

SIM: Um Modelo Semântico-Inferencialista para Sistemas de Linguagem Natural

Vlória Pinheiro
Universidade Federal do Ceará
Campus do Pici - UFC
Departamento de Computação
+55-85-33669841
vladia@lia.ufc.br

Tarcisio Pequeno
Universidade de Fortaleza
Av Washington Soares, 1321
Bairro Edson Queiroz
+55-85-34773000
tarcisio@unifor.br

Vasco Furtado
Universidade de Fortaleza¹ e ETICE²
¹Av Washington Soares, 1321
²Av. Pontes Vieira 220
+55-85-34773000¹ / +55-85-31016601²
vasco@unifor.br

Thiago Assunção
Universidade de Fortaleza
Av Washington Soares, 1321
Bairro Edson Queiroz
+55-85-34773000
thiagooa_unifor@yahoo.com.br

Emanoel Freitas
Universidade de Fortaleza
Av Washington Soares, 1321
Bairro Edson Queiroz
+55-85-3477 3000
emanoeltadeu@gmail.com

ABSTRACT

One of the growing needs related to systems of Natural Language Processing (NLP) is that such systems must be able to perform enriched textual inferences. We argue that one reason for the current limitation of the inferences generated by these systems is that—for the most part—they are based on the characteristics of the things represented by names, and seek to draw inferences based on such characteristics. In this work, we propose the Semantic Inferentialism Model (SIM), which follows a natural path and represents a new paradigm: it seeks to express the inferential capacity of concepts and how these concepts, combined in sentence structures, contribute to the inferential power of sentences. We present a SIM-based Information Extraction System and a pre-evaluation of the results.

RESUMO

Uma crescente necessidade relacionada com os sistemas de Processamento de Linguagem Natural (PLN) é que estes sejam capazes de realizar inferências mais ricas sobre textos. Argumentamos que uma razão para a atual limitação das inferências geradas por tais sistemas é que eles, em sua maioria, tomam como base as características das coisas representadas pelas palavras e, a partir delas, buscam realizar inferências. Neste trabalho, propomos o Semantic Inferentialism Model (SIM) que segue um caminho natural e representa um novo paradigma: busca expressar a capacidade inferencial de conceitos e como estes, combinados em estruturas de frases, contribuem para as premissas e conclusões de sentenças. Apresentamos uma aplicação do SIM em um sistema de extração de informação e uma pré-avaliação dos resultados.

Categories and Subject Descriptors

I.2.7 [Artificial Intelligence]: Natural Language Processing – *language models, language understanding, text analysis.*

General Terms

Algorithms, Design, Experimentation, Languages.

Keywords

Análise Semântica, Inferência Textual, Linguagem Natural.

1. INTRODUÇÃO

Existe uma necessidade crescente por agentes capazes de entender e trabalhar com o grande volume de informações textuais à nossa mão. Porém, a despeito de toda evolução dos sistemas de Processamento de Linguagem Natural (PLN), há um consenso de que a tecnologia para entendimento de textos em linguagem natural tem alcançado resultados superficiais [1]. Argumentamos que uma razão para tal superficialidade é que a maioria dos sistemas de PLN segue abordagens semânticas que tomam como base as características das coisas representadas pelas palavras e, a partir delas, buscam realizar inferências. Nesse modelo semântico representacionalista, o significado é dado trocando-se palavras e expressões por descrições externas à prática linguística ou usos da linguagem. No entanto, é dentro da prática linguística que as circunstâncias para usar uma palavra e as consequências do uso podem ser apreendidas e, ao desconsiderá-las, perde-se muito do que poderia ser inferido.

Ter a capacidade de processar, entender e argumentar sobre um texto requer habilidade para realizar inferências textuais oriundas, às vezes, de relações implícitas. Por exemplo, de “Eu farei prova de matemática amanhã” inferir que o sujeito terá que estudar e não poderá ir ao cinema. Inferências como estas só podem ser realizadas se o agente tiver o domínio do uso de conceitos: se souber quando aplicá-los, se souber com o que se compromete ao usá-los em uma dada sentença junto com outros conceitos, e como esta combinação contribui para as premissas e conclusões da sentença. Esta é a abordagem semântica das teorias inferencialistas de Sellars, Dummett e Brandom. Para Dummett, saber o significado de uma sentença é saber a justificativa para o falante tê-la proferido: “Nós não explicamos o sentido de uma declaração estipulando seu valor-verdade em termos dos valores-verdade de seus constituintes, mas sim estipulando quando ela pode ser afirmada em termos das condições sobre as quais seus constituintes podem ser afirmados” [2]. A teoria semântica inferencialista de Brandom [3] sedimenta a visão de Dummett e fornece uma redução da visão pragmática da linguagem, de Wittgenstein [4], para um racionalismo pragmático onde a tônica são os usos inferenciais de conceitos. Para Wittgenstein, o significado de uma palavra ou expressão são os diversos usos das mesmas em jogos de linguagem, não sendo possível uma sistematização destes usos. Brandom traz de volta uma possibilidade computacional para a semântica da linguagem natural, conservando seu viés pragmático, na medida em que sistematiza a apreensão do significado em jogos racionais, onde o componente inferencial é o mais importante.

Neste trabalho, propomos um modelo baseado na teoria semântica inferencialista de Brandom – o Semantic Inferentialism Model (SIM), o qual define os principais requisitos para expressar e manipular conhecimento lingüístico de forma a capacitar os sistemas de PLN a realizarem inferências, a princípio, de forma mais natural. Com a evolução de pesquisas e aplicações baseadas neste novo paradigma, argumentamos que a habilidade dos agentes para tratar linguagem natural será alavancada, pois os estamos capacitando com um arcabouço inferencial. Arcabouço este que contém o relacionamento inferencial entre conceitos e o conhecimento de como combinar estes relacionamentos em estruturas de sentenças a fim de realizar inferências mais complexas. Após a descrição do modelo, está descrita uma aplicação do SIM em um sistema de extração de informação e uma pré-avaliação dos resultados.

2.UM MODELO SEMÂNTICO INFERENCIALISTA PARA PLN

A maioria dos atuais sistemas de PLN expressa o significado de conceitos em uma rede semântica ou ontologias simples, onde conceitos são ligados a outros através de relações que denotam suas propriedades e uma taxonomia. Bases semânticas típicas como WordNet [5] and CYC [6] são formadas por uma rede de palavras ligadas através de relações léxicas e semânticas que descrevem as coisas que representam. Um problema inicial é que algumas relações entre conceitos não vêm das características das coisas representadas por palavras e expressões, mas sim das circunstâncias e conseqüências do uso de tais expressões apreendidas dentro de práticas lingüísticas, obrigando os

sistemas de PLN a projetar métodos *ad hoc* e específicos do domínio.

O caminho natural é inverso ao dos atuais sistemas de PLN. Uma criança não aprende uma língua decorando um dicionário ou léxico contendo descrições de como as coisas são no mundo, aprende estruturas de sentenças e como combinar as palavras nestas estruturas, a partir do que a autoriza a usar uma palavra e do que pode concluir se usá-la em uma dada estrutura. Portanto, o que deve ser expresso é a capacidade inferencial de conceitos (pré e pós-condições de uso), e como estes, combinados em estruturas de frases, contribuem para as pré-condições e pós-condições de uma sentença que possui certa estrutura. Para endereçar este novo paradigma de expressão semântica, propomos um novo modelo - Semantic Inferentialism Model (SIM), composto de:

1. Uma base conceitual, contendo pré e pós-condições de uso de conceitos. Um conceito pode ser um nome, verbo, advérbio, adjetivo, ou conceitos mais complexos, como “prova de matemática” e “cometer crime”. Pré e pós-condições de uso de um conceito são expressas através de relações entre o conceito e outro, formando uma rede inferencial entre conceitos. Portanto, ao proferir um conceito temos à mão todas as relações inferenciais (não representacionais) deste com outros conceitos, na forma de suas premissas e conclusões;
2. Uma base de sentenças-padrões, que são estruturas de frases de uma língua, p.ex., “X ser assassinar com Y”. Uma sentença-padrão tem partes variáveis, p.ex., X e Y, que serão instanciadas com elementos da base conceitual, e partes fixas, contendo preposições, elementos de ligação e verbos. A uma sentença-padrão associamos outras sentenças-padrões como suas pré ou pós-condições, definindo como os elementos conceituais articulados contribuem para as pré-condições e pós-condições da sentença que segue o padrão. Além disso, possibilita associarmos conteúdo inferencial que não pode ser inferido dos conceitos, tais como, regras e práticas sociais influenciadas pela cultura local;
3. Uma base de regras para raciocínio prático, que combina o conteúdo dos conceitos (base conceitual) articulados em uma sentença e gera novas pré e pós-condições para a mesma. Brandom [3] exemplifica alguns padrões de raciocínio prático, endossados por vocabulários normativos como “preferir”, “ser obrigado a”, “dever”, p.ex., se “X é um bancário” e “Bancários são obrigados a usar gravatas” então “X deve usar gravata”. Uma regra de raciocínio prático pode então ser criada para gerar, como pós-condição da sentença “X é um bancário”, obrigações expressas como pós-condições do conceito “bancário”;
4. Um cálculo de similaridade inferencial para descobrir quais conceitos da base conceitual são usados na sentença e com quais pré e pós-condições contribuem para o significado da sentença. Algumas heurísticas podem ser usadas. Por exemplo, a partir do(s) conceito(s) associado(s) ao termo subsentencial e de sua classe gramatical, calcula-se um índice de similaridade entre suas pré e pós-condições e as pré e pós-condições dos demais conceitos já fixados na

sentença. Através do índice, pode-se descartar pré e pós-condições do(s) conceito(s), irrelevantes para compor o significado da sentença.

O SIM pode ser usado como base para diversas aplicações de lingüística computacional: geradores de textos, sistemas de recuperação de informação baseados em conteúdo, sistemas de extração de informação, sistemas de Question&Answer, dentre outras. Com a base de conhecimento provida pelo SIM, o diferencial será uma melhor habilidade inferencial, possibilitando a descoberta de informações implícitas em textos.

3.APLICAÇÃO DO SIM EM UM SISTEMA DE EXTRAÇÃO DE INFORMAÇÃO

Uma necessidade do projeto WikiCrimes¹ é extrair informações sobre crimes a partir de notícias publicadas em jornais, na web. Para isso, poderiam ser usadas abordagens usuais para extração de informação (Information Extraction - IE) e/ou text mining [7]. No entanto, muitas vezes as informações sobre crimes (local, tipo do crime, arma utilizada etc) estão implícitas no texto jornalístico ou, no melhor caso, estão dispersas ao longo do texto, e algumas inferências mais complexas precisam ser realizadas para se concluir os atributos do crime descrito.

Seja o seguinte texto descritivo de um crime: “Uma mulher foi executada com um tiro na cabeça, em Piraquara, na madrugada de ontem, pelo amante Edilson Bezerra Pinto. Os policiais Leandro e Vitor foram até a Rua Santa Catarina e encontraram o corpo da mulher”. Nós, proficientes da língua portuguesa, facilmente conseguimos inferir que foi um crime passionai, que provavelmente ocorreu na Rua Santa Catarina com uso de arma de fogo. Se nos fosse pedido para dar razões sobre tais conclusões, daríamos explicações do tipo: armas de fogo são usadas para dar tiros; o autor do crime é amante da vítima e, por isso, o crime foi passionai; policiais são capazes de prender suspeitos de crime; o local provável do crime é onde o corpo foi encontrado etc. Tais explicações são oriundas de um conjunto de inferências que somos capazes de fazer a partir do nosso conhecimento sobre (i) as circunstâncias e conseqüências dos conceitos articulados nas sentenças, e sobre (ii) as estruturas de sentenças da língua, que determinam como cada conceito contribui para o significado da sentença, de acordo com a função sintática exercida pelo conceito na sentença. Dispomos, ainda, de uma base de regras de senso comum, que nos permite gerar informações a partir de conhecimentos associados à nossa cultura: por exemplo, que é errado matar nestas circunstâncias (não existe uma guerra) e que o assassino deve ser preso. Além disso, temos a capacidade de filtrar quais inferências são relevantes, a partir dos outros conceitos articulados e do que nos

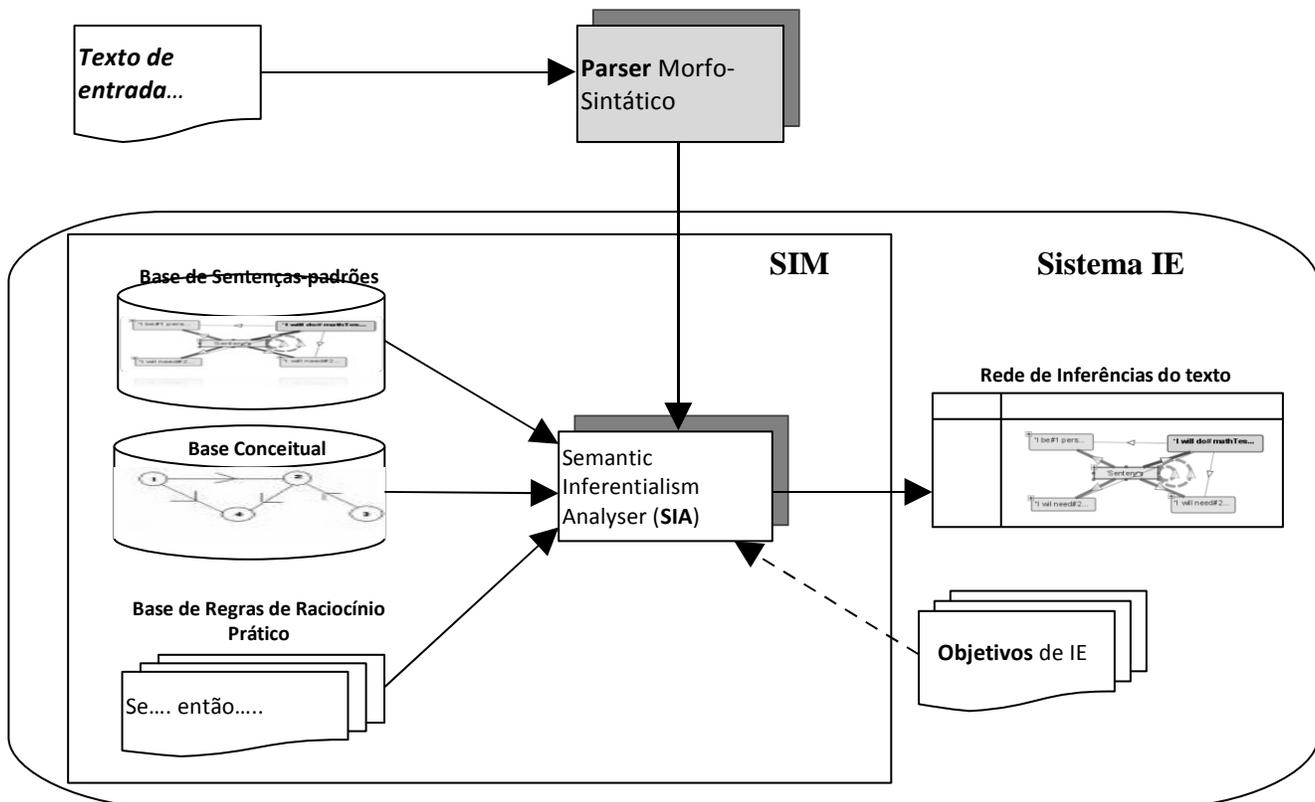


Figura 1. Arquitetura e Processo do Sistema de Extração de Informações (IE) baseado no Semantic Inferentialism Model (SIM).

¹ <http://www.wikicrimes.org>, acessado em 09/08/2008.

está sendo questionado. Ora, o SIM provê justamente todo este arcabouço.

A figura 1 apresenta a arquitetura e processo do sistema de IE baseado no SIM. O SIM recebe um texto de entrada, analisado sintaticamente por um parser, realiza a análise semântica a partir de suas bases (base de sentenças-padrões, base conceitual e base de regras de raciocínio prático) e gera a rede de inferências do texto, filtrando-as, opcionalmente, pelos objetivos de extração de informação. O componente principal do framework SIM é o Semantic Inferentialism Analyser (SIA), que implementa um algoritmo de análise semântica genérico. Este algoritmo combina as estruturas de sentenças do texto de entrada com sentenças-padrões da base, decide sobre quais conceitos da base conceitual foram usados na sentença e com quais pré e pós-condições contribui, através do cálculo de similaridade inferencial, e gera as pré e pós-condições da sentença (rede inferencial) a partir de todo o conhecimento inferencial e de regras de raciocínio prático. Opcionalmente, objetivos do sistema de IE são considerados para filtrar a rede de inferência a ser gerada pelo SIA, por exemplo, precisamos saber tipo do crime, local e tipo de arma utilizada. Logo, o SIA irá gerar apenas inferências relacionadas aos conceitos presentes nos objetivos. O SIA não será detalhado neste artigo devido a restrições de espaço. Por ser a base do paradigma, nos concentramos principalmente na construção das bases do framework SIM.

O PROJETO DAS BASES DO SIM

Um primeiro desafio foi construir as bases do modelo em português. A base conceitual foi construída a partir da ConceptNet [1], uma larga base de conhecimento de senso comum com 1.6 milhões de afirmações em inglês, e de um corpus lingüístico CRIMES2008, montado com notícias policiais publicadas em páginas da web de jornais locais. A rede semântica da ConceptNet é formada por nós, representando conceitos, ligados através de propriedades (capazDe, usadoPara, ehUm, efeitoDe etc.). Necessitávamos expressar pré e pós-condições de conceitos, não uma ontologia tradicional, e a estratégia foi definir, para cada propriedade da ConceptNet, que tipo de relação inferencial ela expressaria: uma pré ou uma pós-condição para o conceito. Por exemplo, a propriedade “usadoPara” foi definida como expressando uma pós-condição e, no caso da relação “*gun* >> *usedfor* >> *commit crime*”, seria gerada uma pós-condição para o conceito “arma”: “ser usada para cometer crime”. O corpus foi anotado sintaticamente pelo parser morfossintático PALAVRAS² e XTRACTOR³. A partir do

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

TIL '08, October 27-28, 2008, Vila Velha, ES, Brasil.

Copyright 2008 ACM 1-58113-000-0/00/0004...\$5.00.

² <http://beta.visl.sdu.dk/visl/pt/parsing/automatic/>, acessado em 21/07/2008

³ <http://abc.di.uevora.pt/xtractor/>, acessado em 21/07/2008

treebank gerado, selecionamos nomes e adjetivos, traduzimos para o inglês e buscamos na ConceptNet seus relacionamentos com outros conceitos, gerando-os como pré ou pós-condições daqueles (nomes e adjetivos). A base conceitual do SIM contém 61022 conceitos, dos quais 3,20% são conceitos articulados nos textos do corpus, ou seja, são conceitos do domínio de crimes. Os 96,80% restantes são conceitos relacionados por pré ou pós-condições aos conceitos do corpus e traduzidos para o português. Por exemplo, o conceito “crime” foi gerado com 52 pré-condições e 22 pós-condições. Do conjunto de nomes e adjetivos articulados no corpus (4295), 55% (2344) não foi encontrado na ConceptNet.

Para a base de sentenças-padrões do SIM foi aplicada solução de *text mining* sobre o *treebank* que pesquisou ocorrência de verbos principais e preposições. Cada ocorrência distinta era gerada como uma sentença padrão e sua frequência foi calculada: as sentenças com maiores frequências foram “X ser Y” (837) e “X ter Y” (496). Foram geradas 5963 sentenças-padrões de 22085 sentenças do corpus, ou seja, 73% das sentenças do corpus são estruturalmente iguais a pelo menos uma sentença-padrão.

A primeira versão do analisador semântico (SIA) do framework SIM não contemplou a base de raciocínio prático e o cálculo de similaridade inferencial, ou seja, todas as pré e pós-condições dos conceitos e sentenças-padrões, articulados no texto, foram combinadas para gerar inferências do texto. A tabela 1 apresenta algumas inferências geradas (de um total de 997) como resultado da análise semântico-inferencialista da sentença “Uma *mulher* foi executada com um *tiro* na *cabeça*, em Piraquara, na *madrugada* de *ontem*, pelo *amante* Edilson Bezerra Pinto”. Os conceitos sublinhados foram selecionados e apenas os conceitos “tiro” e “ontem” não possuem entradas na base conceitual, devido a problemas na tradução para o inglês e conseqüente falha na pesquisa à ConceptNet. Inferências como “Edilson Bezerra Pinto ser assassino” e “Edilson Bezerra Pinto matar uma mulher” foram geradas a partir de pós-condições diretas da sentença-padrão selecionada: “X ser executar por Y”. As pós-condições diretas desta sentença são: “X ser vítima” e “Y matar X”. Já a inferência “Uma mulher foi executada em algo que CapazDe conter cérebro, orelha, boca e nariz.” foi gerada a partir da seguinte pós-condição do conceito “cabeça”: “cabeça >> CapazDe >> conter cérebro, orelha, boca e nariz”.

4.DIRECIONANDO UMA AVALIAÇÃO DO SIM

Direcionar uma avaliação qualitativa de um modelo inferencialista para entendimento de língua natural implica em responder questões tais como: Quantas inferências que as pessoas fariam ao ler um texto, o SIA consegue inferir? Quantas inferências que o SIA gera estão “corretas”? Como decidir se uma inferência é correta ou não? Uma solução é tentar medir isto manualmente e de forma indireta através do sistema de IE para Wikicrimes: quantos crimes ocorridos foram identificados (*recall*)? Quantos crimes identificados eram, de fato, crimes ocorridos (*precision*)? Quantos dos atributos dos crimes foram definidos e quantos o foram corretamente? Esta avaliação está prevista dentro do projeto do sistema de IE para Wikicrimes.

Uma avaliação intuitiva é que o SIM capacita um agente a realizar inferências sobre textos de forma mais direta e natural, visto que já provê conhecimento inferencial em suas bases. No entanto, via de regra, o sucesso da abordagem depende da qualidade da representação semântica. Nesse sentido, realizamos uma pré-avaliação manual sobre as inferências geradas na análise semântica do corpus CRIMES2008. A partir daí, foi possível elencar um conjunto de melhorias na construção e projeto das bases do SIM, dentre elas a melhoria da abordagem usada para tradução da ConceptNet para o português, pois 55% dos conceitos do corpus não foram encontrados na ConceptNet, prejudicando as inferências sobre os textos.

Tabela 1. Visão parcial das inferências geradas como resultado da análise semântico-inferencialista realizada sobre a sentença “Uma mulher foi executada com um tiro na cabeça, em Piraquara, na madrugada de ontem, pelo amante Edilson Bezerra Pinto”

Inferências gerada pelo Analisador Semântico SIA	Conceito ou sentença-padrão associados
Uma mulher ser vítima	X ser executar por Y
Edilson Bezerra Pinto ser assassino	X ser executar por Y
Edilson Bezerra Pinto matar uma mulher	X ser executar por Y
Edilson Bezerra Pinto cometer um crime	X matar
Uma mulher ser executar em Piraquara	X ser executar em Y
Uma mulher é CapazDe gastar dinheiro	Mulher
Uma mulher ÉUm humano feminino	Mulher
Uma mulher foi executada em algo que CapazDe conter cérebro, orelha, boca e nariz.	Cabeça
Uma mulher foi executada em algo que DefinidaComo parte do corpo com pescoço	Cabeça
Uma mulher foi executada em algo que CapazDeReceberAcao sustentação	Cabeça
Uma mulher foi executada por algo que CapazDeReceberAcao atrair	Amante
Uma mulher foi executada por algo que DefinidaComo arquétipo do homem	Amante
Uma mulher foi executada em algo que CapazDeReceberAcao não ir para o trabalho	Madrugada
Uma mulher foi executada por algo que CapazDe ver televisão (*)	Amante

Outro fato identificado foi um número grande de inferências geradas, 342 em média. Para este ponto de melhoria, temos a proposta, no SIM, de um cálculo de similaridade inferencial. Este cálculo pode ser usado para eliminar inferências *nonsense* ou “sujeiras”, originadas da abordagem livre que a ConceptNet foi construída, e/ou para eliminar inferências irrelevantes para um determinado jogo racional, embora possam ser relevantes em outro. Por exemplo, a inferência “Uma mulher foi executada por algo que CapazDe ver televisao” ((*) na Tabela 1) é irrelevante para o jogo de descrever um crime. Nosso argumento para um novo cálculo é de que os cálculos tradicionais de similaridade, baseados na distância entre os nós da rede semântica [1][8][9][10] ou mesmo aqueles baseados em análise de corpus [11][12][13] são insuficientes para tais fins. Precisamos de uma medida derivada do quanto as pré e pós-condições de um conceito estão relacionadas com as pré e pós-condições já definidas para a sentença. Uma idéia é identificar as inferências relevantes com relação aos objetivos do agente. Este cálculo de similaridade é um diferencial do modelo semântico-inferencialista proposto (SIM) e busca simular nossa capacidade de filtrar o que argumentar com base no que nos está sendo questionado ou com base em nossos objetivos ao ler um texto.

5. TRABALHOS RELACIONADOS

O cerne deste artigo é um novo modelo para expressão semântica – SIM, e, por isso, nosso foco na análise dos trabalhos relacionados foi identificar as bases léxico-semânticas utilizadas e o tipo de inferências textuais providas por tais bases. De Paiva et AL [14] propõem uma lógica para representação semântica e raciocínio sobre textos em língua natural – Textual Inference Logic – TIL2. A ontologia base da TIL2 é uma versão caseira unificada da WordNet/VerbNet. TIL2 provê um método para detectar conclusões válidas e contradições usando raciocínio sobre a hierarquia entre os conceitos. Saias e Quaresma [15] propõem um sistema para leitura de notícias na web baseado na ontologia SENSO, enriquecida com conhecimento de senso-comum da ConceptNet, o que permite que as inferências realizadas não se restrinjam a conhecimento taxonômico. Java et AL [8] apresentam o OntoSem, um framework com várias ferramentas para extração e representação do significado de forma independente de linguagem. Utiliza uma ontologia, contendo 8500 conceitos descritos por meio de 16 propriedades, em média, como recurso central para extrair e representar significado de textos. Mellish e Pan [16] apresentam uma investigação sobre inferências adequadas para geração de textos em linguagem natural que buscam explicar conceitos definidos em uma ontologia. O foco é investigar subsumers de A (A é um conceito definido em uma ontologia) que são adequados para responder, em linguagem natural, à pergunta: “O que é A?”.

Como visto, a maioria dos sistemas de PLN e aplicações realizam análise semântica e inferências textuais a partir de largas bases de dados léxico-semânticas como WordNet e outras ontologias que expressam taxonomias. Este modelo de expressão semântica não prioriza o conhecimento de como as relações léxico-semânticas entre conceitos contribuem para a geração das premissas e conclusões de uma sentença. Os sistemas de linguagem natural, baseados nestes modelos, são, então,

obrigados a usar métodos *ad hoc* para realizar inferências mais interessantes.

O SIM permite flexibilidade para expressar conteúdo inferencial em suas bases e capacitar um agente a dominar uma forma composicional de se produzir e entender sentenças. Expressando conteúdo inferencial em suas bases (o que autoriza a usar um conceito e com o que se compromete ao usá-lo; pré e pós-condições sentenças-padrões, e um cálculo de similaridade inferencial), o SIM habilita o agente a realizar inferências de forma mais natural e direta. Através do seu algoritmo de análise semântica (SIA), habilita o agente com uma forma composicional de gerar as inferências de sentenças, a partir do conteúdo inferencial das bases do SIM e de regras de raciocínio prático. Ou seja, o SIA provê o agente com uma camada superior para raciocínio lingüístico sobre o conteúdo inferencial dos conceitos e sentença-padrão, e não apenas sobre suas formas. Este modelo de expressão semântica é mais adequado como modelo computacional para tratamento de linguagem natural, pois, ao expressar premissas e conclusões, circunstâncias e conseqüências, provê a base para boas argumentações e explicações.

6. CONCLUSÃO

Nossa pesquisa tem fundamentos em teorias semânticas filosóficas modernas e não usuais em sistemas de linguagem natural, por isso se justifica enquanto investigação científica na busca de compreender a natureza do significado e de simular a capacidade humana para entendimento da língua natural através de um novo modelo computacional – o SIM. Neste artigo apresentamos este novo modelo que busca representar em suas bases a capacidade inferencial de conceitos e estruturas de sentenças, provendo um caminho direto e natural para realizar inferências sobre textos. A aplicação do SIM em um sistema de extração de informação para o projeto WikiCrimes está sendo o cenário de avaliação do modelo. Até o momento, foi possível realizar uma pré-avaliação das bases do SIM e de seu analisador semântico – SIA.

7. REFERÊNCIAS

- [1] Liu, H. and Singh, P. (2004) ConceptNet: A Practical Commonsense Reasoning Toolkit. *BT Technology Journal*, Volume 22(4). Kluwer Academic Publishers.
- [2] Dummett, M. (1978) *Truth and Other Enigmas*. Duckworth, London.
- [3] Brandom, R.B. (2000) *Articulating Reasons*. In: *An Introduction to Inferentialism*. Harvard University Press, Cambridge.
- [4] Wittgenstein, L. (1953) *Philosophical Investigations*. Tradução G.E.M. Anscombe, Oxford: Basil Blackwell.
- [5] Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. MIT Press.
- [6] Lenat, D.B. (1995) CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11).
- [7] Cohen, K., Hunter, L. (2008) Getting staterd in text mining. *PLoS Compt Biology*, 4(1).
- [8] Java, A. et al. (2007). Using a Natural Language Understanding System to Generate Semantic Web Content. *International Journal on Semantic Web and Information Systems*. November.
- [9] Hirst, G., St-Onge, D. (1998). Lexical chains as representations of context for the detection and correction of malapropisms. In Fellbaum 1998, pp. 305–332.
- [10] Leacock, C., Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. In Fellbaum 1998, pp. 265–283.
- [11] Resnik, P. (1995). Using information content to evaluate semantic similarity. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 448–453, Montreal.
- [12] Jiang, J., Conrath, D. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of International Conference on Research in Computational Linguistics*, Taiwan.
- [13] Lin, D. (1998). An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning*. Madison, WI.
- [14] De Paiva, V. et al. (2007) *Textual Inference Logic: Take Two*. *Proceedings of the Workshop on Contexts and Ontologies, Representation and Reasoning, CONTEXT 2007*
- [15] Saias, J e Quaresma, P. (2006). A proposal for an ontology supported news reader and question-answer system. *Proceedings of the 2nd Workshop on Ontologies and their Applications*.
- [16] Mellish, C e Pan, J.Z. (2008) Natural language directed inference from ontologies. *Artificial Intelligence*, Volume 172 , Issue 10 (Junho 2008), pp.1285-1315.