articuladoras(N)(49) <b>(USADO)</b>
37	N-V-ART-N-PREP-N	cidades(N)-são(V)-os(ART)-centros(N)-de(PREP)-comando(N)(47) <b>(USADO)</b>
38	N-V-ART-N-PREP-N	planejamento(N)-apóia(V)-as(ART)-obras(N)-como(PREP)-forma(N)(43)
39	N-V-ART-N-PREP-N	planos(N)-regulam(V)-a(ART)-ação”(N)-por(PREP)-planos(N)(31) <b>(USADO)</b>
40	N-V-ART-N-PREP-N	cidades(N)-tornam-(V)-os(ART)-atores(N)-de(PREP)-articulação(N)(22) <b>(USADO)</b>
41	N-PROADJ-N-VAUX-PCP-N	cidades(N)-cada(PROADJ)-vez(N)-são(VAUX)-moldados(PCP)-compradores(N)(29)
42	N-PROADJ-N-VAUX-PCP-N	todo(N)-algumas(PROADJ)-partes(N)-é(VAUX)-encarado(PCP)-entorno(N)(12)

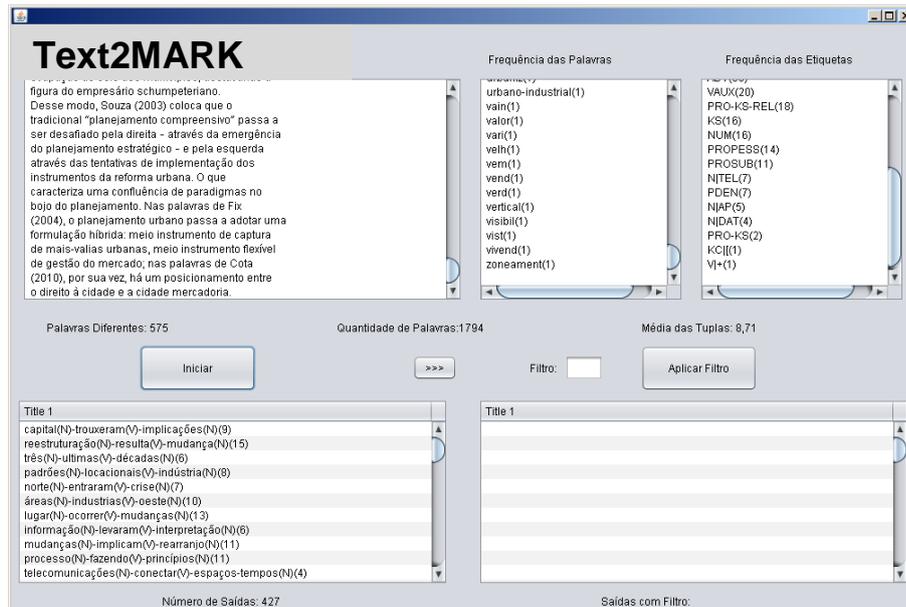


Figura 23. Abaixo à esquerda a seleção inicial das tuplas encontradas automaticamente

Tabela 10. 50% dos predicados não foram encontrados no texto analisado

#	Predicados não encontrados no texto
1	N-VAUX-V-PROADJ-N
2	N-PCP-VAUX-V-N
3	N-VAUX-V-V-N
4	N-ADV-V-V-N
5	N-ADJ-V-N-KC-N
6	N-V-ART-ADJ-ADJ-N
7	N-V-ADJ-V-PROADJ-N
8	N-PREP-N-V-NUM-N
9	N-V-ART-ADJ-ADJ-N
10	PROADJ-N-V-ART-ADJ-ADJ
11	N-PROADJ-N-VAUX-PCP-V-N

Tabela 11. Tuplas marcadas com COMPLEMENTO dão maior detalhe ao mapa

#	Regra	Saídas do Text2MARK após tratamento de exceção
1	N-V-N	reestruturação(N)-resulta(V)-mudança(N)(15) ( <b>USADO</b> )
2	N-V-N	Três(N)-últimas(V)-décadas(N)(6) ( <b>COMPLEMENTO</b> )
3	N-V-N	serviços(N)-incluem(V)-consultorias(N)(8) ( <b>COMPLEMENTO</b> )
4	N-N-N	Microeletrônica(N)-informática(N)-telecomunicações(N)(5) ( <b>COMPLEMENTO</b> )
5	N-N-N	Seguros(N)-distribuição(N)-vendas(N)(3) ( <b>COMPLEMENTO</b> )
6	N-N-N	reestruturação(N)-metrópole(N)-transformações(N)(13) ( <b>COMPLEMENTO</b> )
7	N-N-N	Telefonia(N)-redes(N)-informática(N)(6) ( <b>COMPLEMENTO</b> )
8	N-N-N	empreendimentos(N)-mudanças(N)-mercado(N)(12) ( <b>COMPLEMENTO</b> )
9	N-PREP-N	processos(N)-de(PREP)-desindustrializações(N)(9) ( <b>COMPLEMENTO</b> )
10	N-PREP-N	zonas(N)-de(PREP)-processamento(N)(9) ( <b>COMPLEMENTO</b> )
11	N-PREP-N	tecnologias(N)-de(PREP)-informação(N)(4) ( <b>COMPLEMENTO</b> )
12	N-PREP-N	interpretação(N)-de(PREP)-pós-fordismo(N)(2) ( <b>COMPLEMENTO</b> )
13	N-PREP-N	nichos(N)-de(PREP)-mercado(N)(6) ( <b>COMPLEMENTO</b> )

14	N-PREP-N	Serviços(N)-de(PREP)-limpeza(N)(6) <b>(COMPLEMENTO)</b>
15	N-PREP-N	metrópole(N)-em(PREP)-transformações(N)(7) <b>(COMPLEMENTO)</b>
16	N-PREP-N	Rede(N)-de(PREP)-telefonía(N)(4) <b>(COMPLEMENTO)</b>
17	N-PREP-N	Redes(N)-de(PREP)-informática(N)(5) <b>(COMPLEMENTO)</b>
18	N-PREP-N	atração(N)-de(PREP)-investimentos(N)(6) <b>(COMPLEMENTO)</b>
19	N-VAUX-V-N	entornos(N)-passam(VAUX)-receber(V)-indústrias(N)(18) <b>(USADO)</b>
20	N-VAUX-V-N	Governos(N)-passam(VAUX)-mudar(V)-postura(N)(13) <b>(USADO)</b>
21	N-V-ART-N- PREP-N	globalização(N)-apóia(V)-as(ART)-cidades(N)-como(PREP)- articuladoras(N)(49) <b>(USADO)</b>
22	N-V-ART-N- PREP-N	cidades(N)-tornam-(V)-os(ART)-atores(N)-de(PREP)-articulação(N)(22) <b>(USADO)</b>
23	N-V-ART-N- PREP-N	meio(N)-inclui(V)-a(ART)-rede(N)-de(PREP)-telefonía(N)(11) <b>(COMPLEMENTO)</b>
24	N-V-ART-N- PREP-N	reforma(N)-caracteriza(V)-uma(ART)-confluência(N)-de(PREP)- paradigmas(N)(17) <b>(COMPLEMENTO)</b>

### 5.1.1.2 As inclusões no CMapTools

Durante esta etapa percebeu-se que nem todas as tuplas foram necessárias para a composição do Mapa, como pode ser visto na Tabela 10. Do total de 42 tuplas selecionadas automaticamente, 23 foram utilizadas, resultando em aproximadamente 55%. Da mesma forma, as tuplas escolhidas posteriormente no tratamento de exceção, apesar de darem um sentido a mais no mapa, servem, após análise, para complementar as informações do texto. Ou seja, para uma descrição mais detalhada tais informações seriam bastante úteis. Das 24 tuplas selecionadas como complemento, 5 foram usadas, chegando a aproximadamente 21% do total, conforme a Tabela 11.

Ressalta-se que para a construção, foi necessário acompanhar a leitura do texto, identificando onde as tuplas foram extraídas e inserindo, quando necessário, alguma informação adicional que no contexto não foi alcançado pelo Text2MARK. Esse procedimento serviu também para identificar quais as tuplas foram selecionadas automaticamente e que se complementavam pela inclusão de outros léxicos ou predicados.

O mapa gerado, apesar de abstrair alguns detalhes sobre as técnicas e recursos que sofreram modificações no decorrer do tempo, consegue em sua totalidade representar o conceito central que a distribuição econômica mundial, representada pelas grandes cidades o fazem.

### 5.1.1.3 O resultado do Mapa Gerado

Optou-se por incluir o mapa conceitual na página inteira, maximizando-o, para facilitar sua leitura. Sua representação pode ser vista na Figura 24.

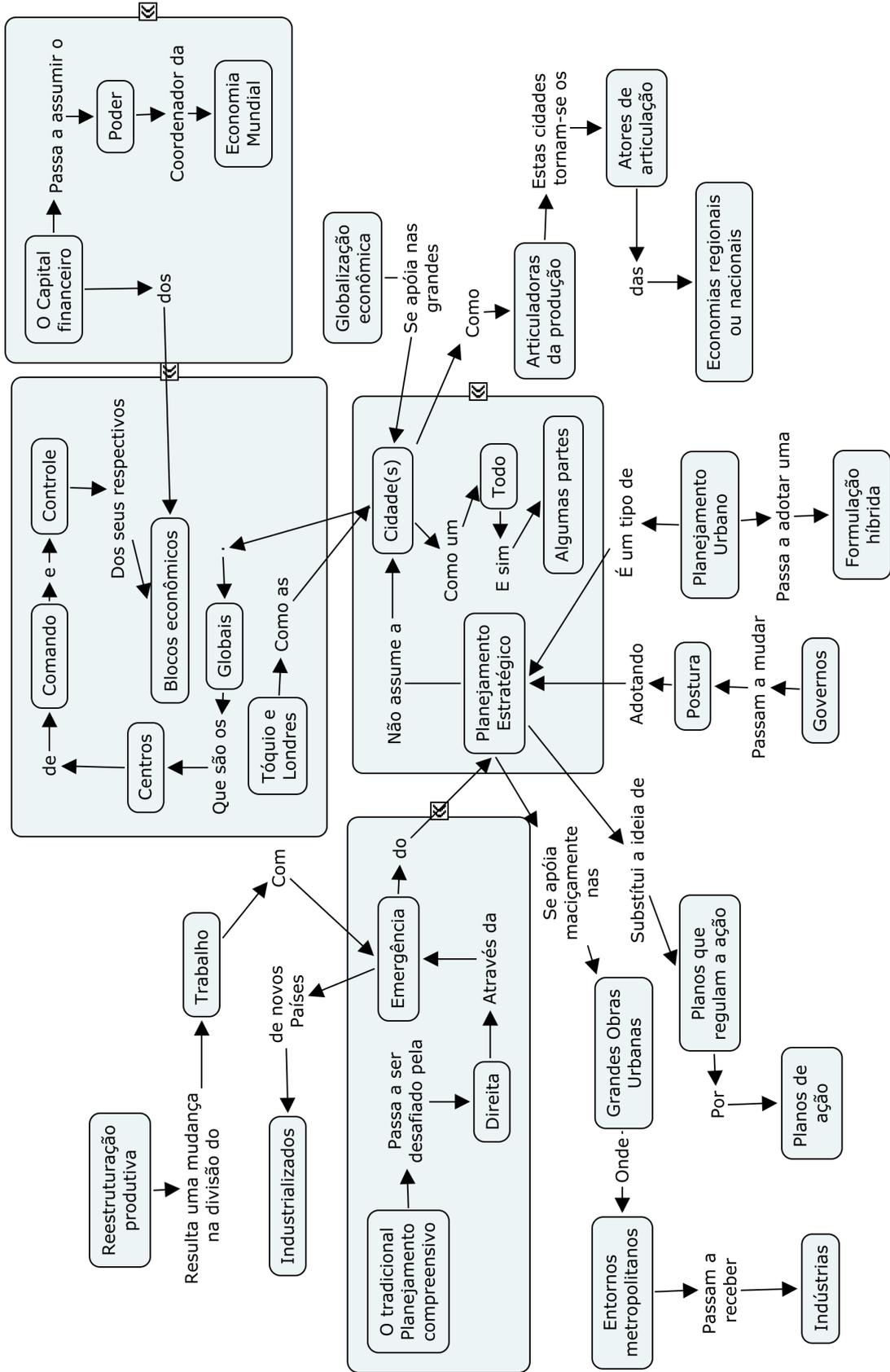


Figura 24. O mapa conceitual do texto com base nas tuplas obtidas no Text2MARK

### 5.1.2 EAD

O segundo texto escolhido, (ANEXO B), tem como autores a Profa. Marianne Kogut Eliasquevici e o Prof. Arnaldo Corrêa Prado Junior; Eliasquevici et al (2008). Da área de educação à distância, Ciências Humanas, da subárea Letras. A Seção do artigo escolhido, de modo aleatório, foi a segunda, com o título: “A incerteza pela perspectiva da ciência”. O texto selecionado tem 760 palavras.

Tabela 12. Frequência dos léxicos mais relevantes no texto de EAD

#	Frequência dos léxicos	#	Frequência dos léxicos
1	N(210)	11	VAUX(16)
2	ART(148)	12	KS(14)
3	PREP (76)	13	PRO-KS-REL(14)
4	ADJ(75)	14	ADV(12)
5	PREP(75)	15	PROPESS(9)
6	V(63)	16	PDEN(6)
7	NPROP(37)	17	PROSUB(6)
8	KC(30)	18	N AP(1)
9	PCP(23)	19	N TEL(1)
10	PROADJ(18)	20	

#### 5.1.2.1 Os resultados das extrações pelo Text2MARK

O texto analisado apresentou um total de 852 palavras, sendo 324 diferentes entre si. A soma total das tuplas obtidas teve como média o peso de 4,16. Apenas as tuplas que ficaram acima da média foram inicialmente selecionadas. O total destas foi de 208 saídas. Após aplicar o filtro com cinco tuplas de cada predicado, o resultado final foram 41 tuplas, sendo estas escolhidas na ordem decrescente da soma do seu peso. Com isso as palavras com maior frequência foram selecionadas, conforme a Tabela 13.

Ressalta-se que nem todos os predicados tiveram resultados, como pode ser visto na Tabela 14. Devido ao processo automatizado, algumas saídas não representaram o contexto escolhido, sendo estas desprezadas na construção do Mapa Conceitual, desta forma, o tratamento de exceção se faz necessário. Na interface do Text2MARK, é possível que sejam selecionadas várias tuplas que representam o texto para serem adicionadas no escopo final (Figura 25). Vinte tuplas adicionais são escolhidas para compor o Mapa, conforme a Tabela 15.

Tabela 13. Tuplas selecionadas automaticamente no texto de EAD

#	Regra	Saídas do Text2MARK
1	N-V-N	século(N)-foi(V)-conhecimento(N)(23) <b>(USADO)</b>
2	N-V-N	Métodos(N)-levariam(V)-conhecimento(N)(14) <b>(USADO)</b>
3	N-V-N	século(N)-foi(V)-processo(N)(14) <b>(USADO)</b>
4	N-V-N	Incerteza(N)-era(V)-valores(N)(13) <b>(USADO)</b>
5	N-V-N	Natureza(N)-transforma(V)-sistema(N)(12) <b>(USADO)</b>
6	N-N-N	ciências(N)-abordagem(N)-conhecimento(N)(19) <b>(USADO)</b>
7	N-N-N	conhecimento(N)-humanidade(N)-progressos(N)(16) <b>(USADO)</b>
8	N-N-N	desenvolvimento(N)-homem(N)-Conhecimento(N)(16) <b>(USADO)</b>
9	N-N-N	Métodos(N)-conhecimento(N)-realidade(N)(15) <b>(USADO)</b>
10	N-PREP-N	Incerteza(N)-como(PREP)-valores(N)(21) <b>(USADO)</b>
11	N-PREP-N	ciência(N)-como(PREP)-objetiva(N)(20)
12	N-PREP-N	abordagem(N)-para(PREP)-conhecimento(N)(18) <b>(USADO)</b>
13	N-PREP-N	leis(N)-por(PREP)-razão(N)(17) <b>(USADO)</b>
14	N-PREP-N	Pensamento(N)-por(PREP)-mudanças(N)(16) <b>(USADO)</b>
15	N-KC-PROADJ-N	Incerteza(N)-e(KC)-seu(PROADJ)-tratamento(N)(10) <b>(USADO)</b>
16	N-KC-PROADJ-N	Contradição(N)-e(KC)-esta(PROADJ)-forma(N)(6) <b>(USADO)</b>
17	N-KC-PROADJ-N	Mecanismos(N)-e(KC)-mesmas(PROADJ)-relações(N)(5) <b>(USADO)</b>
18	N-V-ART-N	século(N)-foi(V)-o(ART)-conhecimento(N)(23) <b>(USADO)</b>
19	N-V-ART-N	razão(N)-era(V)-uma(ART)-dádiva(N)(15) <b>(USADO)</b>
20	N-V-ART-N	Métodos(N)-levariam(V)-o(ART)-conhecimento(N)(14) <b>(USADO)</b>
21	N-V-ART-N	Incerteza(N)-era(V)-os(ART)-valores(N)(13) <b>(USADO)</b>
22	N-V-ART-N	Natureza(N)-transforma(V)-um(ART)-sistema(N)(12) <b>(USADO)</b>
23	N-VAUX-V-N	certeza(N)-foi(VAUX)-é(V)-indestrutibilidade(N)(9)
24	N-VAUX-V-N	Indivíduo(N)-poderia(VAUX)-encontrar(V)-respostas(N)(8) <b>(USADO)</b>
25	N-VAUX-V-N	homem(N)-deveria(VAUX)-estar(V)-corpo(N)(6)
26	N-ADV-V-N	conhecimento(N)-não(ADV)-é(V)-equivalente(N)(19) <b>(USADO)</b>
27	N-ADV-V-N	razão(N)-não(ADV)-acomoda(V)-contradição(N)(12) <b>(USADO)</b>
28	N-ADV-V-N	princípios(N)-sempre(ADV)-estiveram(V)-crise(N)(6)
29	N-VAUX-VAUX-PCP-N	incertezas(N)-podem(VAUX)-ser(VAUX)-reduzidas(PCP)-pesquisas(N)(18) <b>(USADO)</b>
30	N-VAUX-VAUX-PCP-N	Mudanças(N)-podem(VAUX)-ser(VAUX)-encontradas(PCP)-discussões(N)(13) <b>(USADO)</b>
31	N-VAUX-VAUX-PCP-N	Problemas(N)-vem(VAUX)-sendo(VAUX)-desenvolvida(PCP)-autores(N)(8)
32	N-V-ART-ADJ-N	pesquisas(N)-dizer(V)-A(ART)-MAIOR(ADJ)-CONTRIBUIÇÃO(N)(6) <b>(USADO)</b>
33	N-V-ART-ADJ-N	aspectos(N)-tendem(V)-as(ART)-tradicionalis(ADJ)-práticas(N)(5) <b>(USADO)</b>
34	N-V-ART-N-PREP-N	certeza(N)-considerando(V)-a(ART)-existência(N)-de(PREP)-incertezas(N)(15) <b>(USADO)</b>
35	N-V-ART-N-PREP-N	modelos(N)-movem(V)-a(ART)-ciência(N)-à(PREP)-medida(N)(11) <b>(USADO)</b>
36	N-V-ART-N-PREP-N	Positivismo(N)-é(V)-um(ART)-procedimento(N)-de(PREP)-entendimento(N)(9) <b>(USADO)</b>
37	N-V-ART-N-PREP-N	autores(N)-defendem(V)-a(ART)-necessidade(N)-de(PREP)-reestruturação(N)(5) <b>(USADO)</b>
38	N-V-ART-N-PREP-N	Experiência(N)-é(V)-a(ART)-utilização(N)-de(PREP)-métodos(N)(4) <b>(USADO)</b>
39	N-PROADJ-N-VAUX-PCP-N	Início(N)-seus(PROADJ)-princípios(N)-sendo(VAUX)-questionados(PCP)-crise(N)(13)
40	N-PROADJ-N-VAUX-PCP-N	fenômenos(N)-mesmo(PROADJ)-prisma(N)-são(VAUX)-entendidos(PCP)-fenômenos(N)(12) <b>(USADO)</b>
41	N-PROADJ-N-VAUX-PCP-N	progresso(N)-Essa(PROADJ)-doutrina(N)-ter(VAUX)-ultrapassado(PCP)-estágio(N)(10) <b>(USADO)</b>

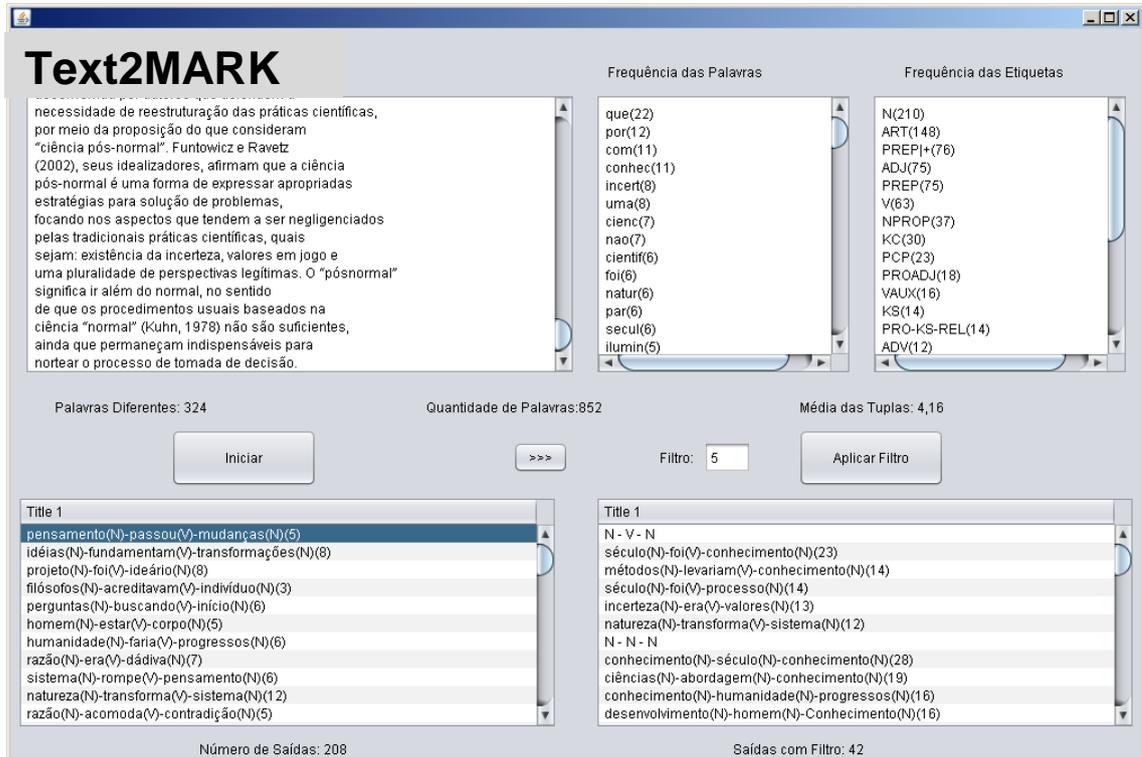


Figura 25. O Text2MARK permite selecionar uma ou várias tuplas e adicionar ao final

Tabela 14. 54% dos predicados não foram encontrados no texto analisado

#	Predicados não encontrados no texto
1	N-PREP-NUM-N
2	N-VAUX-V-PROADJ-N
3	N-PCP-VAUX-V-N
4	N-VAUX-V-V-N
5	N-ADV-V-V-N
6	N-ADJ-V-N-KC-N
7	N-V-ART-ADJ-ADJ-N
8	N-V-ADJ-V-PROADJ-N
9	N-PREP-N-V-NUM-N
10	N-V-ART-ADJ-ADJ-N
11	N-PROADJ-N-V-ART-ADJ-ADJ
12	N-PROADJ-N-VAUX-PCP-V-N

Tabela 15. Tuplas complemento são requeridas para dar mais detalhe ao MC

#	Regra	Saídas do Text2MARK após tratamento de exceção
1	N-V-N	razão(N)-acomoda(V)-contradição(N)(5) ( <b>COMPLEMENTO</b> )
2	N-N-N	métodos(N)-conhecimento(N)-realidade(N)(15) ( <b>COMPLEMENTO</b> )
3	N-N-N	Ciência(N)-forma(N)-estratégias(N)(11) ( <b>COMPLEMENTO</b> )
4	N-PREP-N	Construção(N)-de(PREP)-idéias(N)(3) ( <b>COMPLEMENTO</b> )
5	N-PREP-N	Sistema(N)-com(PREP)-pensamento(N)(16) ( <b>COMPLEMENTO</b> )
6	N-PREP-N	racional(N)-em(PREP)-essência(N)(3) ( <b>COMPLEMENTO</b> )
7	N-PREP-N	relações(N)-de(PREP)-causa(N)(2) ( <b>COMPLEMENTO</b> )
8	N-PREP-N	utilização(N)-de(PREP)-métodos(N)(3) ( <b>COMPLEMENTO</b> )

9	N-PREP-N	Criação(N)-de(PREP)-leis(N)(3) ( <b>COMPLEMENTO</b> )
10	N-PREP-N	procedimento(N)-de(PREP)-entendimento(N)(4) ( <b>COMPLEMENTO</b> )
11	N-PREP-N	idéias(N)-por(PREP)-vez(N)(15) ( <b>COMPLEMENTO</b> )
12	N-PREP-N	necessidade(N)-de(PREP)-reestruturação(N)(3) ( <b>COMPLEMENTO</b> )
13	N-PREP-N	forma(N)-de(PREP)-estratégias(N)(4) ( <b>COMPLEMENTO</b> )
14	N-PREP-N	Solução(N)-de(PREP)-problemas(N)(3) ( <b>COMPLEMENTO</b> )
15	N-PREP-N	pluralidade(N)-de(PREP)-perspectivas(N)(3) ( <b>COMPLEMENTO</b> )
16	N-V-ART-N	Positivismo(N)-é(V)-um(ART)-procedimento(N)(7) ( <b>USADO</b> )
17	N-V-ART-N	modelos(N)-movem(V)-a(ART)-ciência(N)(10) ( <b>COMPLEMENTO</b> )
18	N-V-ART-N	Realidade(N)-complica(V)-as(ART)-transformações(N)(6) ( <b>COMPLEMENTO</b> )
19	N-V-ART-N	incertezas(N)-expandem(V)-um(ART)-aumento(N)(10) ( <b>COMPLEMENTO</b> )
20	N-V-ART-N	Autores(N)-defendem(V)-a(ART)-necessidade(N)(4) ( <b>USADO</b> )

### 5.1.2.2 As inclusões no CMapTools

Este texto em específico se mostrou um grande desafio, na sua estrutura. Palavras importantes no texto apareceram apenas uma vez. Logo, como sua frequência foi baixa, não entra nas regras existentes no Text2MARK. Apesar de ter feito o tratamento de exceção com palavras que o complementam e dão mais detalhe ao Mapa, uma leitura detalhada do texto se fez necessária.

O foco do texto era ao redor do substantivo CONHECIMENTO, logo, ordenar as ideias e os conceitos que faziam relação com este NOME mostrou-se uma tarefa desafiadora. Como o texto trata de conceitos diversos e vários encadeamentos, para facilitar sua leitura foram colocados marcadores circulares para identificar a ordem que deve ser seguida pelo leitor do Mapa. Características que não se fizeram necessárias no texto analisado no item 6.1.1.

Nesta extração, do total das 41 tuplas selecionadas automaticamente, 35 foram utilizadas diretamente na construção do Mapa, representando mais de 85% do total. As tuplas complementares foram no total de 20, sendo apenas 2 utilizadas, ou seja, 10% do total. Ressalta-se que as tuplas complementares não utilizadas servem para complementar o mapa deixando-o mais detalhado.

### 5.1.2.3 O resultado do Mapa Gerado

A figura 26 mostra a representação do mapa e das complexidades encontradas na sua construção. Neste modelo houve grande acerto do Text2MARK, porém, houve necessidade de ordenar suas entradas para facilitar a compreensão.

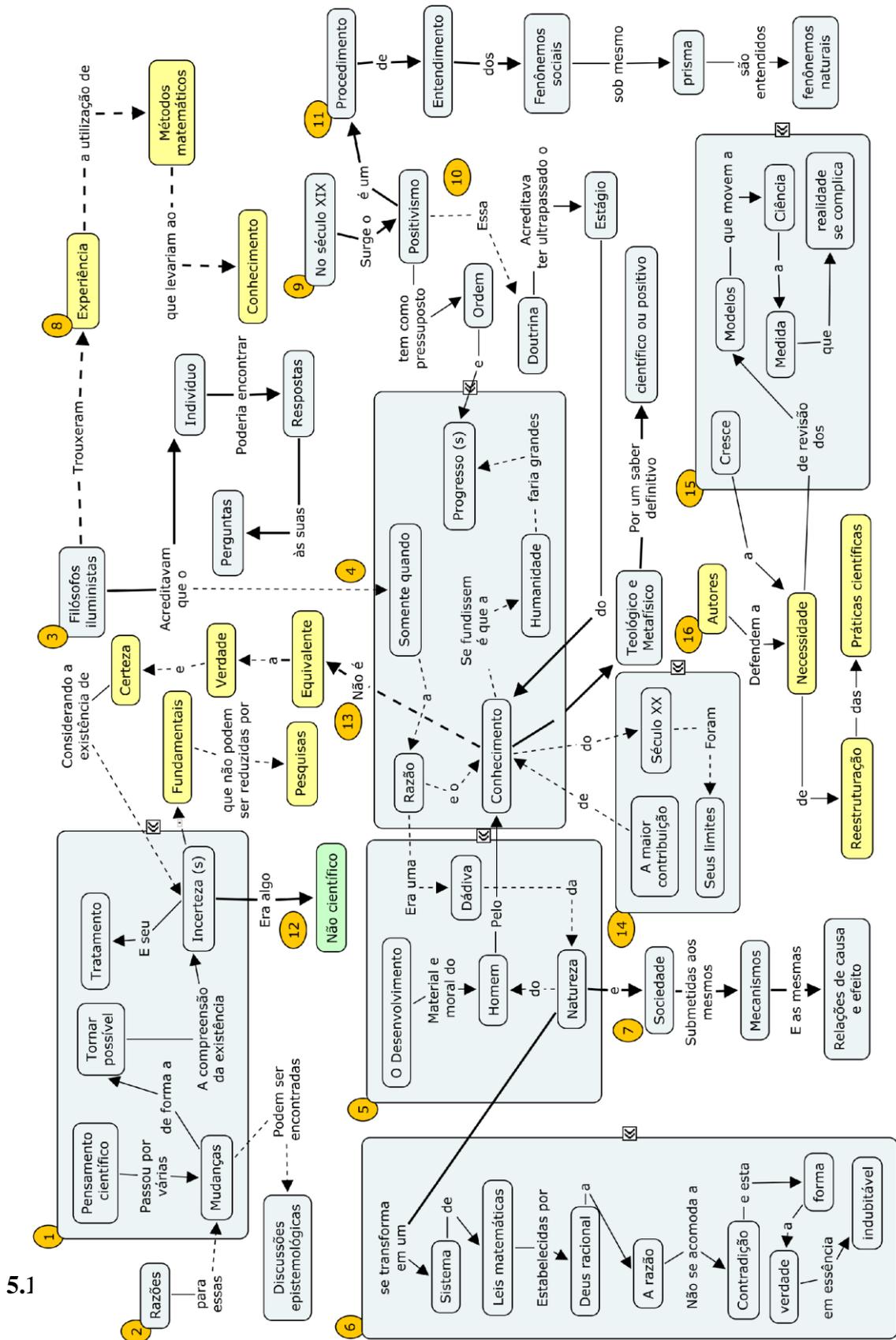


Figura 26. Círculos em laranja orientam o usuário sobre a ordem de leitura do MC

O terceiro texto escolhido, (ANEXO C), tem como autores o Prof. João Marcos Cardoso de Sousa e a Profa. Ida Lucia Machado; Sousa et al (2011), da área de Educação, Ciências Humanas, da subárea Linguística. A Seção do artigo escolhido, de modo aleatório, foi à quinta, item 5.1, com o título: “A transgressão simples”. O texto selecionado tem 947 palavras.

Tabela 16. Frequência dos léxicos mais relevantes no texto de Linguística

#	Frequência dos léxicos	#	Frequência dos léxicos
1	N(250)	13	PRO-KS-REL(17)
2	ART(150)	14	PROSUB(13)
3	PREP(107)	15	VAUX(13)
4	V(103)	16	KS(12)
5	PREP (98)	17	PDEN(7)
6	ADJ(53)	18	NUM(5)
7	NPROP(45)	19	N AP(4)
8	PROADJ(37)	20	PRO-KS(3)
9	ADV(34)	21	N DAT(2)
10	KC(31)	22	IN(1)
11	PCP(23)	23	N TEL(1)
12	PROPESS(18)	24	

### 5.1.3.1 Os resultados das extrações pelo Text2MARK

O texto analisado apresentou um total de 1079 palavras, sendo 360 diferentes entre si. A soma total das tuplas obtidas teve como média o peso de 5,44. Apenas as tuplas que ficaram acima da média foram inicialmente selecionadas. O total destas foi de 243 saídas. Após aplicar o filtro com 6 tuplas de cada predicado, o resultado final foram 50 tuplas, sendo estas escolhidas na ordem decrescente da soma do seu peso, com isso as palavras com maior frequência foram selecionadas, conforme Tabela 17.

Ressalta-se que nem todos os predicados tiveram resultados, como pode ser visto na Tabela 18. Devido ao processo automatizado, algumas saídas não representaram o contexto escolhido, sendo estas desprezadas na construção do Mapa Conceitual, desta forma, o tratamento de exceção se faz necessário.

Tabela 17. Tuplas selecionadas automaticamente no texto de Linguística

#	Regra	Saídas do Text2MARK
1	N-V-N	Herói(N)-tem(V)-transgressão(N)(30) ( <b>USADO</b> )
2	N-V-N	herói(N)-pulveriza(V)-imagens(N)(17) ( <b>USADO</b> )
3	N-V-N	herói(N)-fazer(V)-percurso(N)(17) ( <b>USADO</b> )
4	N-V-N	herói(N)-passar(V)-sentença(N)(17) ( <b>USADO</b> )

5	N-V-N	herói(N)-é(V)-assassino(N)(17) <b>(USADO)</b>
6	N-V-N	multiplicidade(N)-enunciativa(V)-herói(N)(16) <b>(USADO)</b>
7	N-N-N	imagem(N)-herói(N)-herói(N)(32) <b>(USADO)</b>
8	N-N-N	herói(N)-herói(N)-voz(N)(29)
9	N-N-N	herói(N)-transgressão(N)-missão(N)(24) <b>(USADO)</b>
10	N-N-N	imagem(N)-herói(N)-discurso(N)(22) <b>(USADO)</b>
11	N-N-N	imagem(N)-herói(N)-imagens(N)(20) <b>(USADO)</b>
12	N-N-N	sujeito(N)-função(N)-herói(N)(20) <b>(USADO)</b>
13	N-PREP-N	herói(N)-por(PREP)-voz(N)(36) <b>(USADO)</b>
14	N-PREP-N	função(N)-como(PREP)-herói(N)(31) <b>(USADO)</b>
15	N-PREP-N	transgressão(N)-por(PREP)-missão(N)(31) <b>(USADO)</b>
16	N-PREP-N	Herói(N)-como(PREP)-pessoa(N)(31) <b>(USADO)</b>
17	N-PREP-N	Comunicação(N)-Por(PREP)-efeito(N)(30)
18	N-PREP-N	assassino(N)-por(PREP)-leitor(N)(27) <b>(USADO)</b>
19	N-KC-PROADJ-N	transgressão(N)-ou(KC)-seu(PROADJ)-efeito(N)(19) <b>(USADO)</b>
20	N-KC-PROADJ-N	sujeito(N)-ou(KC)-sua(PROADJ)-função(N)(11) <b>(USADO)</b>
21	N-KC-PROADJ-N	sentença(N)-ou(KC)-sua(PROADJ)-liberdade(N)(8) <b>(USADO)</b>
22	N-KC-PROADJ-N	crime(N)-e(KC)-diversos(PROADJ)-diálogos(N)(7) <b>(USADO)</b>
23	N-KC-PROADJ-N	comunicação(N)-e(KC)-própria(PROADJ)-homonímia(N)(5) <b>(USADO)</b>
24	N-V-ART-N	herói(N)-tem(V)-a(ART)-transgressão(N)(30) <b>(USADO)</b>
25	N-V-ART-N	herói(N)-tratar(V)-um(ART)-herói(N)(30)
26	N-V-ART-N	herói(N)-é(V)-o(ART)-assassino(N)(17) <b>(USADO)</b>
27	N-V-ART-N	comunicação(N)-tem(V)-uma(ART)-frequência(N)(16) <b>(USADO)</b>
28	N-V-ART-N	multiplicidade(N)-enunciativa(V)-o(ART)-herói(N)(16) <b>(USADO)</b>
29	N-V-ART-N	recurso(N)-provocar(V)-o(ART)-efeito(N)(13) <b>(USADO)</b>
30	N-VAUX-V-N	herói(N)-deverá(VAUX)-passar(V)-sentença(N)(20) <b>(USADO)</b>
31	N-VAUX-V-N	gênero(N)-Podemos(VAUX)-perceber(V)-obra(N)(12) <b>(USADO)</b>
32	N-VAUX-V-N	assassino(N)-é(VAUX)-É(V)-leitor(N)(6)
33	N-ADV-V-N	Herói(N)-não(ADV)-tratar(V)-herói(N)(40) <b>(USADO)</b>
34	N-ADV-V-N	recurso(N)-intensamente(ADV)-provocar(V)-efeito(N)(15) <b>(USADO)</b>
35	N-ADV-V-N	Velho(N)-Não(ADV)-tocarei(V)-Irmão(N)(15) <b>(USADO)</b>
36	N-ADV-V-N	trama(N)-Assim(ADV)-percebemos(V)-imagem(N)(15) <b>(USADO)</b>
37	N-ADV-V-N	Intenção(N)-notadamente(ADV)-tem(V)-russo(N)(10)
38	N-ADV-V-N	honradíssima(N)-Assim(ADV)-começou(V)-Mítia(N)(8) <b>(USADO)</b>
39	N-PREP-NUM-N	nota(N)-em(PREP)-376(NUM)-anos(N)(4) <b>(USADO)</b>
40	N-VAUX-V-V-N	frequência(N)-podemos(VAUX)-perceber(V)-sentir(V)-enunciado(N)(11) <b>(USADO)</b>
41	N-V-ART-ADJ-N	Mítia(N)-atrapalhou(V)-as(ART)-primeiras(ADJ)-palavras(N)(7) <b>(USADO)</b>
42	N-V-ART-ADJ-N	sentido(N)-confirmamos(V)-a(ART)-intensa(ADJ)-atividade(N)(6) <b>(USADO)</b>
43	N-ADV-V-V-N	nome(N)-provavelmente(ADV)-ouviu(V)-falar(V)-vez(N)(10) <b>(USADO)</b>
44	N-ADV-V-V-N	frequência(N)-bem(ADV)-perceber(V)-sentir(V)-enunciado(N)(9) <b>(USADO)</b>
45	N-ADV-V-V-N	juízo(N)-mesmo(ADV)-verá(V)-haverá(V)-homem(N)(8) <b>(USADO)</b>
46	N-V-ART-N-PREP-N	Herói(N)-tem(V)-a(ART)-transgressão(N)-por(PREP)-missão(N)(52) <b>(USADO)</b>
47	N-V-ART-N-PREP-N	ironia(N)-lampeja(V)-o(ART)-efeito(N)-de(PREP)-verdade(N)(14) <b>(USADO)</b>
48	N-V-ART-N-PREP-N	personagens(N)-provoca(V)-o(ART)-leitor(N)-a(PREP)-sensação(N)(9) <b>(USADO)</b>
49	N-V-ART-N-PREP-N	obra(N)-requer(V)-um(ART)-estado(N)-de(PREP)-abertura(N)(7) <b>(USADO)</b>
50	N-V-ART-N-PREP-N	Fiódor(N)-despojou(V)-a(ART)-herança(N)-de(PREP)-mãe(N)(4) <b>(USADO)</b>

Tabela 18. 50% dos predicados não foram encontrados no texto analisado

#	Predicados não encontrados no texto
1	N-VAUX-V-PROADJ-N
2	N-PCP-VAUX-V-N
3	N-VAUX-VAUX-PCP-N
4	N-ADJ-V-N-KC-N
5	N-V-ART-ADJ-ADJ-N
6	N-V-ADJ-V-PROADJ-N
7	N-PROADJ-N-VAUX-PCP-N
8	N-PREP-N-V-NUM-N
9	N-V-ART-ADJ-ADJ-N
10	PROADJ-N-V-ART-ADJ-ADJ
11	N-PROADJ-N-VAUX-PCP-V-N

Tabela 19. Menor quantidade de tuplas adicionais e maior utilização das mesmas

#	Regra	Saídas do Text2MARK após tratamento de exceção
1	N-V-N	transgressão(N)-provoca(V)-diálogos(N)(9) <b>(USADO)</b>
2	N-V-N	personagens(N)-provoca(V)-leitor(N)(8) <b>(USADO)</b>
3	N-PREP-N	troca(N)-de(PREP)-palavras(N)(4) <b>(USADO)</b>
4	N-PREP-N	sujeito(N)-como(PREP)-enunciador(N)(21) <b>(USADO)</b>
5	N-PREP-N	preço(N)-por(PREP)-deslize(N)(23) <b>(USADO)</b>
6	N-PREP-N	efeito(N)-de(PREP)-transgressão(N)(15) <b>(USADO)</b>
7	N-PREP-N	ato(N)-como(PREP)-assassinato(N)(22) <b>(USADO)</b>
8	N-PREP-N	ação(N)-de(PREP)-transgressão(N)(13) <b>(USADO)</b>
9	N-PREP-N	herói(N)-em(PREP)-imagens(N)(16) <b>(USADO)</b>
10	N-PREP-N	herói(N)-em(PREP)-discurso(N)(18) <b>(USADO)</b>
11	N-PREP-N	morte(N)-para(PREP)-libertação(N)(12) <b>(USADO)</b>
12	N-V-ART-N	Fiódor(N)-despojou(V)-a(ART)-herança(N)(3) <b>(USADO)</b>
13	N-V-ART-N	transgressão(N)-provoca(V)-os(ART)-diálogos(N)(9) <b>(USADO)</b>

### 5.1.3.2 As inclusões no CMapTools

O texto analisado apresenta características distintas dos anteriores. Neste cenário são tratadas citações diretas de um autor, o russo Dostoievski. A análise científica central do texto é a linguística e suas complexidades. São ressaltadas características gramaticais que podem pertencer a um mesmo sujeito em diversas ocasiões, neste caso, configurando-se como uma transgressão. Devido à complexidade deste texto é que foram escolhidas 6 tuplas de cada predicado, buscando o máximo de representação dentro de cada média.

No geral vários fragmentos do texto são identificados diretamente pelo Text2MARK, porém, palavras que recaem sobre um trecho específico, apresentaram maior frequência, logo, maior importância dentro do contexto central. Devido isso foi necessário replicar essas palavras dentro do mapa para que sua leitura não ficasse confusa. As palavras HERÓI e

TRANSGRESSÃO foram ressaltadas em cores diferentes. Por ser visual, decidiu-se utilizar a cor e não outra forma para representar essa duplicação.

Utilizaram-se ainda formas diferentes de pontilhados nas linhas de ligação entre os conceitos. Serve para destacar quando uma ligação utiliza conceito existente em seu contexto. Ressalta-se a utilização de um léxico do tipo VERBO sendo representado por uma elipse, o verbo HAVERÁ. Dentro dos predicados, há uma regra de dois léxicos VERBOS próximos. Como a notação padrão da construção de tuplas é o VERBO como elo de ligação entre os NOMES, essa ocorrência precisou ser representada de forma diferente.

No que concerne a extração, o Text2MARK conseguiu gerar 45 tuplas válidas no total de 50, ou seja, 90% de aproveitamento. Já nas tuplas adicionais foram utilizadas 13 de 13, ou seja, 100% do total, conforme Tabela 19.

#### *5.1.3.3 O resultado do Mapa Gerado*

A construção do mapa conceitual deste texto foi um dos mais interessantes e desafiadores para a ferramenta. Como o texto trata basicamente de regras gramaticais, em específico a transgressão na linguística, esperava-se que os resultados obtidos automaticamente pelo Text2MARK fossem muito próximos da totalidade, conforme a Figura 27.



### 5.1.4 Meio Ambiente

O quarto texto escolhido, (ANEXO D), tem como autora Joaquina Barata Teixeira; Teixeira (2000), da área de Meio Ambiente, Ciências Humanas, da subárea Sustentabilidade. A Seção do artigo escolhido, de modo aleatório, foi a segunda, item sem numeração, com o título: “O Conceito de Desenvolvimento e os rumos para a auto-sustentabilidade”.

Tabela 20. Frequência dos léxicos mais relevantes no texto de Meio Ambiente

#	Frequência dos léxicos	#	Frequência dos léxicos
1	N(192)	13	PROADJ(10)
2	ART(130)	14	VAUX(9)
3	ADJ(80)	15	PROPESS(7)
4	PREP(72)	16	PROSUB(6)
5	PREP +(65)	17	PRO-KS-REL(5)
6	V(47)	18	NUM(4)
7	KC(31)	19	ADV-KS-REL(3)
8	ADV(22)	20	PDEN(3)
9	NPROP(20)	21	KC (1)
10	KS(19)	22	N AP(1)
11	PCP(17)	23	N DAT(1)
12	N TEL(10)	24	PRO-KS(1)

#### 5.1.4.1 Os resultados das extrações pelo Text2MARK

O texto analisado apresentou um total de 781 palavras, sendo 329 diferentes entre si. A soma total das tuplas obtidas teve como média o peso de 3,88. Apenas as tuplas que ficaram acima da média foram inicialmente selecionadas. O total destas foi de 173 saídas. Após aplicar o filtro com 5 tuplas de cada predicado, o resultado final foram 32 tuplas, sendo estas escolhidas na ordem decrescente da soma do seu peso, com isso as palavras com maior frequência foram selecionadas, conforme Tabela 21.

Ressalta-se que nem todos os predicados tiveram resultados, como pode ser visto na Tabela 22. Devido ao processo automatizado, algumas saídas não representaram o contexto escolhido, sendo estas desprezadas na construção do Mapa Conceitual, desta forma, o tratamento de exceção se faz necessário.

Tabela 21. Tuplas selecionadas automaticamente no texto de Meio Ambiente

#	Regra	Saídas do Text2MARK
1	N-V-N	Desenvolvimento(N)-pôs(V)-ação(N)(11) <b>(USADO)</b>
2	N-V-N	desenvolvimento(N)-afirma(V)-solução(N)(11) <b>(USADO)</b>
3	N-V-N	sociedade(N)-conhecerem(V)-devastação(N)(11) <b>(USADO)</b>
4	N-V-N	sociedades(N)-têm(V)-história(N)(10) <b>(USADO)</b>
5	N-V-N	sociedades(N)-evoluem(V)-etapas(N)(9) <b>(USADO)</b>
6	N-N-N	conceito(N)-desenvolvimento(N)-idéia(N)(13) <b>(USADO)</b>
7	N-N-N	mundo(N)-teoria(N)-desenvolvimento(N)(13)
8	N-N-N	atuação(N)-mercado(N)-desenvolvimento(N)(13) <b>(USADO)</b>
9	N-N-N	movimento(N)-horizonte(N)-desenvolvimento(N)(12)
10	N-N-N	desenvolvimento(N)-ação(N)-princípio(N)(12) <b>(USADO)</b>
11	N-PREP-N	sociedade(N)-para(PREP)-devastação(N)(16) <b>(USADO)</b>
12	N-PREP-N	gerações(N)-como(PREP)-princípio(N)(14) <b>(USADO)</b>
13	N-PREP-N	branco(N)-como(PREP)-patamar(N)(13) <b>(USADO)</b>
14	N-PREP-N	fora(N)-como(PREP)-jeito(N)(13)
15	N-PREP-N	sócio-cultural(N)-com(PREP)-manutenção(N)(13) <b>(USADO)</b>
16	N-KC-PROADJ-N	dependência(N)-mas(KC)-suas(PROADJ)-raízes(N)(5) <b>(USADO)</b>
17	N-V-ART-N	sociedades(N)-têm(V)-uma(ART)-história(N)(15) <b>(USADO)</b>
18	N-V-ART-N	desenvolvimento(N)-afirma(V)-a(ART)-solução(N)(11) <b>(USADO)</b>
19	N-V-ART-N	sociedade(N)-conhecerem(V)-a(ART)-devastação(N)(11) <b>(USADO)</b>
20	N-V-ART-N	economia(N)-expressando(V)-uma(ART)-bipolaridade(N)(9) <b>(USADO)</b>
21	N-V-ART-N	tribo(N)-estaria(V)-um(ART)-estágio(N)(6) <b>(USADO)</b>
22	N-VAUX-V-N	branco(N)-são(VAUX)-estariam(V)-patamar(N)(8) <b>(USADO)</b>
23	N-VAUX-V-N	sensu(N)-pode(VAUX)-descolar(V)-debate(N)(4) <b>(USADO)</b>
24	N-ADV-V-N	auto-sustentabilidade(N)-Como(ADV)-viu(V)-introdução(N)(14) <b>(USADO)</b>
25	N-ADV-V-N	sensu(N)-Não(ADV)-descolar(V)-debate(N)(10) <b>(USADO)</b>
26	N-ADV-V-N	visão(N)-extremamente(ADV)-diga-(V)-passagem(N)(5) <b>(USADO)</b>
27	N-VAUX-VAUX-PCP-N	fatores(N)-possam(VAUX)-ser(VAUX)-desenvolvidos(PCP)-atenção(N)(14) <b>(USADO)</b>
28	N-V-ART-ADJ-N	Diferenças(N)-situam-(V)-o(ART)-plano(ADJ)-ídeo-político(N)(5) <b>(USADO)</b>
29	N-ADV-V-V-N	sociedades(N)-não(ADV)-sustentáveis?(V)-são(V)-formas(N)(20) <b>(USADO)</b>
30	N-V-ART-N-PREP-N	sociedades(N)-têm(V)-uma(ART)-história(N)-de(PREP)-reprodução(N)(17) <b>(USADO)</b>
31	N-V-ART-N-PREP-N	pluralismo(N)-envolvendo(V)-a(ART)-tecnologia(N)-de(PREP)-mão(N)(6) <b>(USADO)</b>
32	N-V-ART-N-PREP-N	Congelamento(N)-é(V)-o(ART)-combate(N)-à(PREP)-injustiça(N)(4) <b>(USADO)</b>

Tabela 22. 60% dos predicados não foram encontrados no texto analisado

#	Predicados não encontrados no texto
1	N-PREP-NUM-N
2	N-VAUX-V-PROADJ-N
3	N-PCP-VAUX-V-N
4	N-VAUX-V-V-N
5	N-ADJ-V-N-KC-N
6	N-V-ART-N-PREP-N
7	N-V-ART-ADJ-ADJ-N
8	N-V-ADJ-V-PROADJ-N
9	N-PROADJ-N-VAUX-PCP-N
10	N-PREP-N-V-NUM-N
11	N-V-ART-ADJ-ADJ-N
12	PROADJ-N-V-ART-ADJ-ADJ
13	N-PROADJ-N-VAUX-PCP-V-N

Tabela 23. Todas tuplas adicionais necessárias na construção do MC

#	Regra	Saídas do Text2MARK após tratamento de exceção
1	N-V-N	conceitos(N)-atravessaram(V)-estratégias(N)(6) <b>(USADO)</b>
2	N-V-N	neoliberalismo(N)-passamos(V)-modernização(N)(4) <b>(USADO)</b>
3	N-V-N	ecodesenvolvimento(N)-coloca(V)-limites(N)(4) <b>(USADO)</b>
4	N-PREP-N	rumos(N)-para(PREP)-auto-sustentabilidade(N)(10) <b>(USADO)</b>
5	N-PREP-N	conceito(N)-de(PREP)-desenvolvimento(N)(12) <b>(USADO)</b>
6	N-PREP-N	modernização(N)-para(PREP)-desconstrução(N)(10) <b>(USADO)</b>
7	N-PREP-N	paradigma(N)-de(PREP)-desenvolvimento(N)(10) <b>(USADO)</b>
8	N-PREP-N	Reflexão(N)-sobre(PREP)-especificidades(N)(5) <b>(USADO)</b>
9	N-PREP-N	conhecimento(N)-sobre(PREP)-ecossistemas(N)(6) <b>(USADO)</b>
10	N-PREP-N	estratégias(N)-de(PREP)-auto-sustentação(N)(3) <b>(USADO)</b>
11	N-PREP-N	regulação(N)-sobre(PREP)-mercado(N)(7) <b>(USADO)</b>
12	N-PREP-N	Biosfera(N)-em(PREP)-risco(N)(2) <b>(USADO)</b>
13	N-PREP-N	fato(N)-de(PREP)-insustentabilidade(N)(3) <b>(USADO)</b>
14	N-V-ART-N	neoliberalismo(N)-passamos(V)-a(ART)-modernização(N)(4) <b>(USADO)</b>
15	N-V-ART-N	pluralismo(N)-envolvendo(V)-a(ART)-tecnologia(N)(5) <b>(USADO)</b>

#### 5.1.4.2 As inclusões no CMapTools

O texto apresenta um conceito sobre as relações existentes entre MEIO AMBIENTE e a SOCIEDADE. Há claramente a identificação de vários conceitos no texto que são identificados pelo Text2MARK. Um ponto interessante que vale ressaltar nesta análise é que na Seção escolhida, algumas referências são feitas a seções anteriores do estado-da-arte, mesclados com opiniões do autor do artigo.

Apesar dessa dicotomia entre os momentos do texto, as identificações do Text2MARK se mostraram completas nas relações do tema. As tuplas formadas se complementaram em vários momentos, gerando uma construção diferente, agrupando conceitos de forma binária, terciária e quaternária, onde em vários momentos, apenas o elemento de ligação era necessário.

Há também no texto um conjunto de critérios, identificado com marcadores (bullets), outro ponto desafiador, porém, apenas um dos oito conceitos não foi identificado pelo Text2MARK, mas de forma sem prejudicar a compreensão do texto. Esse conjunto de critérios foi agrupado no lado direito inferior, do total cinco foram agrupados em nós e dois ficaram entre eles.

Este mapa apresenta como característica central a descentralização de palavras-chave, logo, se foca nos conceitos do tema e se apresenta de forma distribuída. Sua leitura apesar de não centralizada, traduz a essência do que o autor do artigo quer passar e nas problemáticas encontradas nas relações entre MEIO AMBIENTE e SOCIEDADE. Nesta seleção, o

Text2MARK identificou 29 das 32 tuplas válidas, aproximadamente 91% e das tuplas complementares 15 de 15 foram usadas, 100% de aproveitamento, conforme Tabela 23.

#### *5.1.4.3 O resultado do Mapa Gerado*

Mapa que teve a menor quantidade de tuplas automáticas devido ao seu escopo, mas que obteve também alto grau de tuplas válidas, conforme pode ser visto na Figura 28.



### 5.1.5 Pedagogia

O último texto escolhido, ANEXO D, tem como autor Gilberto Lacerda Santos; Santos (2011), da área de Educação, Ciências Humanas, da subárea Pedagogia. A Seção do artigo escolhido, de modo aleatório, não tem numeração, com o título: “Recomendações para a formação docente”. O texto selecionado tem 556 palavras.

Tabela 24. Frequência dos léxicos mais relevantes no texto de Pedagogia

#	Frequência dos léxicos	#	Frequência dos léxicos
1	N(166)	13	ADV(15)
2	PREP(93)	14	PROADJ(14)
3	ART(82)	15	VAUX(12)
4	ADJ(44)	16	KS(11)
5	PREP+(41)	17	PRO-KS-REL(10)
6	V(36)	18	NUM(6)
7	KC(22)	19	PROPESS(3)
8	PCP(21)	20	PROSUB(2)
9	NPROP(17)	21	PDEN(1)

#### 5.1.5.1 Os resultados das extrações pelo Text2MARK

O texto analisado apresentou um total de 607 palavras, sendo 233 diferentes entre si. A soma total das tuplas obtidas teve como média o peso de 3,8. Apenas as tuplas que ficaram acima da média foram inicialmente selecionadas. O total destas foi de 165 saídas. Após aplicar o filtro com 6 tuplas de cada predicado, o resultado final foram 44 tuplas, sendo estas escolhidas na ordem decrescente da soma do seu peso, com isso as palavras com maior frequência foram selecionadas, conforme a Tabela 25.

Ressalta-se que nem todos os predicados tiveram resultados, como pode ser visto na Tabela 26. Devido ao processo automatizado, algumas saídas não representaram o contexto escolhido, sendo estas desprezadas na construção do Mapa Conceitual, desta forma, o tratamento de exceção se faz necessário.

Tabela 25. Tuplas selecionadas automaticamente no texto de Pedagogia

#	Regra	Saídas do Text2MARK
1	N-V-N	formação(N)-são(V)-subsídios(N)(11) <b>(USADO)</b>
2	N-V-N	professor(N)-é(V)-ferramenta(N)(10) <b>(USADO)</b>
3	N-V-N	professores(N)-é(V)-escola(N)(10) <b>(USADO)</b>
4	N-V-N	trabalho(N)-decorre(V)-construção(N)(10) <b>(USADO)</b>
5	N-V-N	professor(N)-conceber(V)-material(N)(9) <b>(USADO)</b>
6	N-V-N	cidadania(N)-contribuir(V)-formação(N)(9) <b>(USADO)</b>
7	N-N-N	ferramenta(N)-formação(N)-formação(N)(17) <b>(USADO)</b>
8	N-N-N	formação(N)-professores(N)-voga(N)(15) <b>(USADO)</b>
9	N-N-N	professores(N)-educação(N)-NTICE(N)(13) <b>(USADO)</b>
10	N-N-N	tecnologias(N)-organização(N)-trabalho(N)(13) <b>(USADO)</b>
11	N-N-N	trabalho(N)-Conhecimento(N)-Psicologia(N)(12) <b>(USADO)</b>
12	N-N-N	forma(N)-criatividade(N)-professor(N)(12) <b>(USADO)</b>
13	N-PREP-N	professores(N)-para(PREP)-educação(N)(22) <b>(USADO)</b>
14	N-PREP-N	ferramenta(N)-para(PREP)-formação(N)(21) <b>(USADO)</b>
15	N-PREP-N	professor(N)-para(PREP)-material(N)(19) <b>(USADO)</b>
16	N-PREP-N	educação(N)-para(PREP)-criatividade(N)(18) <b>(USADO)</b>
17	N-PREP-N	educação(N)-como(PREP)-todo(N)(16) <b>(USADO)</b>
18	N-PREP-N	Elementos(N)-para(PREP)-manuseio(N)(15) <b>(USADO)</b>
19	N-KC-PROADJ-N	compreensão(N)-e(KC)-seus(PROADJ)-impactos(N)(5) <b>(USADO)</b>
20	N-KC-PROADJ-N	forma(N)-e(KC)-Tais(PROADJ)-recursos(N)(5) <b>(USADO)</b>
21	N-KC-PROADJ-N	2002(N)-e(KC)-tal(PROADJ)-escolha(N)(4) <b>(USADO)</b>
22	N-KC-PROADJ-N	item(N)-e(KC)-nossa(PROADJ)-meta-investigação(N)(3) <b>(USADO)</b>
23	N-V-ART-N	conhecimento(N)-fosse(V)-uma(ART)-prioridade(N)(14) <b>(USADO)</b>
24	N-V-ART-N	trabalho(N)-decorre(V)-a(ART)-construção(N)(10) <b>(USADO)</b>
25	N-V-ART-N	cidadania(N)-contribuir(V)-a(ART)-formação(N)(9) <b>(USADO)</b>
26	N-V-ART-N	formação(N)-é(V)-um(ART)-instrumento(N)(8) <b>(USADO)</b>
27	N-V-ART-N	novo(N)-pondera(V)-as(ART)-dissertações(N)(6) <b>(USADO)</b>
28	N-V-ART-N	tarefa(N)-possibilitar(V)-a(ART)-aprendizagem(N)(5) <b>(USADO)</b>
29	N-VAUX-V-N	cidadania(N)-poderá(VAUX)-contribuir(V)-formação(N)(13) <b>(USADO)</b>
30	N-VAUX-V-N	informações(N)-precisa(VAUX)-repensar(V)-modo(N)(7) <b>(USADO)</b>
31	N-VAUX-V-N	critérios(N)-devem(VAUX)-pautar(V)-escolha(N)(6) <b>(USADO)</b>
32	N-VAUX-V-N	voga(N)-são(VAUX)-É(V)-consenso(N)(4) <b>(USADO)</b>
33	N-ADV-V-N	educação(N)-não(ADV)-garante(V)-mudanças(N)(11) <b>(USADO)</b>
34	N-ADV-V-N	conhecimento(N)-não(ADV)-fosse(V)-prioridade(N)(11) <b>(USADO)</b>
35	N-ADV-V-N	resposta(N)-não(ADV)-existe(V)-forma(N)(9) <b>(USADO)</b>
36	N-ADV-V-N	escola(N)-tradicionalmente(ADV)-tem(V)-tarefa(N)(6) <b>(USADO)</b>
37	N-ADV-V-N	voga(N)-automaticamente(ADV)-É(V)-consenso(N)(3) <b>(USADO)</b>
38	N-PREP-NUM-N	Recomendações(N)-para(PREP)-dois(NUM)-indicadores(N)(15) <b>(USADO)</b>
39	N-PREP-NUM-N	consideração(N)-de(PREP)-três(NUM)-variáveis(N)(5) <b>(USADO)</b>
40	N-VAUX-V-PROADJ-N	informações(N)-precisa(VAUX)-repensar(V)-seu(PROADJ)-modo(N)(10) <b>(USADO)</b>
41	N-VAUX-VAUX-PCP-N	trabalho(N)-pode(VAUX)-ser(VAUX)-respondida(PCP)-conteúdo(N)(18) <b>(USADO)</b>
42	N-VAUX-VAUX-PCP-N	elementos(N)-podem(VAUX)-ser(VAUX)-identificados(PCP)-integrantes(N)(12) <b>(USADO)</b>
43	N-VAUX-VAUX-PCP-N	escolha(N)-deve(VAUX)-ser(VAUX)-baseada(PCP)-consideração(N)(11) <b>(USADO)</b>
44	N-V-ART-N-PREP-N	tarefa(N)-possibilitar(V)-a(ART)-aprendizagem(N)-de(PREP)-conceitos(N)(6) <b>(USADO)</b>

Tabela 26. 55% dos predicados não foram encontrados no texto analisado

#	Predicados não encontrados no texto
1	N-PCP-VAUX-V-N
2	N-VAUX-V-V-N
3	N-V-ART-ADJ-N
4	N-ADV-V-V-N
5	N-ADJ-V-N-KC-N
6	N-V-ART-ADJ-ADJ-N
7	N-V-ADJ-V-PROADJ-N
8	N-PROADJ-N-VAUX-PCP-N
9	N-PREP-N-V-NUM-N
10	N-V-ART-ADJ-ADJ-N
11	PROADJ-N-V-ART-ADJ-ADJ
12	N-PROADJ-N-VAUX-PCP-V-N

Tabela 27. Tuplas adicionais quando a média da soma deixa de fora termos importantes

#	Regra	Saídas do Text2MARK após tratamento de exceção
1	N-V-N	escola(N)-tem(V)-tarefa(N)(5) <b>(USADO)</b>
2	N-V-N	tempo(N)-instaura-(V)-paradigma(N)(3) <b>(USADO)</b>
3	N-V-N	poder(N)-criar(V)-alternativas(N)(6) <b>(USADO)</b>
4	N-V-N	professores-pesquisadores(N)-adotaram(V)- pressupostos(N)(3) <b>(USADO)</b>
5	N-V-N	educação(N)-garante(V)-mudanças(N)(7) <b>(USADO)</b>
6	N-N-N	aula(N)-reinvenção(N)-materiais(N)(4)
7	N-N-N	critérios(N)-escolha(N)-professor(N)(10)
8	N-N-N	capacidade(N)-discernimento(N)-professor(N)(9) <b>(USADO)</b>
9	N-N-N	conteúdo(N)-investigações(N)-professores- pesquisadores(N)(8)
10	N-N-N	construção(N)-conhecimentos(N)-socialização(N)(9)
11	N-PREP-N	mecanismos(N)-de(PREP)-formação(N)(8) <b>(USADO)</b>
12	N-PREP-N	consenso(N)-em(PREP)-trabalhos(N)(7)
13	N-PREP-N	Estado(N)-de(PREP)-aprendizagem(N)(4) <b>(USADO)</b>
14	N-PREP-N	ferramentas(N)-de(PREP)-desenvolvimento(N)(4) <b>(USADO)</b>
15	N-PREP-N	Conhecimento(N)-em(PREP)-Psicologia(N)(6) <b>(USADO)</b>
16	N-PREP-N	alternativas(N)-para(PREP)-abordagem(N)(13) <b>(USADO)</b>
17	N-PREP-N	utilização(N)-por(PREP)-parte(N)(10) <b>(USADO)</b>
18	N-PREP-N	sala(N)-de(PREP)-aula(N)(2) <b>(USADO)</b>
19	N-PREP-N	capacidade(N)-de(PREP)-discernimento(N)(2) <b>(USADO)</b>
20	N-PREP-N	construção(N)-de(PREP)-conhecimentos(N)(8) <b>(USADO)</b>

### 5.1.5.2 As inclusões no CMapTools

Neste texto destaca-se inicialmente a utilização das 44 tuplas selecionadas automaticamente pelo Text2MARK, ou seja, 100% de aproveitamento. E das tuplas complementares o uso de 15 das 20, totalizando 75% de uso, conforme Tabela 27. Tal resultado reforça a permanência de um predicado inicialmente duvidoso nos critérios de busca, o predicado NOME-NOME-NOME, não há nas regras gramaticais essa construção,

porém, durante o levantamento do estado-da-arte percebeu-se a importância do léxico NOME no contexto da extração automática.

Esse texto também traz seu desenvolvimento em torno de palavras-chave como PROFESSOR e FORMAÇÃO. Logo, os assuntos estão dispostos a mostrar dependências diretas dessas palavras. Não houve preocupação neste mapa em deixar próximos os parágrafos representados graficamente. O intuito é de perceber como a disposição pode influenciar na leitura do mapa e na compreensão do texto. Porém, percebeu-se que não houve muita influência a ausência dessa proximidade.

Algumas palavras também se repetiram e se optou por replicar as mesmas dentro do mapa para facilitar a construção e ligação dos seus elementos. Sempre que foi possível se utilizou o recurso de aproveitamento da palavra para dar sentido a outro trecho do texto. Nesta situação, optou-se por usar a coloração diferente das linhas, já usada anteriormente nos mapas construídos.

Vários parágrafos foram agrupados em nós e alguns parágrafos seguintes foram colocados próximos dos conceitos tratados. Assim como no texto anterior, este texto apresentou uma série de critérios identificados por asterisco, cinco no total, e todos tiveram seus textos extraídos automaticamente pelo Text2MARK.

### *5.1.5.3 O resultado do Mapa Gerado*

O teste final do Text2MARK apresenta na Figura 29 as instruções que devem ser seguidas por professores, segundo a perspectiva da Pedagogia. O mapa conceitual por ser um instrumento pedagógico torna-se, pelo resultado alcançado, uma forma de levar essa informação e compreensão aos docentes.



## 5.2 Os grupos de avaliação dos mapas conceituais

Após a obtenção dos mapas, os mesmos foram avaliados por três grupos distintos de alunos. O primeiro grupo composto por treze alunos de segundo semestre do curso de Letras da Universidade Federal do Amapá - UNIFAP. O segundo grupo foi composto por um mestrando e uma doutora, ambos da Universidade Federal do Pará – UFPA. O terceiro grupo composto de 24 alunos do sétimo semestre do curso de Letras também da UNIFAP.

A avaliação foi feita de forma subjetiva, inicialmente foi feita uma rápida explanação sobre o que eram mapas conceituais e uma breve leitura de apenas um trecho do mapa 01 do texto de geografia. Houve cuidado para não induzir de nenhuma forma os alunos na leitura dos textos.

Os grupos deveriam inicialmente ler os textos e em seguida ler os mapas. Feito isso deveriam atribuir uma nota de zero a dez. Essa nota tem relação de quanto o mapa representa a ideia central do texto lido e não os seus parágrafos.

Como se trata de avaliação subjetiva de grupos que não têm familiaridade com mapas conceituais foram usadas seis métricas estatísticas para melhorar a interpretação dos dados, são elas: i) média aritmética; ii) desvio padrão; iii) distorção; iv) modo; v) mediana e vi) intervalo de confiança.

As métricas i e ii são mais conhecidas, porém se resolveu explicar as demais métricas. A distorção é a caracterização do grau de assimetria da distribuição sobre sua média. O modo retorna o valor mais repetido ou de maior frequência na amostra. A mediana é o número central de um determinado conjunto de números e por fim é o intervalo de confiança que determina uma média da população.

A Tabela 28 mostra a avaliação individual dos alunos do segundo semestre de Letras da UNIFAP para cada mapa conceitual gerado e ao final é demonstrada a distribuição estatística. Todos os alunos submetidos ao teste não tiveram contato anterior com mapas conceituais ou sua construção.

Tabela 28. A avaliação subjetiva feita pelos alunos de letras da UNIFAP

Avaliador	Curso	Sem	Instituição	MC 01	MC 02	MC 03	MC 04	MC 05
1	Letras	2	UNIFAP	9,0	9,0	10,0	10,0	10,0
2	Letras	2	UNIFAP	9,5	9,5	9,5	10,0	10,0
3	Letras	2	UNIFAP	9,0	6,0	8,0	7,0	10,0
4	Letras	2	UNIFAP	9,0	8,0	9,0	8,0	9,0
5	Letras	2	UNIFAP	9,0	8,0	8,0	9,0	8,0
6	Letras	2	UNIFAP	7,0	8,0	8,0	9,0	7,0
7	Letras	2	UNIFAP	9,0	8,0	7,0	5,0	6,0
8	Letras	2	UNIFAP	3,0	7,0			
9	Letras	2	UNIFAP	3,0	7,0	8,0	8,0	8,0
10	Letras	2	UNIFAP	7,5	8,0	8,5	8,5	9,0
11	Letras	2	UNIFAP	7,0	8,0	8,0	7,5	8,0
12	Letras	2	UNIFAP	6,0	9,0	8,0	7,0	8,0
13	Letras	2	UNIFAP	9,0	10,0	7,0	10,0	10,0
Média				7,5	8,1	8,3	8,3	8,6
Desvio padrão				2,2	1,1	0,9	1,5	1,3
Distorção				-1,3	-0,1	0,6	-0,7	-0,5
Modo				9,0	8,0	8,0	10,0	10,0
Mediana				9,0	8,0	8,0	8,3	8,5
Intervalo de confiança				0,4490	0,1871	0,1684	0,2806	0,2619

A Tabela 29 mostra a avaliação feita na UFPa que teve um pequeno grupo e com isso não foi possível determinar algumas métricas estatísticas, porém o resultado apresentado nas avaliações parece consistente com pequenas variações entre os avaliadores.

Tabela 29. A avaliação subjetiva feita na UFPa

Avaliador	Curso	Sem	Instituição	MC 01	MC 02	MC 03	MC 04	MC 05
1	Computação	Mestrando	UFPa	8,0	9,0	8,0	9,0	8,0
2	Computação	Doutor	UFPa	7,0	8,0	7,0	8,0	9,0
Média				7,5	8,5	7,5	8,5	8,5
Desvio padrão				0,7	0,7	0,7	0,7	0,7
Distorção								
Modo								
Mediana				7,5	8,5	7,5	8,5	8,5
Intervalo de confiança				0,1363	0,1363	0,1363	0,1363	0,1363

A Tabela 30 mostra a avaliação do último grupo feita na UNIFAP por alunos do sétimo semestre de letras. Os mesmos não possuíam conhecimento algum sobre mapa conceitual e foi necessária uma breve introdução sobre o tema, seguindo a mesma metodologia aplicada ao primeiro grupo.

Tabela 30. Distribuição mais uniforme e consistente entre os alunos do sétimo semestre

Avaliador	Curso	Sem	Instituição	MC 01	MC 02	MC 03	MC 04	MC 05
1	Letras	7	UNIFAP	10,0	10,0	10,0	10,0	10,0
2	Letras	7	UNIFAP	10,0	9,0	8,0	8,0	8,0
3	Letras	7	UNIFAP	6,0	8,0	9,0	8,0	7,0
4	Letras	7	UNIFAP	9,0	9,0	9,5	8,0	8,0
5	Letras	7	UNIFAP	7,0	9,0	9,0	8,0	10,0
6	Letras	7	UNIFAP	8,0	7,0	5,0	6,0	7,0
7	Letras	7	UNIFAP	10,0	8,0	9,0	9,0	10,0
8	Letras	7	UNIFAP	8,0	9,0	9,0	8,0	7,0
9	Letras	7	UNIFAP	8,5	7,5	8,5	8,5	9,5
10	Letras	7	UNIFAP	7,5	8,5	6,0	7,0	10,0
11	Letras	7	UNIFAP	9,0	8,0	6,0	5,0	8,0
12	Letras	7	UNIFAP	7,0	9,0	7,0	8,0	9,0
13	Letras	7	UNIFAP	8,0	10,0	9,0	9,0	10,0
14	Letras	7	UNIFAP	9,0	8,0	8,0	7,0	9,0
15	Letras	7	UNIFAP	8,0	9,0	8,0	8,0	8,0
16	Letras	7	UNIFAP	9,0	8,0	8,0	8,0	9,0
17	Letras	7	UNIFAP	8,0	9,0	8,0	8,0	8,0
18	Letras	7	UNIFAP	8,0	8,0	9,0	8,0	8,0
19	Letras	7	UNIFAP	9,0	8,0	8,0	8,0	10,0
20	Letras	7	UNIFAP	8,0	8,0	7,0	7,0	8,0
21	Letras	7	UNIFAP	8,0	8,0	9,0	8,0	9,0
22	Letras	7	UNIFAP	8,0	9,0	9,0	8,0	10,0
23	Letras	7	UNIFAP	8,0	9,0	8,0	8,0	9,0
24	Letras	7	UNIFAP	8,0	7,0	8,0	6,0	6,0
			Média	8,3	8,5	8,1	7,8	8,6
			Desvio padrão	1,0	0,8	1,2	1,0	1,2
			Distorção	0,0	0,1	-1,0	-0,8	-0,4
			Modo	8,0	9,0	8,0	8,0	10,0
			Mediana	8,0	8,3	8,0	8,0	9,0
			Intervalo de confiança	0,1377	0,1101	0,1652	0,1377	0,1652

Tabela 31 trata a comparação de todos os resultados obtidos. Na última coluna apresenta uma média aritmética e todas as demais métricas encontradas. Os resultados apresentados reforçam a aplicabilidade do método de extração automático. Observa-se uma variação muito pequena na média das notas subjetivas e nas demais métricas avaliadas.

Tabela 31. Dados comparativos consolidados

Avaliação de alunos segundo semestre						
	MC 01	MC 02	MC 03	MC 04	MC 05	Média
Média	7,5	8,1	8,3	8,3	8,6	8,1
Desvio padrão	2,2	1,1	0,9	1,5	1,3	1,4
Distorção	-1,3	-0,1	0,6	-0,7	-0,5	-0,4
Modo	9,0	8,0	8,0	10,0	10,0	9,0
Mediana	9,0	8,0	8,0	8,3	8,5	8,4
Intervalo de confiança	0,4490	0,1871	0,1684	0,2806	0,2619	0,2694

Avaliação de alunos mestrado/doutorado						
	MC 01	MC 02	MC 03	MC 04	MC 05	Média
Média	7,5	8,5	7,5	8,5	8,5	8,1
Desvio padrão	0,7	0,7	0,7	0,7	0,7	0,7
Distorção						
Modo						
Mediana	7,5	8,5	7,5	8,5	8,5	8,1
Intervalo de confiança	0,1363	0,1363	0,1363	0,1363	0,1363	0,1363

Avaliação de alunos sétimo semestre						
	MC 01	MC 02	MC 03	MC 04	MC 05	Média
Média	8,3	8,5	8,1	7,8	8,6	8,3
Desvio padrão	1,0	0,8	1,2	1,0	1,2	1,0
Distorção	0,0	0,1	-1,0	-0,8	-0,4	-0,4
Modo	8,0	9,0	8,0	8,0	10,0	8,6
Mediana	8,0	8,3	8,0	8,0	9,0	8,3
Intervalo de confiança	0,1377	0,1101	0,1652	0,1377	0,1652	0,1432

Média das avaliações dos Mapas Gerados com saídas do Automap

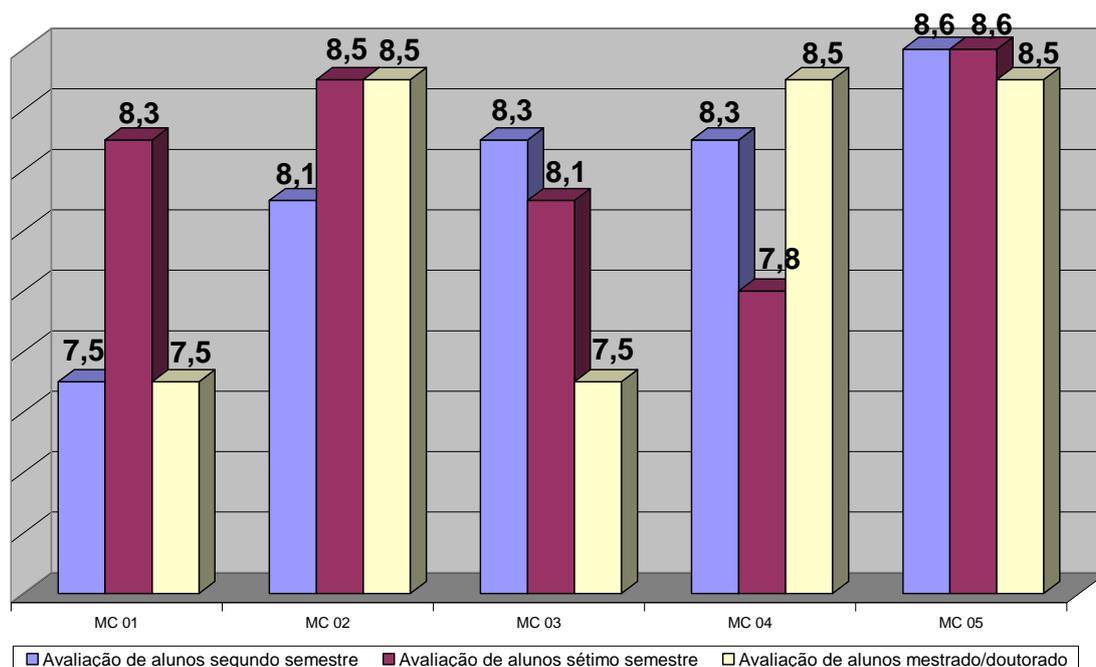


Gráfico 1. Médias obtidas nas avaliações dos mapas gerados com auxílio do Text2MARK

### 5.3 Considerações finais do Capítulo

Na construção dos mapas conceituais, convencionou-se em utilizar setas para indicar a direção de leitura dos conceitos e seus elementos de ligação. Tentou-se manter um padrão de construção onde a leitura se faz de cima para baixo e da esquerda para direita, porém, como a ideia do mapa é apontar para conceitos já existentes tornou-se complicado manter esse padrão de escrita.

A união de conceitos também foi feita através da criação de nós, grandes grupos que concentram parágrafos inteiros ou seções inteiras. Alguns conceitos que estão relacionados a um determinado grupo têm suas ligações feitas nestes grupos, permitindo ao leitor entender que há uma relação direta entre os dois conceitos.

A leitura feita do mapa gerado consegue demonstrar por meio das relações dos conceitos existentes a essência da ideia que se apresenta no texto analisado. Não obstante a isso, a união de conceitos permitido pelo CMapTools auxiliou para destacar aqueles que, segundo nossa análise, são importantes no texto fornecido.

O complemento de elementos léxicos nas tuplas também foi algo positivo na construção das mesmas. Por exemplo, o ARTIGO após o VERBO, completando algumas tuplas no formato NOME-VERBO-NOME. Outro elemento, a PREPOSIÇÃO entre NOMES também auxilia na compreensão da singularidade do texto analisado. Esses são apenas dois de alguns exemplos identificados durante as inclusões.

Algumas construções dos mapas conceituais se mostraram verdadeiros desafios, não somente para a ferramenta, mas para a proposta desta dissertação, por ser a primeira vez que se realizou a construção de mapas distintos a área da computação e de textos que não eram da normalização de banco de dados.

Ressalta-se a notória dificuldade que acadêmicos e não-especialistas em mapas devem ter para encontrar os elementos necessários para a construção dos mesmos. Saber qual elemento textual é importante e em qual contexto ele deve ser inserido ou abstraído da construção é definitivamente um desafio.

Conclui-se que apesar da notação de orientação da leitura utilizada no mapa 02 não ser convencional em mapas conceituais, praticamente todos os acadêmicos de letras acharam o mapa mais fácil de ler. Associa-se a isto a falta de prática no uso dos MCs.

No que envolve a avaliação de desempenho do Text2MARK, os números apresentados podem ser vistos na Tabela 32. A média das tuplas primárias válidas é de 84,2% e das tuplas secundárias válidas são 61,2% de utilização. Pela abordagem da proposta em não limitar a uma forma textual escrita, os resultados apresentados conseguem representar a essência dos textos, como podem ser visto nos mapas. Alguns mapas com mais detalhes do que outros, mas todos capturados na representação gráfica.

Tabela 32. O percentual de tuplas válidas encontradas automaticamente pelo Text2MARK

#	Item teste	Tuplas primárias	Tuplas secundárias
1	Geografia	55%	21%
2	EAD	85%	10%
3	Linguística	90%	100%
4	Meio Ambiente	91%	100%
5	Pedagogia	100%	75%

Sobre esses percentuais é preciso fazer algumas considerações. A razão de haver algumas tuplas que não foram utilizadas é devido a sua estrutura lógica formada pela sequência do predicado, uma vez que escrever na forma preconizada pela linguística requer experiência, ou seja, formava trechos sem nexos ou com segmentos distintos. Assim como, outras retrataram apenas um pequeno trecho do texto e sem tuplas complementares, com isso, para não ter que realizar uma inclusão completa do parágrafo, optou-se por não utilizar tais tuplas.

O Gráfico 1 permitiu mostrar as médias das avaliações, ressaltando que nem sempre os mapas mais completos permitem melhor leitura. De tal forma que os mapas mais simples consigam passar a ideia central do texto proposto, tiveram um aceite visual melhor.

## 6 CONCLUSÕES

O ato de simplificar o processo de extração automatizada para construção de mapas conceituais se mostrou um dos vários pontos de acerto da proposta. Partindo-se do princípio que “simples é melhor”, foi proposta uma metodologia agradável em comparação aos modelos existentes no estado-da-arte. A complexidade e as regras necessárias para o processo automatizado ficou embutido no código da interface e todos seus recursos são executados localmente com baixo custo computacional que obtivemos após 38 versões.

Percebeu-se também a necessidade do tratamento de exceções em todos os casos. Deve-se a isso a regra padrão da exclusão por estarem acima da média associada à frequência das palavras. Com isso, alguns trechos do texto se perdem e precisam ser recuperados pelo usuário no ato da construção do mapa conceitual.

Era esperado durante a construção dos predicados que nem todos fossem utilizados, pois a proposta aborda um tratamento gramatical que permite que qualquer texto livre de qualquer área seja analisado. Isso não significa que o mapa não tenha obtido o sentido na tradução gráfica do texto, pelo contrário, mesmo não sendo especialista no assunto e não tendo escolha direta do mesmo, a compreensão do mapa após sua leitura mostra a viabilidade da proposta.

A proposta também permitiu algo inesperado quando realizado o tratamento de exceção as tuplas complementares aumentaram o detalhe do mapa se tornando uma forma interessante de explorar e refinar ainda mais a representatividade dos mesmos. Isto se torna mais um diferencial para o usuário, pois, o referencial teórico já mostrou que quanto mais especialista no assunto o usuário for mais detalhes ele pode incluir ou omitir no mapa.

Conclui-se que a escolha e construção dos predicados, baseados nas regras gramaticais, associados a todos os critérios definidos na proposta, permitiram que o Text2MARK pudesse alcançar de forma significativa resultados expressivos, representando um texto de livre escrita como um Mapa Conceitual.

Conclui-se, após todas as avaliações, que o Text2MARK se mostrou válido na sua proposta de auxiliar a construção de mapas conceituais com a seleção automática de tuplas baseado em regras híbridas. Assim como permitiu sua utilização em texto de qualquer natureza, independente da área de conhecimento e/ou concentração.

## 6.1 Trabalhos futuros

Buscar um método de melhorar a definição estatística dos léxicos que serão considerados na seleção das tuplas. Percebeu-se que algumas palavras-chave que ocorriam apenas uma vez no texto não foram incluídas, fazendo-se necessário sua inclusão manualmente.

Fomentar a utilização do Text2MARK na análise de requisitos para construção gráfica de bancos de dados através de textos normalizados ou não. Uma vez que o escopo inicial foi de 100% de utilização das tuplas encontradas para construir uma representação de um diagrama entidade-relacionamento.

Desenvolver um método para determinação de sinônimos com o intuito de reduzir as palavras existentes no texto original e também para reduzir sua necessidade na representação gráfica, sem perder a essência do texto. Associados a isso, utilizar um parser em Prolog para revalidar os encadeamentos gramaticais criados nas regras dos predicados.

Adaptar para que o Text2MARK possa ser utilizado em idiomas diferentes do português, tornando-se uma ferramenta de auxílio multilíngue na construção de mapas conceituais. Assim como buscar uma possível integração direta do CMapTools para gerar automaticamente os mapas conceituais a partir da seleção das tuplas.

## REFERÊNCIAS BIBLIOGRÁFICAS

AL-KAMHA, R. Conceptual XML for Systems Analysis. PhD dissertation, Brigham Young University, Department of Computer Science, June 2007.

ALVES, A. O.; Pereira, F. C., & Cardoso, A. (2001). Automatic Reading and Learning from Text. In: Proceedings of the International Symposium on Artificial Intelligence (ISAI'2001), (pp. 302-310). Fort Panhala (Kolhapur), India.

ARAÚJO, Ana Marina Teixeira, Crediné Silva de Menezes, Davidson Cury (2002) “Um Ambiente Integrado para Apoiar a Avaliação da Aprendizagem Baseado em Mapas Conceituais”. Departamento de Informática – Universidade Federal do Espírito Santo (UFES). XIII Simpósio Brasileiro de Informática na Educação – SBIE – UNISINOS 2002.

ARLOTTA L., L; CRESCENZI, V., MECCA, G; MERALDO, P. Automatic annotation of data extracted from large web sites. In Proceedings of the Sixth International Workshop on the Web and Databases (WebDB 2003), pages 7–12, San Diego, California, June 2003.

AUSUBEL, D.P. (2003). Aquisição e Retenção de Conhecimentos: Uma Perspectiva Cognitiva. Lisboa: Plátano Edições Técnicas.

AUSUBEL, D.P., NOVAK, J.D e HANESIAN, H. Educational psychology: a cognitive view. 2nd Ed. New York: Hold Rinehart and Winston, 1978. Trad. p/ português de Eva Nick et al. Psicologia educacional. Rio de Janeiro: Interamericano, 1980.

BAADER, F.; NUTT, W. Basic description logics. In F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors, The Description Logic Handbook, chapter 2, pages 43–95. Cambridge University Press, Cambridge, UK, 2003.

BAI, S.-M., & Chen, S.-M. (2008). Automatically constructing concept maps based on fuzzy rules for adapting learning systems. In: Expert Systems with Applications, 35 (1), 41-49.

BOGDEN, Christopher A. The use of concept mapping as a possible strategy for instructional design and evaluation in college genetics. M.Sc. dissertation, Cornell University, Ithaca, N.Y., 1977.

BUITELAAR, P.; OLEJNIK, D.; SINTEK, M. OntoLT: A Protege plug-in for ontology extraction from text based on linguistic analysis. In Proceedings of the First European Semantic Web Symposium (ESWS'04), pages 31–44, Heraklion, Greece, May 2004.

CABRAL, Anderson Ricardo Yanzer; OLIVEIRA, Taiana Rosales de. Como criar mapas conceituais utilizando o CmapTools. Versão 3.x. Universidade Luterana do Brasil – ULBRA Guaíba. Guaíba, 2003.

CIMIANO, P.; VOLKER, J.. Text2Onto—a framework for ontology learning and data-driven change discovery. In Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB'05), pages 227–238, Alicante, Spain, June 2005.

COHN, Mike. Desenvolvimento de software com Scrum – Aplicando Métodos Ágeis com Sucesso. Ed. Bookman. 2011. 496 p.

CLARIANA, R. B., & Koul, R. (2004). 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.

CLARIANA, R. B., & Koul, R. (2004). 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.

CHANG, T.-H., Tam, H.-P., Lee, C.-H., & Sung, Y.-T. (2008). Automatic Concept Map Constructing using top-specific training corpus. In: Proceedings of the Asia-Pacific Educational Research Association Board Meeting (APER'A'2008). Singapore.

CHEN, N.-S., Kinshuk, Wei, C.-W., & Chen, H.-J. (2008). Mining e-Learning domain concept map from academic articles. In: Computers & Education, 50 (3), 1009-1021.

CHOMSKY, Noam. (1965) “Aspects of the Theory of syntax”. The MIT PRESS. Cambridge, Massachusetts. Coleção STVDIVM. Temas filosóficos, jurídicos e sociais.

DALMOLIN, Luiz Claudio Duarte, Silvia Modesto Nassar, Rogério Cid Bastos (2009a) “Extrator de Conceitos e Termos Conectores para Criação de Cursos Baseados em Mapas Conceituais”. Instituto de Informática e Estatística – Universidade Federal de Santa Catarina (UFSC)

DALMOLIN, Luiz Claudio Duarte, Silvia Modesto Nassar, Rogério Cid Bastos, Gustavo Pereira Mateus. (2009b) “A Concept Map Extractor tool for Teaching and Learning. Federal University of Santa Catarina (UFSC). 2009 Ninth IEEE International Conference on Advanced Learning Technologies

DALMOLIN, Luiz Claudio Duarte, Silvia Modesto Nassar, Rogério Cid Bastos (2008) “Uma ferramenta de mineração de documentos para a construção de mapas conceituais”. Instituto de Informática e Estatística – Universidade Federal de Santa Catarina (UFSC). V Congresso Brasileiro de Ensino Superior a Distância, Gramado. (2008).

DOMINGUES, Miriam L., Eloi L. Favero, Ivo P. Medeiros (2007) “Etiquetagem de Palavras para o Português do Brasil”. TIL – V Workshop em tecnologia da informação e da linguagem humana. Rio de Janeiro (2007).

DOMINGUES, Miriam L. 2011.140f. Abordagem para o desenvolvimento de um etiquetador de alta acurácia para o português do Brasil. Originalmente apresentado como Tese de Doutorado. Universidade Federal do Pará, Programa de Pós-Graduação em Engenharia Elétrica, Belém-PA 2011.

ELIASQUEVICI, Marianne K. JÚNIOR, Arnaldo Corrêa Prado. “O papel da incerteza no planejamento de sistemas de educação a distância”. Educação e Pesquisa, São Paulo, v.34, n.2, p. 309-325, maio/ago. 2008 311.

IHMC CmapTools – Institute for Human and Machine Cognition (IHMC), disponível em (<http://cmap.ihmc.us/>) acessado em 018 de janeiro 2013.

GAINES, B. R., & Shaw, M. L. (1994). 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.

GOULI, Evangelia, Agoritsa Gogoulou, Kyparisia Papanikolaou & Maria Grigoriadou (2004) “COMPASS: AN ADAPTIVE WEB-BASED CONCEPT MAP ASSESSMENT TOOL”. Concept Maps: Theory, Methodology, Technology. Proc. of the First Int. Conference on Concept Mapping. A. J. Cañas, J. D. Novak, F. M. González, Eds. Pamplona, Spain 2004

GOMES, Patrícia Silva. Entre o planejamento “estratégico” e a regulação urbanístico-ambiental: um olhar sobre a expansão urbana recente no Vetor Norte da metrópole belo-horizontina. ISSN 0103-8427 Caderno de Geografia, v.21, n.36, 2011.

GRAUDINA, V., & Grundspenkis, J. (2008). 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.

GRILLO, Marlene. LIMA, Valdez Marina do Rosário (2008). “Mapas conceituais e sua utilização na Educação”.

HARTFIELD, Bradley (1991) “The Relation of Design to Human Concerns”. Apple Computer. CH2961-1/91/0000/0494\$01.00 1991 IEEE.

HEUSER, Carlos Alberto. Projeto de Banco de Dados. 4ª Edição, número 4. Porto Alegre: Sagra Luzzatto, 2001. Cap. 3 p. 69-70 e Cap. 7 p. 172-174.

HODGES, Andrew, 1997. Turing. A natural Philosopher. Phoenix, Orion Publishing Group Ltd.

JEAN-MARYA, Yves R; SHIRONOSHITAA, E. Patrick ; KABUKA, Mansur R. Ontology matching with semantic verification. University of Miami, Coral Gables, FL 33124, USA. Abril 2009.

KIM, Jung-Min; CHOI, Byoung-II; SHIN, Hyo-Phil ; KIM, Hyoung-Joo. A methodology for constructing of philosophy ontology based on philosophical texts. Computer Standards & Interfaces 29 (2007) 302–315.

KORNILAKIS, Harry, Kyparisia A.Papanikolaou, Evangelia Gouli, Maria Grigoriadou (2004a) “Using Natural Language Generation to Support Interactive Concept Mapping”. Department of Informatics & Telecommunications, University of Athens.

KORNILAKIS Harry, Maria Grigoriadou, Kyparisia A.Papanikolaou, Evangelia Gouli (2004b) “Using WordNet to Support Interactive Concept Map Construction”. Department of Informatics & Telecommunications, University of Athens, Greece

KOWATA, Juliana H., CURY, Davidson, BOERES, Maria Claudia Silva (2009) “Caracterização das Abordagens para Construção (Semi) Automática de Mapas Conceituais”. Universidade Federal do Espírito Santo (UFES). XX Simpósio Brasileiro de Informática na Educação (2009)

KOWATA, Juliana H., CURY, Davidson, BOERES, Maria Claudia Silva. (2012). Em direção à construção automática de Mapas Conceituais a partir de textos. Revista Brasileira de Informática na Educação, Volume 20, Número 1, 2012.

KUMAZAWA, T., Saito, O., Kozaki, K., Matsui, T., & Mizoguchi, R. (2009). Toward knowledge structuring of sustainability science based on ontology engineering. In: Sustainability Science, 99-116.

LAU, R. Y., Chung, A. Y., Song, D., & Huang, Q. (2008). 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.

LEE, C.-H., Lee, G.-G., & Leu, Y. (2009). Application of automatically constructed concept map of learning to conceptual diagnosis of elearning. In: *Expert Systems with Applications*, 36 (2), 1675-1684.

LINGUATECA. CETENFolha. In: LINGUATECA. 2013. Disponível em: <<http://www.linguateca.pt/CETENFolha>>. Acesso em: 10 Maio. 2013.

MANZANO-MACHO, D., & Gómez-Pérez, A. (2005). An overview of methods and tools for ontology learning from texts. In: *The Knowledge Engineering Review*, 19 (3), 187-212.

MICHELSON, M.; KNOBLOCK, C.A. Unsupervised information extraction from unstructured, ungrammatical data sources on the world wide web. *International Journal of Document Analysis and Recognition*, 10(3-4):211-226, 2007.

MOREIRA, Marco Antonio (2006) “MAPAS CONCEITUAIS (Concept maps)”. Instituto de Física, UFRGS. Trabalho utilizado em um "workshop" sobre mapas conceituais oferecido no Segundo Congresso Internacional sobre Investigação em Didática das Ciências & das Matemáticas, Valência, Espanha, 23 a 25 de setembro de 1987. Adaptado de uma conferência proferida na Terceira Reunião Nacional de Educação em Física, Córdoba, Argentina, 5 a 8 de outubro de 1983. Publicado em CONTACTOS, México, 3(2):38-57, 1988, em Monografias do Grupo de Ensino, Série Enfoques Didáticos, N° 2, 1991 e na Série Textos de Apoio ao Professor de Física, N° 3, 1992. Revisado, atualizado e ampliado em 2006.

MOREIRA, Marco Antonio (1997) “MAPAS CONCEITUAIS E APRENDIZAGEM SIGNIFICATIVA” Instituto de Física – UFRGS. Adaptado e atualizado, em 1997, de um trabalho com o mesmo título publicado em O ENSINO, Revista Galáico Portuguesa de Sócio-Pedagogia e Sócio-Linguística, Pontevedra/Galícia/Espanha e Braga/Portugal, N° 23 a 28: 87-95, 1988.

MOREIRA, Marco A. e BUCHWEITZ, Bernardo. *Novas estratégias de ensino e aprendizagem: os mapas conceituais e o Vê epistemológico*. Lisboa: Plátano Edições Técnicas, 1993. 114p.

MOREIRA, Marco A. y NOVAK, Joseph D. Investigación en enseñanza de las ciencias en La Universidad de Cornell: sistemas de referencia teóricos, cuestiones foco y abordajes metodológicos. Trabajo presentado em El II Congreso Internacional sobre Investigación em la Didáctica de las Ciencias y de las Matemáticas, Valência, España, Septiembre de 1987. *Enseñanza de las Ciencias*, 6 (1):3-18,1988.

MOREIRA, Marco A. Uma abordagem cognitivista ao ensino da Física: a teoria de Ausubel como sistema de referência para a organização do ensino de ciências. Porto Alegre: Editora da Universidade, 1983.

MOREIRA, Marco Antonio; MASINI, Elcie Salzano. Aprendizagem significativa: a teoria de David Ausubel. São Paulo: Moraes, 1982.

MOREIRA, Marco A. Concept maps as tool for teaching. *Journal of College Science Teaching*, Washington, 8 (5): 283-86, 1979.

MOREIRA, Marco A. An Ausubelian approach to physics instructions: an experiment in an introductory college course in electromagnetism. Ph.D. thesis, Cornell University, Ithaca, N.Y., 1977.

MPS.BR: SOFTEX: MPS.BR – Guia Geral, Versão 2011, MPS.BR:2011. [Online]. <http://www.softex.br/mpsbr> acessado em 02/04/2012.

NOVAK, Joseph D. Conocimiento y aprendizaje: los mapas conceptuales como herramientas facilitadoras para escuelas y empresas. Editora: Alianza Editorial - Ano de publicação: (1998) - País: España - ISBN: 84-206-2901-4

NOVAK, Joseph D. and GOWIN, D. B. Learning how to learn New York: Cambridge University Press, 1984. Trad. p/ português por Carla Valadares, Aprender a aprender. Lisboa: Plátano Edições Técnicas, 1996.

NOVAK, Joseph D. A Theory of education Thaca, N.Y.: Cornell University Press, 1977. Trad. p/português de M.A.Moreira, Uma teoria de educação. São Paulo, Pioneira, 1981.

PALMEIRA, Clay, Rafael Chaves, Hamilton Cavalcante, Eloi Favero (2012). A Requirements Elicitation and Analysis Aided by Text Mining. *IJCSNS– International Journal of Computer Science and Network Security*. Vol. 12 No. 6 pp. 122-128. Published on-line on [http://paper.ijcsns.org/07\\_book/201206/20120616.pdf](http://paper.ijcsns.org/07_book/201206/20120616.pdf)

PARK A, Jinsoo; CHO B, Wonchin ; RHO, Sangkyu Evaluating ontology extraction tools using a comprehensive evaluation framework. *Data & Knowledge Engineering* 69 (2010) 1043–1061.

PÉREZ, Cláudia Camerini Corrêa, Renata Vieira (2005) “Mapas Conceituais: geração e avaliação”. Programa Interdisciplinar de Pós-Graduação em Computação Aplicada (PIPCA) – Universidade do Vale do Rio dos Sinos (UNISINOS). Anais do III Workshop em Tecnologia

da Informação e da Linguagem Humana (TIL 2005) em XXV Congresso da Sociedade Brasileira de Computação. São Leopoldo, RS.

PÉREZ, C. C., & Vieira, R. (2004). 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, MA

PERINI, Mário A. (1995) “Gramática Descritiva do Português”. ISBN 85 08 05550 1. Editora Ática S.A.

RICHARDSON, W. R., & Fox, E. A. (2007). Using Concept Maps in NDLTD as a Cross-Language. In: 10th International Symposium on Electronic Theses and Dissertations (ETD 2007). Uppsala, Sweden.

RICHARDSON, W. R., Srinivasan, V., & Fox, E. A. (2008). Knowledge discovery in digital libraries of electronic theses and dissertations: an NDLTD case study. In: International Journal on Digital Libraries , 9 (2), 163-171.

ROBINSON, Genessa, Marcelo Scopel, Leticia Rafaela Rheinheimer, Junior Martins, Sérgio Crespo C S Pinto, Lúcia M. M. Giraffa, Crediné Menezes (2004). “Modelando Requisitos Especificados com Mapas conceituas através da UML-MC”. XV Simpósio Brasileiro de Informática na Educação – SBIE – UFAM 2004.

ROLIM, Luiz Henrique Martins Lins(2006) “Utilização de Mapas Conceituais em Engenharia de Software: Projetando uma ferramenta case”. Anais do 12º Encontro de Iniciação Científica e Pós-Graduação do ITA – XII ENCITA / 2006 Instituto Tecnológico de Aeronáutica, São José dos Campos, SP, Brasil, Outubro, 16 a 19, 2006.

ROSSON, Mary Beth, Hansa Sinha, Mithu Bhattacharya and Dejin Zhao (2007) “Design Planning in End-User Web Development”. The Pennsylvania State University. 2007 IEEE Symposium on Visual Languages and Human-Centric Computing

SANTOS, Gilberto L. Uma Pesquisa Longitudinal sobre Professores e Computadores. Educ. Real., Porto Alegre, v. 36, n. 3, p. 837-848, set./dez. 2011. Disponível em: <[http://www.ufrgs.br/edu\\_realidade](http://www.ufrgs.br/edu_realidade)>.

SARAWAGI, S. Information extraction. Foundations and Trends in Databases, 1(3):261–377, 2008.

SAYÃO, Miriam, Arndt Von Staa, Julio Cesar Sampaio do Prado Leite (2003). “Qualidade em requisitos” Monografias em Ciência da Computação no 47/03. ISSN 0103-9741.

Departamento de Informática. Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ). Outubro (2003).

SCHMID, H.. Improvements in part-of-speech tagging with an application to German, In: ACL SIGDAT-WORKSHOP. Proceedings... [S.l. : S. n.], March, 1995.

STEWART, James; VAN KIRK, Judy and ROWELL, Richard M. Concept maps: a tool for use in biology teaching. *The American Biology Teacher*. 41 (3): 171-75. 1979.

SILVA, Vivian dos Santos; SILVA, João C.P. da Silva; CAMPOS, Maria Luiza M. Alinhamento de Ontologias através do algoritmo de alinhamento local de caminhos. SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL. Niterói, Rio de Janeiro. Universidade Federal Fluminense. Agosto 2009

SILVA, Rodrigo Cezario da (2011) “Uma abordagem para o reuso de requisitos baseada em padrões e rastreabilidade” Dissertação de mestrado em computação aplicada. Univali (2011).

SIMPER, Elena, Reusing ontologies on the Semantic Web: A feasibility study. *Data & Knowledge Engineering* 68 (2009) 905–925.

SOMMERVILLE, Ian. Engenharia de Software – 9a edição. Ed. Pearson Education –BR, 2011. 568 p

SOUSA, J.M.C.; MACHADO, I.L. A função transgressiva dos múltiplos sujeitos nos gêneros discursivos /The transgressive function of multiple subjects in speech genres. A função transgressiva dos múltiplos sujeitos nos gêneros discursivos. *Bakhtiniana*, São Paulo, v. 1, n.6, p. 111-128, 2º semestre 2011.

STRAVRIANOU, Anna, Periklis Andritsos, Nicolas Nicoloyannis (2007) “Overview and Semantic Issues of Text Mining”. *SIGMOD Record*, September 2007 (Vol. 36, No. 3)

TAVARES, Romero (2007) “Aprendizagem significativa em um ambiente multimídia”. Departamento de Física – UFPB. *Indivisa, Boletín de Estudios e Investigación*, 2007, Monografía VIII, pp. 551-561. ISSN: 1579-3141

TAO, C.; EMBLEY, D.W. Automatic hidden-web table interpretation, conceptualization, and semantic annotation. *Data & Knowledge Engineering*, 2009. in press.

TEIXEIRA, Joaquina B. Desenvolvimento auto-sustentado e meio ambiente no universo temático e tático-operativo do serviço social. Original completo elaborado no ano 2000, para

o Módulo 4 do Curso de Especialização à Distância promovido pelo CFESS – foi reduzido para ajustar-se ao número de páginas determinado para a publicação

T.H. Tse, L. Pong (1991) “An examination of requirements specification languages” *The Computer Journal* (1991) 34 (2): 143-152. doi: 10.1093/comjnl/34.2.143

TSENG, S.-S., Sue, P.-C., Su, J.-M., Weng, J.-F., & Tsai, W.-N. (2007). A new approach for constructing the concept map. In: *Computers & Education* , 49 (3), 691-707.

VALERIO, A., & Leake, D. (2006). 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.

WANG, Y.; VOLKER, J.; HAASE, P. Towards semi-automatic ontology building supported by large-scale knowledge acquisition. In *AAAI Fall Symposium On Semantic Web for Collaborative Knowledge Acquisition*, volume FS-06-06, pages 70–77, Arlington, Virginia, October 2006.

YAMASAKI. Kazuya, Hiroyuki FUKUDA, Tsukasa HIRASHIMA, & Hideo FUNAOI (2010) “Kit-Build Concept Map and Its Preliminary Evaluation”. Graduate school of Engineering, Hiroshima University. S. L. Wong et al. (Eds.) (2010). *Proceedings of the 18th International Conference on Computers in Education*. Putrajaya, Malaysia: Asia-Pacific Society for Computers in Education.

ZOUAQ, A., & Nkambou, R. (2008). Building Domain Ontologies from Text for Educational Purposes. In: *IEEE Transactions on Learning Technologies* , Volume 1 (1), p. 49-62.

ZOUAQ, A., & Nkambou, R. (2009). Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project. In: *IEEE Transactions on Knowledge and Data Engineering* (10.1109/TKDE.2009.25).

## **ANEXO A – TRECHO CORRESPONDENTE A SEÇÃO 6.1.1**

### **Geografia (A reestruturação produtiva e os novos papéis assumidos pelo planejamento e regulação urbanos)**

As mudanças estruturais derivadas da crise do modelo fordista-keynesiano e emergência da globalização econômica, fundada na mobilidade geográfica das firmas multinacionais e do capital financeiro, trouxeram várias implicações às economias nacionais (a brasileira, em particular) e às cidades e regiões.

A reestruturação produtiva resulta em toda uma mudança na divisão internacional do trabalho com a emergência de novos países industrializados e processos de desindustrializações (de espaços da indústria fordista) e reindustrializações (compatíveis com a industrialização pós-fordista) nas cidades e regiões, especialmente dos países centrais.

Nas palavras de Diniz e Crocco (1996) ao longo das três últimas décadas os padrões locais da indústria começaram a ser fortemente alterados. Em primeiro lugar, as velhas regiões industriais do nordeste dos Estados Unidos e do norte da Inglaterra entraram em crise, inaugurando o chamado processo de desindustrialização (Bluestone, Harrison, 1982; Massey, Meegan, 1982), seguido pela emergência de novas áreas industriais no oeste, montanhas e sul dos EUA (Markusen, 1985), e no sul da Inglaterra. Em segundo lugar, passou a ocorrer mudanças na divisão internacional do trabalho com emergência dos NICs (newly industrialized countries) e de zonas de processamento de exportação em países não industrializados (Frobel et. al., 1997; Balassa, 1981, Lipietz, Laborgne, 1984). Em terceiro lugar, a crise do chamado padrão fordista de produção, as mudanças na organização dos processos de trabalho e as alterações tecnológicas (lideradas pela microeletrônica e pelas tecnologias de informação) levaram à interpretação de que estas foram levadas ao pós-fordismo. Em quarto lugar, tanto o aumento dos fluxos internacionais e a tendência à globalização, quanto a sua reação, através da crescente integração inter-blocos, e seus efeitos

na mobilidade dos fatores, têm despertado a atenção dos teóricos do comércio internacional sobre a problemática locacional (DINIZ; CROCCO, 1996, p.78).

Essas mudanças implicam no rearranjo do processo produtivo, fazendo com que os princípios de integração vertical e divisão técnica do trabalho da grande empresa fordista, fossem substituídos por uma estrutura de organização das cadeias produtivas mais horizontalizada e flexível (HARVEY, 1992). Todo esse rearranjo é interligado pela emergência de um novo padrão tecnológico, liderado pelo paradigma da microeletrônica, informática e telecomunicações, capazes de conectar os espaços-tempos de produção e consumo, as novas fronteiras industriais e os centros de comando e controle.

A acumulação flexível faz surgir também um amplo leque de serviços avançados, criando novos nichos de mercado altamente especializados. Esses serviços incluem consultorias, financeiras, assistência jurídica, publicidade, contabilidade, projetos, seguros, distribuição de vendas por atacado, serviços de limpeza, segurança, bancários, imobiliários, produção de pesquisa e inovação, etc. que se destinam a suprir a indústria pós-fordista ligando o circuito terciário superior (HARVEY, 1992).

Na acumulação flexível o capital financeiro passa a assumir o poder coordenador da economia mundial em detrimento ao capital corporativo (produtivo), apesar de não suplantá-lo na realização da mais-valia (SOJA, 1993). Assim, o capital financeiro circula em nível nacional, e, sobretudo, transnacional a procura de investimentos lucrativos.

Todo esse processo derivado da globalização econômica se apóia nas grandes cidades como articuladoras da produção, fornecedoras de serviços e aceleradoras da circulação do capital e do consumo, tendo em certas cidades seletas (as cidades globais) os pontos nodais de articulação global da economia.

Sassen (1991) foi quem primeiro cunhou o termo “cidades globais” para referir-se aos pontos nodais dos fluxos financeiros, a partir dos quais se obtém o controle global dos mercados financeiros secundários e sítios de produção dispersos, dado que o investimento financeiro direto ocorre cada vez mais através do mercado de ações e de títulos (SASSEN, 1991).

Sassen (1991) identifica Nova Iorque, Tóquio e Londres como as cidades globais que são os centros de comando e controle dos seus respectivos blocos econômicos. Mesmo situadas com papéis secundários nessa rede global, as grandes cidades tornam-se os atores

privilegiados de articulação das economias regionais ou mesmo nacionais aos espaços-tempos mundiais.

Como destaca Soja (2000), Sassen (1991), Monte-Mór (2006) e Monte-Mór e Roriz (2010), além de outros autores, a reestruturação produtiva e espacial da metrópole contemporânea resulta em várias transformações aos espaços intra-urbanos e urbano-regionais, influenciando-os e articulando-os em diversas escalas.

O termo “cidade-região” tem sido proposto por alguns autores (MAGALHÃES, 2008; MONTE-MÓR, 2006) para descrever a reestruturação do tecido urbano das metrópoles industrializadas, implodidas e explodidas em contextos históricos anteriores (conforme síntese lefebrviana) e que se prenuncia como explicação para os processos urbano-regionais resultantes da reestruturação produtiva (MAGALHÃES, 2008) e da “urbanização extensiva” (MONTE-MÓR, 1994). Magalhães (2008) destaca que a cidaderegião incorpora os “entornos metropolitanos” que passam a receber indústrias que se deslocam à procura de custos mais baixos, mas dentro ainda da zona de influência da metrópole prestadora de serviços que lhe dão suporte.

Magalhães et. al. (2006) colocam que o “entorno metropolitano” se caracteriza pela expansão do MTCI. Segundo os autores, desde a metropolização fordista o *modus vivendi* urbano vem se derramando sobre as regiões circunvizinhas e também sobre os espaços regionais mais distantes. O tecido urbano avançou sobre o antigo espaço rural, redefinindo-o e integrando-o ao sistema urbano-industrial centrado na aglomeração metropolitana e nos centros urbanos principais na sua área de influência.

Dessa forma, as CGPs expandidas para o Brasil nos últimos trinta anos e que levou ao espaço regional o meio técnico-científico (SANTOS, 2005) passa agora a levar o meio técnico-científico-informacional (MTCI) que inclui a rede de telefonia móvel e fixa, redes de informática e internet, etc. No entanto, essas condições, que têm sua origem nos centros urbanos, atingem todo o espaço regional de forma desigual.

Diniz e Crocco (1996) analisam que as transformações econômicas com a cessão do estado desenvolvimentista e emergência da globalização econômica implicaram em várias consequências às políticas regionais. Segundo os autores, a partir da década de 1980, há enfraquecimento do papel do Estado brasileiro em termos de investimento direto, dessa forma as políticas regionais do tipo keynesiana características das décadas anteriores passam a ser substituídas por políticas de competitividade, materializadas, muitas vezes na “guerra fiscal”

assumida entre as cidades e regiões para atrair investimentos dentro do contexto da reestruturação produtiva, apontando que a inserção do Brasil na economia mundial globalizada tende a ser amplamente diferenciada, alimentando as históricas e profundas desigualdades interregionais.

Todo o cenário nos anos 90 de crises e transformações em diversas escalas sugere uma redefinição do papel das cidades, pois, de um lado, há a diminuição dos aportes de recursos em função da crise geral do Estado dos anos 80, e, por outro, há a emergência da globalização econômica (dando maior visibilidade ao local) e a promulgação da Constituição Federal de 1988 (transferindo maiores responsabilidades aos municípios). Com isso, várias municipalidades e também Governos Estaduais passam a mudar a postura na condução das políticas públicas adotando o planejamento estratégico – cuja matriz teórica é o planejamento empresarial sistematizado na Harvard Business School.

O planejamento estratégico modifica os conteúdos do planejamento e dos projetos urbanos. No que se refere ao planejamento, há o abandono da sua idéia como instrumento de ordenamento territorial – baseado no controle do uso e ocupação do solo –, para a idéia de gestão.

Nesse sentido, conforme defendem Borja e Castells (1997) há uma substituição da idéia de “planos que regulam a ação” por “planos de ação”, o que equivale dizer, conforme analisa Souza (2003), que há uma substituição de um “planejamento forte” por um “planejamento fraco”. Esse “plano de ação” direciona-se a criar um “projeto de cidade” capaz de promover as transformações necessárias à atração de investimentos, sintetizada em um “urbanismo de resultados” (OLIVEIRA, 1999). No entanto, o planejamento estratégico não assume a cidade como um todo e sim algumas partes seletas enquanto o “resto” é encarado como “entorno social e ambiental”, através da criação de uma “coesão social” – termo repetido diversas vezes nos textos de Borja e Castells (1997) – e de um “meio ambiente favorável aos negócios” – nesse intuito, os problemas ambientais, em boa medida, frutos da pobreza urbana, são vistos como empecilhos à cidade; contudo, na outra ponta, aparece redefinida a “cidade sustentável”, como uma das roupagens discursivas do planejamento estratégico, na qual as áreas verdes e os recursos hídricos entram como infraestrutura requerida.

No que se refere aos projetos urbanos, o planejamento estratégico se apóia maciçamente nas grandes obras urbanas como forma de aparelhar certos espaços frente às novas CGPs requeridas pelos circuitos globais. Dessa forma, a ação se faz através da dinamização de novas áreas/localizações estratégicas ou através da requalificação ou revitalização de áreas

obsoletas agora refuncionalizadas e revitalizadas (SANCHÉZ, 2001); segundo Sánchez a própria noção de revitalização é representação carregada de valor. Assim, a imagem (síntese) da cidade é passada a partir da reestruturação de certos tecidos urbanos estratégicos, resultando em (lugares das) cidades cada vez mais iguais, porque são moldados por compradores com as mesmas necessidades (VAINER, 2000).

A concorrência interurbana se expressa também na busca pelos municípios de uma flexibilização na regulação urbanística, nesse intuito, vários municípios mudam Leis de Zoneamento para permitir a entrada de determinados empreendimentos. Nessas mudanças espaciais o mercado imobiliário tem uma posição central, e sua ação, acaba por demandar determinados padrões de uso e ocupação do solo aos municípios, destacando a figura do empresário schumpeteriano.

Desse modo, Souza (2003) coloca que o tradicional “planejamento compreensivo” passa a ser desafiado pela direita – através da emergência do planejamento estratégico – e pela esquerda através das tentativas de implementação dos instrumentos da reforma urbana. O que caracteriza uma confluência de paradigmas no bojo do planejamento. Nas palavras de Fix (2004), o planejamento urbano passa a adotar uma formulação híbrida: meio instrumento de captura de mais-valias urbanas, meio instrumento flexível de gestão do mercado; nas palavras de Cota (2010), por sua vez, há um posicionamento entre o direito à cidade e a cidade mercadoria.

## **ANEXO B – TRECHO CORRESPONDENTE A SEÇÃO 6.1.2**

### **EAD (A incerteza numa perspectiva da ciência)**

O pensamento científico passou por várias mudanças paradigmáticas, de forma a tornar possível a compreensão da existência da incerteza e do seu tratamento. Razões para essas mudanças podem ser encontradas nas discussões epistemológicas ocorridas entre os séculos XVIII e XX, a partir do período conhecido como Iluminismo com a construção de idéias que fundamentam transformações sociais, econômicas e políticas relacionadas a uma nova compreensão de mundo.

De acordo com Suassuma (1999), o projeto iluminista foi o ideário da modernidade, época marcada pela revolta contra o poder da Igreja, do rei e da aristocracia e pelo início das discussões dos direitos humanos. Os filósofos iluministas acreditavam que o próprio indivíduo poderia encontrar respostas às suas perguntas, buscando, no início, inspiração na lógica carte-siana segundo a qual o homem deveria estar dividido em corpo e espírito. Somente quando a razão e o conhecimento se fundissem é que a humanidade faria grandes progressos — o desenvolvimento material e moral do homem pelo Conhecimento. A razão era uma dádiva da natureza ao homem.

A objetividade da modernidade se distancia de Descartes com o sistema newtoniano, que rompe com o pensamento de uma natureza oculta e insondável. “A natureza se transforma em um sistema de leis matemáticas estabelecidas por um Deus racional. A razão não se acomoda à contradição, e desta forma, a verdade científica/racional é, em essência, indubitável [...]” (Becker; Gomes, 1993, p.150). A natureza e a sociedade ficam então submetidas aos mesmos mecanismos e às mesmas relações de causa e efeito.

O movimento iluminista trouxe consigo o ideal da investigação sistemática, a matematização da experiência, isto é, a utilização de métodos matemáticos quantitativos que levariam ao conhecimento seguro e certo da realidade.

No século XIX, surge o Positivismo, de Augusto Comte, que tinha como pressupostos a criação de leis universais, a ordem e o progresso. Essa doutrina acreditava ter ultrapassado o estágio do conhecimento teológico e o metafísico por um saber definitivo científico ou positivo. O Positivismo é um procedimento teórico-metodológico de entendimento dos fenômenos sociais sob o mesmo prisma com que são entendidos os fenômenos naturais. É reflexo das novas idéias instituídas pelo Iluminismo que, por sua vez, fundamenta a modernidade. Nessa época, o método quantitativo das ciências naturais era a abordagem adotada para se alcançar o conhecimento objetivo. A incerteza era algo considerado como não científico.

Embora os valores positivistas fossem dominantes ao longo do século XIX e início do XX, seus princípios epistemológicos sempre estiveram sendo questionados. Entretanto, para Van Asselt e Rotmans (2000), a primeira grande crise do Positivismo foi originada na Matemática, quando contradições nos princípios lógicos desta foram reveladas no início do século XX. Einstein teve um papel importante com a proposta de uma nova Física (teoria da relatividade). A emergência da Estatística intensificou o estudo da incerteza. O princípio da incerteza de Heisenberg na década de 30 do século XX foi significativo nesse processo.

Além da Matemática e da Física, manifestações em outras áreas também começavam a traduzir a insatisfação com o modelo positivista da ciência (Biologia: Humberto Maturana e Francisco Varela; Teoria geral dos sistemas: Ludwig von Bertalanffy; Cibernética: Norbert Wiener; etc.). O reflexo desse movimento foi a percepção da ciência como não puramente objetiva e de que o conhecimento não é equivalente à verdade e certeza, considerando a existência de incertezas fundamentais que não podem ser reduzidas por meio de mais pesquisas.

Logo, no dizer de Morin (2001): A MAIOR CONTRIBUIÇÃO de conhecimento do século XX foi o conhecimento dos limites do conhecimento. A maior certeza que nos foi dada é a da indestrutibilidade das incertezas, não somente na ação, mas também no conhecimento. (p. 55)

Cresce a necessidade de uma revisão dos modelos que movem a ciência à medida que a realidade se complica, que as transformações se aceleram, que as incertezas se expandem e que um aumento notável nos riscos perpassa à humanidade.

Uma relevante perspectiva transdisciplinar para o estudo de problemas complexos vem sendo desenvolvida por autores que defendem a necessidade de reestruturação das práticas

científicas, por meio da proposição do que consideram “ciência pós-normal”. Funtowicz e Ravetz (2002), seus idealizadores, afirmam que a ciência pós-normal é uma forma de expressar apropriadas estratégias para solução de problemas, focando nos aspectos que tendem a ser negligenciados pelas tradicionais práticas científicas, quais sejam: existência da incerteza, valores em jogo e uma pluralidade de perspectivas legítimas. O “pósnormal” significa ir além do normal, no sentido de que os procedimentos usuais baseados na ciência “normal” (Kuhn, 1978) não são suficientes, ainda que permaneçam indispensáveis para nortear o processo de tomada de decisão.

## ANEXO C – TRECHO CORRESPONDENTE A SEÇÃO 6.1.3

### Linguística (Transgressão Simples)

A ocorrência da transgressão simples no seio da comunicação tem uma frequência bem acima daquela que podemos perceber ou sentir num dado enunciado oral ou escrito. Assim, desde a troca de palavras e de pequenos referentes dêiticos, sobreposição de pequenos temas ao tema proposto na comunicação e, até mesmo a própria homonímia e a polissemia das palavras, configuram-se, em muitos casos, na ação da transgressão simples do gênero. Podemos perceber na obra dostoiévskiana inúmeros episódios de transgressão simples do gênero agenciados pelo autor ao longo da desenvoltura da narrativa. Por exemplo, no sutil ato falho provocado por Dimítri num diálogo com Kuzmá Kuzmitch ao pronunciar o nome de Grúchenhka:

- O respeitável Kuzmá Kuzmitch provavelmente ouviu falar mais de uma vez de minhas desavenças com meu pai, Fiódor Pávlovitch Karamázov, que me despojou da herança de minha mãe... porque isso é assunto de todas as conversas, as pessoas se metem naquilo que não lhes compete [...] pode também ter sido informado por Grúchenhka [...] perdoe, por Agrafienaó Alieksándrovna [...] pela honradíssima e respeitadíssima Agripina Alieksándrovna. Assim começou Mítia, que se atrapalhou desde as primeiras palavras [...] (Dostoiévski, p. 2001, 376, grifo nosso).

Além do tom irônico presente na cena, é perceptível a transgressão ocorrida na esfera da identidade que o sujeito comunicante propõe como enunciador do seu dizer. O preço por tal deslize foi o de, justamente, não ter êxito em sua negociação de empréstimo. Outro recurso intensamente utilizado por Dostoiévski para provocar o efeito de transgressão simples é o uso da locução adverbial talvez ao longo dos diálogos proferidos na trama. Reportamo-nos a alguns deles:

Liberal dos anos 40 e 50, livre pensador e ateu, teve neste caso uma participação extraordinária, por tédio, talvez, ou para se divertir. Foi tomado subitamente pelo desejo de ver o mosteiro e o 'santo' (p. 40) (fala do narrodor sobre a posição e interesse de Ivan no encontro no mosteiro.)

- Então, matarei. Não vou suportar isso. (Dimítri)

- Irmão quem matará? (Aliócha)

- O velho. Não tocarei nela.

- Irmão, que está dizendo?

- Não sei, não sei... Talvez mate, talvez não mate. [...]

Aliócha, pensativo, foi para a casa de seu pai. (2001, p. 136, grifo nosso).

- Ora, talvez eu não creia em Deus. (Aliócha)

(p. 2001, 234, grifo nosso).

- Talvez não tenha ainda perdoado - disse ela, com ar ameaçador, de olhos baixos, como que falando a si mesma. - Talvez meu coração pense somente em perdoar (Grúchenhka) (2001, p. 360, grifo nosso).

Apontamos para o efeito que tal transgressão provoca nos diálogos. Além de fomentar um estado de dúvida sobre a identidade ou caráter daquele que fala ou de quem se fala entre os personagens, provoca no leitor a sensação de desconstrução pela dúvida, ao longo da trama, acerca das intenções e de quem seria capaz de participar ou praticar um ato tão imoral como o assassinato.

Por conseguinte, observamos que grande parte da ação de transgressão simples ocorre ou tem seu efeito através da instância de identidade, imagens, dos parceiros na comunicação. Por isso, tratarmos do efeito polifônico como algo ou um acontecimento em constante ação no ato de linguagem. Dessa forma, verificamos como Dostoiévski utilizou-se desse recurso quando, ao longo de sua obra, a imagem discursiva do herói se pulveriza, polifonicamente, em inúmeras imagens, no funcionamento dos diversos discursos que povoam a trama. Assim, percebemos, a partir da imagem enunciativa de Aliócha, a ação de transgressão genérica no ato de linguagem pela multiplicidade enunciativa que o herói desempenha na trama: irmão; amigo; messias; puritano; amante; filho; ateu; religioso; revolucionário; noviço e assassino.

Nesse sentido, confirmamos a intensa atividade transgressiva que ocorre no ato comunicacional da qual Dostoiévski, ao utilizar o recurso de desarranjo - transgressão - da unidade do sujeito nos discursos, brinca com o não-um da subjetividade e cria na insuficiência

e impotência da imagem do herói em cada discurso a possibilidade desse herói fazer percurso para a busca de certa verdade e saber que lhe falta.

A dicotomia já está implantada dentro de Aliócha, o assassinato, a acusação e absolvição se organizaram como elementos que fazem parte da estrutura de sujeito ou de sua função como herói. Todo o humano deverá passar por sua sentença para descobrir ou elaborar sua liberdade: “- Tem razão, é impossível decidir antes da sentença. Depois do julgamento, você mesmo verá; haverá em você um homem novo que decidirá” (fala de Aliócha a Dimítri) (Dostoiévski, 2001, p. 584). É necessário o encontro das contradições, a reunião das disjunções, o espírito carnavalesco para a criação de verdade como função de real num saber que liberta e opera sujeito. A imagem do herói é carnavalizada, o herói é satânico; messiânico; é o bobo da corte. Pura ironia que lampeja o efeito de verdade. Na montagem desse herói carnavalizado, que tem a transgressão por missão, Dostoiévski transgride o imaginário literário da figura do herói como uma pessoa possuída pela bondade; o herói é o assassino do pai, ama o crime, fomentou, e trabalhou, nos diversos diálogos o acontecimento da morte do pai, morte necessária para libertação e acontecimento de seu devir como herói. O assassino é amado e cortejado por todos. É assim que o leitor deverá trabalhar a construção de quem é esse herói, pois, além de não se tratar de um herói escrito e determinado por uma única voz de escritor, o efeito de verdade que será transmitido pelo herói na obra requer um estado de abertura, aceitação e participação (elaboração) da subjetividade da instância leitora.

## **ANEXO D – TRECHO CORRESPONDENTE A SEÇÃO 6.1.4**

### **Meio Ambiente (O conceito de Desenvolvimento e os rumos para a auto-sustentabilidade)**

Como você viu na introdução, é sabido que o conceito de desenvolvimento desponta com ROSTOW e se baseia inicialmente na idéia de que as sociedades evoluem por etapas. Para esta visão (extremamente equivocada, diga-se de passagem), uma tribo indígena estaria num estágio inferior e a sociedade norte-americana no topo.

Esta visão desferiu um golpe fatal na diversidade cultural do planeta, porque todos os outros grupos étnicos (que não o branco ocidental cristão capitalista) são olhados como “atrasados”, como “rudimentares” (estariam no patamar inferior), devendo todos alcançar o modelo da civilização ocidental industrializada de consumo.

É desta matriz teórica que derivam vários termos usuais da economia política, expressando uma bipolaridade simplificadora e discriminatória: desenvolvidos/subdesenvolvidos, centro/periferia, primeiro mundo/terceiro mundo, etc...

Diz-se que Rostow foi ultrapassado na teoria do desenvolvimento desde o período da ascensão da teoria da dependência (esta também considerada obsoleta hoje), mas suas raízes, banidas no plano dos conceitos, atravessaram as estratégias da modernização conservadora imposta por organizações financeiras internacionais (FMI, Banco Mundial) aos governos da América Latina, da África, da Ásia, enfim, aos países portadores de dívida externa até o advento do neoliberalismo, quando passamos da modernização conservadora para a desconstrução destrutiva.

Para romper com esse paradigma de desenvolvimento exógeno (a partir de fora), entendido como “o jeito americano de viver”, impõe-se repensar um movimento endógeno (a partir de dentro) para o horizonte de um desenvolvimento auto-sustentado (claro que sem a negação sectária do intercâmbio e de fatores exógenos), onde possam ser desenvolvidos:

- a atenção e reflexão sobre as especificidades nacionais e regionais, onde o quadro cultural (antes ignorado e até discriminado), ascenda a um valor e a uma relevância sem precedentes;
- o combate ao esfacelamento cultural nativo;
- a negação da uniformização sócio-cultural, com a manutenção e até resgate da diversidade (sem congelamento, é claro);
- o combate à injustiça, à pobreza, à miséria e à exclusão;
- o amplo e aprofundado conhecimento sobre os ecossistemas nacionais e regionais e o estoque de biodiversidade;
- o envolvimento dos cidadãos locais, no planejamento das estratégias de auto-sustentação;
- a operacionalização do pluralismo tecnológico, envolvendo tanto a tradicional tecnologia de mão de obra intensiva (geradora de ocupação e emprego), quanto as de ponta;
- a regulação estatal sobre o mercado predador e aético;

Sem dúvida que foi ampliado o horizonte das pesquisas e propostas no rumo da superação do desenvolvimento capitalista, o qual “pôs em ação um princípio de autodestruição capaz de liquidar o sutil equilíbrio físico-químico e ecológico do planeta e devastar a biosfera, pondo assim em risco a continuidade do experimento da espécie homo sapiens e demens.” (BOFF, 1999: 20)

Embora se reconheçam alguns pontos comuns entre as propostas do ecodesenvolvimento e as do desenvolvimento sustentável, a exemplo: a defesa ao direito das gerações futuras como um princípio básico e a criação de uma sociedade sustentável, é preciso estar atento às suas diferenças, que, segundo Layrargues, situam-se no plano ídeo-político:

“Enquanto o ecodesenvolvimento coloca limites à livre atuação do mercado, o desenvolvimento sustentável afirma que a solução da crise ambiental virá com a instalação do mercado total na economia das sociedades modernas” (LAYRARGUES, op. cit.: 10)

O importante , na verdade, é estar-se atento ao fato de que a insustentabilidade do ambiente e da vida é questão fundamentalmente sócio-política e não ambiental strito sensu.

Não se pode, portanto, descolar o debate da sustentabilidade do campo das lutas sociais e dos projetos societários pretendidos.

Estamos de acordo com Acselrad quando diz que não há sentido em separar-se “meio ambiente” das configurações históricas das sociedades, porque “sustentáveis” ou “não sustentáveis” são as formas de apropriação da natureza para a reprodução social e não o maior ambiente em si.

Um exemplo disso é o fato de que muitas das formas sociais não capitalistas de produção comprovadamente sustentáveis (indígenas, quilombolas), tornaram-se insustentáveis, ao serem destruídas pela formação social capitalista envolvente e dominante.

As sociedades indígenas têm uma história de reprodução sócio-ambiental de mais de 10.000 anos. Bastaram 500 anos de contato com a sociedade capitalista envolvente para conhecerem a devastação, a miséria, a pobreza e a exclusão.

## **ANEXO D – TRECHO CORRESPONDENTE A SEÇÃO 6.1.5**

### **Pedagogia (Recomendações para a formação docente)**

Considerando os dois indicadores anteriores, os mecanismos de formação de professores atualmente em voga são automaticamente questionados. É consenso em todos os trabalhos que a escola, que tradicionalmente tem como tarefa exclusiva possibilitar a aprendizagem de conceitos e informações, precisa repensar seu modo de funcionamento, uma vez que, com a emergência de uma sociedade altamente baseada no trânsito de informações em tempo real, instaura-se um novo paradigma segundo o qual é preciso que se esteja em permanente estado de aprendizagem e de adaptação ao novo, conforme pondera Kenski (2003). Nas seis dissertações, vários elementos podem ser identificados como eventuais integrantes de um currículo de formação de professores para a educação mediada por NTICE. Vejamos os que consideramos principais:

- Elementos formativos para o manuseio de ferramentas de desenvolvimento de materiais didáticos;
- Conteúdos disciplinares referentes à compreensão da Sociedade Informacional e de seus impactos na organização do trabalho pedagógico;
- Conhecimento sólido em Psicologia Cognitiva;
- Conhecimento sólido dos componentes curriculares (conteúdos) a fim de poder criar alternativas para sua abordagem em situação mediada;
- Experiência concreta em aprendizagem mediada por tecnologias.

Com relação às questões de pesquisa anunciadas no item anterior, e diretamente vinculadas à nossa meta-investigação, a primeira delas era sobre os modos de utilização das NTICE, por parte do professor, para conceber seu próprio material didático. As investigações de Feitosa (1999) e Brito (2000) fornecem subsídios para que formulemos a resposta de que não existe forma pré-estabelecida. A criatividade do professor, elemento que tampouco é

explorado em cursos de formação docente, é um instrumento essencial para que seja assegurada a dinamização da sala de aula interativa e a reinvenção dos materiais didáticos, sejam eles convencionais ou inovadores, em função de necessidades específicas de relações educativas específicas.

A segunda questão era sobre os critérios que devem pautar a escolha, pelo professor, de dispositivos tecnológicos de apoio ao seu trabalho pedagógico. As investigações de Neves (2002) e Fonseca (2005) indicam que tal escolha deve ser baseada na consideração de três variáveis: características do conteúdo, características dos alunos, características dos dispositivos. Aí também não existem fórmulas pré-estabelecidas e a capacidade de discernimento do professor é ferramenta essencial, para a qual a formação inicial e a formação continuada são importantes subsídios.

Finalmente, a terceira questão, sobre o impacto das tecnologias educativas na organização do trabalho pedagógico, pode ser respondida a partir do conteúdo das investigações de Iunes (2001) e Zaks (2005). Segundo ambos os professores-pesquisadores, que adotaram pressupostos e contornos distintos, a utilização de recursos tecnológicos na educação não garante mudanças na forma de ensinar e aprender. Tais recursos servem como ferramentas no auxílio da construção de conhecimentos por meio de uma atuação ativa, crítica e criativa por parte de alunos e professores. No entanto, é a escola, entendida como espaço de construção de conhecimentos e de socialização do saber, como um ambiente de discussão, de troca de experiências e de elaboração de uma nova cidadania, que poderá contribuir na formação do indivíduo inserido na Sociedade Informacional, garantindo uma educação voltada para a criatividade, para o prazer, para a autonomia e autorrealização. Todo e qualquer impacto na organização do trabalho pedagógico decorre, portanto, da construção de uma nova concepção da escola, condição ainda trabalhada unicamente em espaços restritos e em determinadas áreas de conhecimento, como se não fosse uma prioridade do campo da educação como um todo.