

Aquisição de Conhecimento de Senso Comum e Inferencialista

Vlória Pinheiro¹, Vasco Furtado¹, Tarcísio Pequeno¹, Wellington Franco¹

¹Universidade de Fortaleza (UNIFOR)

Av. Washington Soares, 1321 – 60.811-905 – Fortaleza – CE – Brasil

{vladiacelia,vasco,tarcisio}@unifor.br, jwellingtonfranco@gmail.com

Abstract. *This paper presents a semi-automated method for the acquisition of common sense and inferentialist concepts in Portuguese. Its innovative feature is a module of reasoning over the existing knowledge that aims to offer original content to the user helping in the expression of semantic relationships and validation of new concepts. This reasoning process is based on heuristics and syntactic analysis of noun phrases. A qualitative evaluation with users who interacted with the system built on the proposed method showed that the interactions made in the process of AC are more productive since the user is reminded about semantic relations and common sense about the new concept.*

Resumo. *Este artigo apresenta um método semi-automático para aquisição de conceitos (AC) de senso comum e inferencialista na língua portuguesa. O método tem como diferencial um módulo de raciocínio sobre conhecimento preexistente que visa oferecer ao usuário conteúdo inicial que o ajude a externar e a validar relações semânticas de novos conceitos. Esse processo de raciocínio se baseia em heurísticas e na análise sintática de sintagmas nominais. Uma avaliação qualitativa com usuários que interagiram com o sistema, construído com base no método proposto, indicou que as interações realizadas no processo de AC são mais produtivas visto que o usuário é lembrado e instigado sobre relações semânticas de senso comum acerca do novo conceito.*

1. Introdução

Conhecimento de senso comum consiste em fatos e conhecimentos espaciais, físicos, sociais, temporais e psicológicos, possuídos pela maioria das pessoas, os quais são frutos da experiência da vida diária [Anacleto et al. 2007] [Liu e Singh 2004]. Muitas vezes, este conhecimento trata-se de um conjunto de suposições implícitas e básicas que suportam e explicam raciocínios necessários para realização de tarefas inteligentes por computadores (p.ex. entendimento de textos em língua natural). Por exemplo, quando alguém fala “Eu comprei doces”, está implícito que usou dinheiro; que o efeito de cair de uma moto é você se machucar; que objetos rolam de superfícies inclinadas; que políticos se envolvem com casos de corrupção e escândalos.

Especialmente na área de Processamento de Linguagem Natural (PLN), há o consenso de que a compreensão de textos por sistemas computacionais depende tanto de conhecimento de mundo como de conhecimento linguístico [Kay 2003, p.19]. No entanto, um dos desafios das pesquisas desta área é a falta de recursos semântico-linguísticos que expressem conhecimento de mundo para suportar tarefas como extração

de informações, recuperação de informação, sistemas de perguntas e respostas, dentre outras. Este desafio é ainda maior quando consideramos a língua portuguesa [Pardo, Caseli e Nunes 2009]. Neste sentido, o recurso InferenceNet [Pinheiro et al. 2010] foi construído contendo relações semânticas de senso comum e inferencialistas sobre conceitos e sentenças, os quais são expressos em língua portuguesa e língua inglesa. As bases semânticas do recurso InferenceNet foram construídas de acordo com o Modelo Semântico Inferencialista (SIM) [Pinheiro 2010] e expressam o caráter pragmático da língua natural através de precondições e conseqüências do uso de conceitos e sentenças.

Assim como acontece com outras bases semânticas, uma dificuldade é garantir uma evolução contínua do recurso linguístico InferenceNet de forma eficaz e com a tempestividade que as aplicações exigem. Métodos de aquisição de conhecimento (AC) automáticos, embora largamente usados em PLN [Che et al. 2009] [Baker et al. 2007], não se mostram eficazes para capturar conhecimento tácito e de senso comum, pois este conhecimento não é comumente derivável das propriedades estruturais e gramaticais de textos disponíveis em *corpus* linguísticos [Kay 2003]. De outro lado, métodos de AC semi-automáticos tradicionais (baseados em modelos), por exemplo os adotados no projeto OMCS [Singh et al. 2002] e OMCS-Br [Anacleto et al. 2006], enfrentam dificuldades em capturar conhecimento de senso comum e pragmático de usuários. Uma dificuldade advém do fato de que as pessoas possuem estes tipos de conhecimento (de senso comum e pragmático), mas não sabem explicitá-lo; o conhecimento está tão arraigado na mente das pessoas que é difícil lembrar e mais ainda externá-lo por meio de relações semânticas estruturadas. Outra questão é que mesmo quando o fazem é difícil garantir a consistência com conceitos já existentes, evitando a duplicação de conteúdo e fortalecendo a conexão da rede semântica.

Neste trabalho, propomos um método de AC semi-automático sobre conceitos de uma língua natural. O diferencial do método proposto é um processo de raciocínio sobre conhecimento preexistente que visa oferecer ao usuário uma base inicial que o ajude a externar e a validar relações semânticas para o novo conceito. O processo de raciocínio se baseia em heurísticas para geração de conteúdo semântico e que são definidas conforme a estrutura gramatical dos nomes dos novos conceitos a serem adquiridos. Por exemplo, para adquirir o conteúdo conceitual de “crime passional” é usada uma estratégia específica para a estrutura gramatical “<nome> <adjetivo>”, a qual aproveita o conhecimento semântico preexistente dos conceitos “crime” e “passional” ou de “crime” e “paixão”. Além disso, o método prevê um processo interativo que privilegia uma melhor precisão na captura e validação das relações semânticas por parte do usuário. O método de AC proposto neste artigo foi implementado e avaliado para a base conceitual bilíngue da InferenceNet.

2. Método de Aquisição de Conhecimento de Senso Comum e Inferencialista

O método de AC proposto consiste em um raciocínio heurístico aplicado sobre conteúdo conceitual preexistente e relacionado, o qual gera uma base inicial de conhecimento para o novo conceito. O método possibilita, adicionalmente, um processo interativo com o usuário que pode incluir novas relações inferenciais e de senso comum e excluir relações propostas, encerrando um mecanismo de validação do valor semântico do novo conceito.

2.1. Heurísticas

As heurísticas são responsáveis pela geração e proposição do conteúdo semântico para a expressão linguística de entrada EXP, a qual nomeia o novo conceito a ser adquirido. De acordo com a estrutura gramatical de EXP, um conjunto de heurísticas busca conteúdo semântico relacionado em uma base de conhecimento preexistente (p.ex.: InferenceNet, OMCS-Br etc.) e gera novas relações semânticas, as quais servem de base para validação pelo usuário e, em seguida, para definição do conteúdo do novo conceito.

Nesse artigo, as heurísticas propostas contemplam apenas os sintagmas nominais por estes serem normalmente usados para denominar as “coisas” do mundo e, por isso, serem primordialmente usados para nomear conceitos da língua natural. A Tabela 1 apresenta as estruturas gramaticais de sintagmas nominais, contempladas pelas heurísticas.

Tabela 1. Principais estruturas gramaticais de sintagmas nominais

Estrutura do Sintagma Nominal EXP	Exemplos
<nome>	vingança, pistolagem
<nome> <adjetivo>	crime passional, impunidade penal
<adjetivo> <nome>	má urbanização
<adjetivo ₁ > <nome> <adjetivo ₂ >	má iluminação pública

- 1) **<nome>** ou **<adjetivo>**— quando EXP não é encontrada na base conceitual, é apresentado para o usuário um conjunto de conceitos preexistentes na base, os quais são: (i) semanticamente relacionados (p.ex. sinônimos); (ii) nomeados com o mesmo radical de <nome> ou <adjetivo>; (iii) nomeados com a forma primitiva de <nome>; (iv) substantivos relacionados a <adjetivo>. Por exemplo, para a expressão linguística “torcedor”, a heurística apresentaria os conceitos “fã”, “torcida” e “torcer”. Para a palavra “passional”, a heurística apresentaria o conceito “paixão”. Em seguida, o usuário seleciona qual, dentre os conceitos apresentados, pode ser usado como base para aquisição do novo conceito. A heurística retorna uma lista de relações semânticas do conceito selecionado, previamente contidas na base.
- 2) **<nome> <adjetivo>** ou **<adjetivo> <nome>** — nestes casos, <nome> é caracterizado por <adjetivo>, indicando-lhe atributo, propriedade, estado, modo de ser ou aspecto. Percebe-se, portanto, um caso de especialização, no qual “<nome> <adjetivo>” ou “<adjetivo> <nome>” expressa uma situação particular ou um tipo de <nome>. Por exemplo, no caso da expressão “crime passional”, o adjetivo “passional” está caracterizando o nome “crime” atribuindo-lhe propriedades relativas à “paixão” e especializando um tipo de “crime”. Esta heurística define os seguintes passos:
 - a. Chamada recursiva à heurística (1) para EXP₁ = <nome> e EXP₂ = <adjetivo>, retornando uma lista de relações semânticas dos conceitos associados a EXP₁ e EXP₂;
 - b. Herança do conteúdo de <nome> para o novo conceito “<nome><adjetivo>” ou “<adjetivo><nome>”, pois em ambos tem-se a expressão de um caso particular ou um tipo de <nome> e, assim sendo, todo o conteúdo de <nome> pode ser transcrito (ou herdado) para “<nome> <adjetivo>” ou

“<adjetivo><nome>”. Por exemplo, a relação inferencial (*capazDe*, ‘crime’, ‘ter vítima’, *Pre*) é transcrita para uma nova relação semântica (*capazDe*, ‘crime passional’, ‘ter vítima’, *Pre*);

- c. Transcrição parcial do conteúdo de <adjetivo> para o novo conceito “<nome><adjetivo>” ou “<adjetivo><nome>”. Neste caso, <adjetivo> está caracterizando <nome> e algumas relações semânticas de <adjetivo> devem ser transcritas para <nome> de forma a atribuir-lhe características ou qualidades. A seguinte metaregra é usada neste passo:

$$\frac{\langle A \rangle \text{ é caracterizado por } \langle B \rangle \ \& \ \langle B \rangle \langle \text{nome_rel} \rangle \ C}{\rightarrow \langle A \text{ caracterizado por } B \rangle \langle \text{nome_rel} \rangle \ C}$$

Para definir quais <nome_rel> tornam esta inferência válida, cada <nome_rel> da base semântica deve ser analisada conforme a natureza da relação semântica. Relações semânticas estruturais (por exemplo, *ehUm*, *feitoDe*, *parteDe*) comumente não devem ser herdadas, pois expressam conteúdo restrito a <adjetivo>. Por exemplo, o fato de que “<paixão> <éUm> <sentimento>” não implica que “<crime passional> <éUm> <sentimento>”. Relações semânticas pragmáticas como relações funcionais, causais, eventuais, motivacionais, comumente suscitam características que são atribuídas de <adjetivo> para <nome>. Por exemplo, o fato de que “<paixão> <feitoDe> <ciúme>” autoriza a geração do conteúdo “<crime passional> <feitoDe> <ciúme>”. Como exemplo, a tabela 2 apresenta os tipos de relações semânticas da InferenceNet definidas para aplicação da metaregra acima. Ao final do processo, a heurística retorna a lista de relações semânticas geradas, as quais foram associadas a “<nome><adjetivo>” ou “<adjetivo><nome>”.

- 3) <adjetivo₁> <nome> <adjetivo₂> — neste caso, o usuário é questionado se <adjetivo₁> está qualificando “<nome> <adjetivo₂>”, por exemplo, como acontece em “má iluminação pública”. Caso o usuário confirme, a heurística (2) é chamada para EXP=“<nome> <adjetivo₂>” e, em seguida, para EXP = “<adjetivo₁> <sn₂>”, com <sn₂> = “<nome><adjetivo₂>”. Caso contrário, a heurística (2) é chamada para EXP=“<adjetivo₁> <nome>” e EXP= “<nome> <adjetivo₂>”. Ao final, a heurística retorna a lista de relações semânticas selecionadas de forma recursiva.

Tabela 2. Tipos de relações semânticas de InferenceNet que serão herdadas de <adjetivo> para <nome><adjetivo> ou <adjetivo><nome>.

Natureza da Relação	Tipo de Relação Semântica	Tipo Relação inferencial
RELATIVA À PROPRIEDADE	PropriedadeDe;	Precondição
RELATIVA À EVENTO	EventoPreRequisitoDe; PrimeiroSubEventoDe;SubEventoDe; UltimoSubEventoDe	Precondição
CAUSAL	EfeitoDe; EfeitoDesejavelDe	Pós-condição
MOTIVACIONAL	MotivacaoDe; DesejoDe	Precondição
FUNCIONAL	UsadoPara; CapazDeReceberAcao	Precondição
RELATIVA À AÇÃO	CapazDe	Precondição

A Figura 1 apresenta o algoritmo que implementa as heurísticas propostas, exemplificando para o novo conceito “crime passional”.

```

relaçõesSemanticas[ ] gerarConteudo (exp)

// 1ª. iteração: exp = "crime passionnal"
// 2ª. iteração: exp = "crime"
// 3ª. iteração: exp = "passional"

Se existeBaseConhecimento(exp) então
  retorna recuperarConteudo(exp);

  // 2ª. iteração: para exp = "crime"
  // recuperada relações semânticas de "crime"
  // Exemplos: capazDe ("crime", "envolver violência", Pre); efeitoDe("crime", "sofrimento", Pos)

senão {
  caso estrutura(exp) = "<nome>" ou "<adjetivo>":
    conceitosRelacionados[ ] = recuperarConceitosRelacionados(exp);
    conceitoRelacionado = selecaoUsuario(conceitosRelacionados[ ]);
    retorna recuperarConteudo(conceitoRelacionado);

    // 3ª. iteração: para exp = "passional"
    // recuperada relações semânticas de "paixão" – conceitoRelacionado selecionado pelo usuário
    // Exemplos: eventoPreRequisitoDe("paixão", "amante", Pre); efeitoDe("paixão", "sofrimento", Pos)
    // usadoPara ("paixão", "romance", Pre); ehUm("paixão", "sentimento", Pos)

  caso estrutura(exp) = "<nome><adjetivo>" ou "<adjetivo><nome>":
    // 1ª. iteração: exp = "crime passionnal"

    exp1 = primeiroTermo(exp); // exp1 = "crime"
    exp2 = segundoTermo(exp); // exp2 = "passional"
    conteudo1 = gerarConteudo(exp1); // chamada recursiva para exp1 = "crime"
    conteudo2 = gerarConteudo(exp2); // chamada recursiva para exp2 = "passional"
    se estrutura(exp1) = "<adjetivo>" então {
      conteudo1 = selecionarConteudo(conteudo1);
    } senão {
      conteudo2 = selecionarConteudo(conteudo2);

      // seleciona relações de conteúdo1 (referente a exp2 = "passional") conforme passo 2.c
      // Exemplos: eventoPreRequisitoDe("paixão", "amante", Pre);
      // efeitoDe("paixão", "sofrimento", Pos)
      // usadoPara ("paixão", "romance", Pre);
      // Obs: a relação "ehUm("paixão", "sentimento", Pos)" não foi selecionada
    }
    retorna conteudo1+conteudo2;

  caso estrutura(exp) = "<adjetivo><nome><adjetivo>":
    se <adjetivo> qualifica "<nome><adjetivo>" então {
      exp1 = <adjetivo>;
      exp2 = <nome> <adjetivo>;
      conteudo1 = gerarConteudo(exp1);
      conteudo2 = gerarConteudo(exp2);
      conteudo1 = selecionarConteudo(conteudo1);
    } senão {
      exp1 = <adjetivo><nome>;
      exp2 = <nome><adjetivo>;
      conteudo1 = gerarConteudo(exp1);
      conteudo2 = gerarConteudo(exp2);
      retorna conteudo1+conteudo2;
    }
  }
  retorna conteudo1+conteudo2; }

```

Figura 1. Algoritmo para geração de conteúdo de conceitos

3. Avaliação

A avaliação realizada visou analisar dois aspectos: (i) o quão as heurísticas facilitam a aquisição de conhecimento conceitual de senso comum e pragmático para língua portuguesa; (ii) a qualidade do conteúdo conceitual gerado pelas heurísticas, ou seja, se o conteúdo proposto realmente expressa o valor semântico do conceito desejado pelo usuário. Nesta avaliação, o algoritmo foi implementado para AC de conceitos para base InferenceNet e utilizou o *parser* PALAVRAS [Bick 2000]. No entanto, o método pode ser aplicado em outras bases de conhecimento de senso comum e pode ser utilizado outro *parser* sintático para língua portuguesa. A metodologia de avaliação seguiu os passos delineados na sequência.

- (1) Seleção de 20 pessoas adultas com experiência em sistemas interativos da Internet e que não tinham conhecimento sobre o método de AC, proposto neste trabalho. As pessoas foram distribuídas aleatoriamente em 2 (dois) grupos de 10 pessoas, um grupo para cada cenário de teste;
- (2) Seleção de conceitos usados na língua portuguesa que não existiam previamente na base InferenceNet: "crime passionnal", "má iluminação pública" e "violência policial";
- (3) Definição de cenários de teste:

Cenário 1 — usuários irão incluir, sem limite de tempo, relações semânticas para os conceitos escolhidos, através do portal www.inferencenet.org, o qual possui uma interface interativa que permite a entrada de relações de senso comum e inferencialista na base InferenceNet.

Cenário 2 — no portal www.inferencenet.org, o usuário informa a expressão linguística EXP correspondente ao conceito e interage com o portal para validar o conteúdo conceitual gerado pelo algoritmo implementado. Os usuários foram orientados a alterar e a excluir relações semânticas caso não concordassem com elas, além de incluir novas relações caso julgassem ainda necessárias, sem limite de tempo.

- (4) Geração de *baseline*, onde um avaliador humano validou as relações semânticas geradas pelo algoritmo e definiu uma *baseline* para os conceitos desta avaliação. A *baseline* serviu para análise qualitativa do conteúdo semântico ao final do processo de AC experimentado pelos 10 usuários no Cenário 2.

Em cada cenário, foi medido o tempo para realização da atividade e quantas relações semânticas foram incluídas e excluídas para cada conceito selecionado. A tabela 3 apresenta os resultados médios coletados. No Cenário 2, o algoritmo implementado gerou as seguintes quantidades de relações para os conceitos em questão: **crime passionnal** — 45 precondições e 17 pós-condições; **má iluminação pública** — 13 precondições; **violência policial** — 67 precondições e 1 pós-condição. Como resultado principal, tem-se que 76% das relações geradas foram validadas por humanos (considerando a média de exclusões da *baseline* e dos 10 usuários que participaram do Cenário 2).

Tabela 3. Resultados coletados nos dois cenários de avaliação e da *baseline*

Cenários	“crime passional”			“má iluminação pública”			“violência policial”		
	Incl	Excl	Tempo	Incl	Excl	Tempo	Incl	Excl	Tempo
Cenário 1	2,9 pre 1,6 pos	-	00:02:55	2,9 pre 1,2 pos	-	00:02:25	4,1 pre 2 pos	-	00:02:31
Cenário 2	0 pre 0 pos	8,2 pre 0,7 pos	00:03:46	0 pre 0,1 pos	6,3 pre 0 pos	00:01:42	0 pre 0,1 pos	17,9 pre 0,4 pos	00:03:52
Baseline	4 pre 1 pos	10 pre 5 pos	n/a	4 pre 3 pos	7 pre 0 pos	n/a	5 pre 1 pos	23 pre 1 pos	n/a

A partir dos resultados coletados, tem-se que o método proposto possibilita interações mais produtivas para AC — no Cenário 1 os usuários levaram, em média, 2min31s para inclusão de 4,9 relações semânticas (média de precondições e pós-condições incluídas para os três conceitos), enquanto que no Cenário 2 os usuários realizaram 11,2 exclusões e inclusões de relações semânticas em 3min6s (tempo médio). Observamos que, no Cenário 1, os usuários encontraram dificuldade em externar relações semânticas de senso comum sobre o conceito e, em alguns casos, até mesmo em se lembrar o que caracterizaria semanticamente aquele conceito. No Cenário 2, o usuário é instigado a interagir com as relações semânticas geradas, resultando em uma melhor relação inclusões/exclusões por minuto (3,61 no Cenário 2 contra 1,96 no Cenário 1). Outro fato interessante, é que a quantidade de relações semânticas geradas pelo método é bem maior do que as inclusões feitas pelos usuários no Cenário 1, mesmo considerando as exclusões realizadas no Cenário 2. Em relação à qualidade do conteúdo conceitual gerado pelas heurísticas, comparamos os grafos conceituais dos três conceitos, após as inclusões e exclusões realizadas pelos usuários, e os grafos conceituais da *baseline*. Como resultado, o método proposto possibilitou 70% de acurácia em média para os três conceitos analisados, considerando as relações mais excluídas pelos usuários. Importante salientar que, no Cenário 2, os usuários se limitaram a excluir as relações que lhes pareceram inválidas para o conceito, e praticamente não incluíram novas relações.

Noutro experimento realizado buscamos medir a cobertura das heurísticas propostas na base InferenceNet, ou seja, o quanto as heurísticas conseguem recuperar conceitos similares a novos conceitos. A lista de novos conceitos foi formada com 500 nomes de marcadores da base de mapas colaborativos criados através da ferramenta WikiMapps (www.wikimapps.com). No primeiro cenário, aplicamos as heurísticas (seção 3.1), sem interação com usuário, para recuperação de conceitos relacionados aos marcadores. No segundo cenário, aplicamos a ferramenta de indexação LUCENE para recuperar conceitos sintaticamente relacionados aos nomes dos marcadores. Por fim, um humano analisou a qualidade dos conceitos recuperados em ambos os cenários, descartando não conformidades. Como resultado, identificamos que as heurísticas conseguiram recuperar conceitos para 81% dos marcadores e a ferramenta de indexação sintática LUCENE recuperou conceitos para 62% dos marcadores. Importante salientar que, neste último caso, os conceitos recuperados foram somente aqueles sintaticamente similares ao nome do marcador. Por exemplo, para o marcador “educação” foi recuperado por LUCENE o conceito “educar”. No caso do processo heurístico de caracterização semântica, foram recuperados conceitos semanticamente relacionados.

Por exemplo, para o marcador “favela” foi recuperado o conceito “gueto”; e para o marcador “desastre ambiental” foram recuperados conceitos “ambiente” e “floresta”.

4. Trabalhos Relacionados

As pesquisas em AC têm se concentrado na aquisição de conhecimento de senso comum a partir de esforço colaborativo de especialistas bem como de usuários diversos da web. O projeto CYC [Lenat 1995] é um dos mais antigos exemplos de base de senso comum construída a partir de especialistas. Em 2000, o projeto *Open Mind Common Sense* (OMCS) [Singh et al. 2002] surgiu com o objetivo de coletar, pela Internet e de colaboradores voluntários, sentenças que expressavam fatos da vida comum. O *corpus* OCMS deu origem às triplas de conhecimento de senso comum da ConceptNet [Liu e Singh 2004]. A nova versão do OMCS [Speer, 2007] já provê funcionalidades para que o usuário ajude a refinar e validar o conhecimento coletado. O projeto Verbosity [von Ahn, Kedia e Blum 2006] é um jogo interativo para AC de conhecimento de senso comum. Cankaya e Moldovan (2009) propõe um método automático para gerar novas triplas de conhecimento a partir de metaregras de senso comum. O algoritmo proposto busca automaticamente em uma base estendida da WordNet¹ por conceitos que têm determinada propriedade e gera novos axiomas usando fatos de senso comum. Para língua portuguesa, são poucos os projetos de AC de senso comum. O principal é o projeto *Open Mind Common Sense - Brasil* (OMCS-Br) para coleta de conhecimento de senso comum em português por colaboradores na web [Anacleto et al. 2006].

O diferencial do método proposto neste artigo é a recuperação de conteúdo similar da base de conhecimento, o qual facilita interações mais produtivas para aquisição de conhecimento de senso comum e inferencialista para novos conceitos. Na Figura 1, para aquisição do novo conceito “crime passionai”, o algoritmo propõe ao usuário relações semânticas recuperadas do conteúdo conceitual de “paixão”, a saber: *eventoPreRequisitoDe* (“paixão”, “amante”, *Pre*); *efeitoDe* (“paixão”, “sofrimento”, *Pos*); *usadoPara* (“paixão”, “romance”, *Pre*), possibilitando maior riqueza no processo de AC do novo conceito.

5. Conclusão

Neste trabalho, propomos um método semi-automático para AC de senso comum e inferencialista. O diferencial do método em relação ao estado da arte é um processo automático de raciocínio que gera novos fatos de senso comum e pragmáticos para conceitos da língua portuguesa, a partir de conteúdo de outros conceitos similares e conforme a estrutura gramatical de sintagmas nominais. Além disso, a interação com usuário final permite uma validação das relações semânticas geradas e, conseqüentemente, melhor qualidade na aquisição de conhecimento desta natureza. O método foi implementado e avaliado para a base de senso comum e inferencialista da língua portuguesa — InferenceNet, e obteve 76% de validação por usuários humanos. Além disso, o método proposto possibilitou interações mais produtivas para AC, pois, com uma base inicial para validação, o usuário é instigado sobre relações semânticas de senso comum acerca do novo conceito. Como trabalhos futuros, tem-se a evolução do algoritmo com novas heurísticas para geração de conteúdo inferencial que contemple demais estruturas gramaticais de sintagmas (por exemplo, “<nome> <preposição> <nome>”) e uma avaliação da complexidade do algoritmo proposto.

¹ <http://xwn.hlt.utdallas.edu/>, acessado em 10/02/2011.

Referências

- Anacleto et al. (2006) Can Common Sense uncover cultural differences in computer applications? In: International Federation for Information Processing, IFIP, 19th World Computer Congress, 217, pp.21–24.
- Anacleto et al. (2007) A Common Sense-Based On-line Assistant for Training Employees. In: Proceedings of INTERACT 2007, LNCS, 4662, pp.243-254.
- Baker, C., Ellsworth, M., Erk, K. (2007) SemEval-2007 Task 19: Frame Semantic Structure Extraction. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic, pp. 99-104.
- Bick, E. The Parsing System "Palavras". (2000) Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus University Press.
- Cankaya, H., Moldovan, D. (2009) Method for extracting commonsense knowledge. In Proceedings of the fifth international conference on Knowledge capture (2009), K-CAP '09, pp. 57-64.
- Che et al. (2009) Multilingual Dependency-based Syntactic and Semantic Parsing. In: Proceedings of the Conference on Computational Natural Language Learning (CoNLL-2009) Shared Task.
- Kay, M. (2003) Introduction to Computational Linguistics. In: Mitkov, R. (ed). Oxford Handbook of Computational Linguistics, Oxford University Press, pp.17-20.
- Lenat, D.B. (1995). CYC: Large-Scale Investment in Knowledge Infrastructure. Communications of the ACM, vol.38, no 11, pp.28-32.
- Liu, H., Singh, P. (2004) "ConceptNet: A Practical Commonsense Reasoning Toolkit", BT Technology Journal, Volume 22(4), pp.211-226. Kluwer Academic Publishers.
- Pardo, T.A.S., Caseli, H.M., Nunes, M.G.V. (2009) Mapeamento da Comunidade Brasileira de Processamento de Línguas Naturais. In: Proceedings of the 7th Brazilian Symposium in Information and Human Language Technology, STIL, Brasil.
- Pinheiro, V. (2010) SIM: Um Modelo Semântico Inferencialista para Expressão e Raciocínio em Sistemas de Linguagem Natural. Tese de Doutorado. Universidade Federal do Ceará, MDCC, Fortaleza.
- Pinheiro, V., Pequeno, T., Furtado, V., Franco, W. (2010) InferenceNet.Br: Expression of Inferentialist Semantic Content of the Portuguese Language. In: T.A.S. Pardo et al. (eds.): PROPOR 2010, LNAI 6001, pp.90-99. Springer, Heidelberg.
- Singh, P. et al. (2002) Open Mind Common Sense: Knowledge acquisition from the general public. In: Proceedings of ODBASE'02, LNCS, No 2519, pp.1223-1237. Heidelberg: Springer-Verlag.
- Speer, R. (2007) Open Mind Commons: An inquisitive approach to learning common sense. In: Proceedings of the Workshop on Common Sense and Intelligent User Interfaces, January 28, 2007, Honolulu, Hawaii.
- von Ahn, L., Kedia, M., Blum, M. (2006) Verbosity: A Game for Collecting Commonsense Knowledge. ACM Conference on Human Factors Computing Systems, CHI Notes 2006, pp. 75-78.