

*Relatório do Segundo Trabalho de Inteligência
Artificial: Opção 1*

Erick Grilo
Max Fratane
Matheus Prado
Vitor Santos

1 *Knowledge Bases*

1.1 Introdução

Uma base de conhecimento (do inglês, *knowledge base*, abreviadas KB) é uma espécie de repositório usada para armazenar dados complexos sobre os mais variados assuntos, que pode ser estruturados ou não, usados por sistemas computacionais. A grosso modo, uma base de dados sobre um determinado assunto, ou sites, como a *Wikipedia* e até sessões de perguntas e respostas (*F.A.Q.*) podem ser consideradas bases de conhecimento. Dentro do aspecto da computação uma KB é vista como um recurso que pode ser lido por máquinas com a intenção de disponibilizar informações. É usada em sistemas a fim de facilitar a coleção, organização e a extração de dados da mesma, por pessoas ou organizações.

O termo "base de conhecimento" foi cunhado em meados da década de 70, a fim de distinguir das bases de dados (bancos de dados). Nessa época, todas as informações eram armazenadas em bancos de dados hierárquicos ou relacionais, e a diferença entre as bases de dados era explícita: A base de dados possuía informações representadas como uma tabela, com *strings* ou números em cada elemento, usuários múltiplos, onde várias pessoas poderiam ter acesso à base de dados ao mesmo tempo, e dados estáticos e persistentes.

Os primeiros sistemas baseados em conhecimento necessitavam de dados que fossem armazenados de uma forma diferente de como os bancos de dados funcionam: necessitavam de ponteiros para outros objetos que, por sua vez, teriam ponteiros para outros dados. Tais sistemas visavam modelar o mundo que conhecemos, então, apenas dados armazenados em *strings* e inteiros não seriam suficientes. A representação ideal de conhecimento em bases de dados é com o uso de ontologia, que é a ideia de modelagem de um objeto, que pode possuir subclasses, interações com outras classes e instâncias.

Em um futuro (não tão distante), espera-se que humanos e bases de dados poderão trabalhar cooperativamente através da web semântica. A web semântica é uma extensão da web atual que procura padronizar a troca de dados e o formato de dados pela web usando o *Resource Description Format* (RDF) (Miller 1998). A Web semântica interliga significados de palavras e, neste âmbito, tem como finalidade conseguir atribuir um significado (sentido) aos conteúdos publicados na Internet de modo que seja perceptível tanto pelo humano como pelo computador.

1.2 Alguns exemplos de *Knowledge Bases*

Em geral, bases de dados on-line são bases de dados que tentam ler a internet a fim de extrair informações relevantes. Alguns exemplos são:

- **YAGO (Yet Another Great Ontology)**: É uma base de conhecimento desenvolvida *Max Planck Institute for Computer Science*, em Saarbrücken, na Alemanha. YAGO é uma base de dados semântica, que deriva (ou seja, absorve conhecimento) da Wikipedia, WordNet e do GeoNames. Atualmente, YAGO possui conhecimento de mais de 10 milhões de entidades (tais como pessoas, empresas, países, etc.). Algumas das especialidades do YAGO consiste em *YAGO, A High-Quality Knowledge Base* (n.d.):
 1. YAGO possui uma precisão de mais de 95% de veracidade de seus conhecimentos, que é avaliada manualmente.
 2. YAGO combina a taxonomia do *WordNet* University (2017) com a categorização da *Wikipedia*, atribuindo as entidades a mais de 350.000 classes.
 3. YAGO é uma ontologia que é âncorada no tempo e no espaço, ou seja, o YAGO atribui uma dimensão de tempo e uma dimensão de espaço para muitos de seus fatos e suas entidades. Além de uma taxonomia, YAGO possui domínios temáticos como "música", "pessoas", ou "ciência", provindos de domínios do *WordNet*.
 4. YAGO extrai e combina entidades e fatos de 10 Wikipédias de diferentes idiomas.

Os dados do YAGO são providenciados na sintaxe de *Turtle* (Terse RDF Triple Language), que é um formato para expressar dados em RDF, representando informações como triplas, que consiste em sujeito, predicado e objeto. Cada um desses itens é expressado como uma URI (Universal Resource Identifier) da Web. Seus dados também são expressados no formato *tsv* (*tab separated values*, um formato de texto simples que representa os dados em uma estrutura tabular). O YAGO também possui ligação com uma outra base de dados, a SUMO (<http://www.site.uottawa.ca:4321/sumo/index.html>)

- **NELL (Never-Ending Language Learning)** : É uma base de conhecimento desenvolvida na Universidade Carnegie Mellon, sob a alcunha de *Read the Web*, cujo principal objetivo é de ler e aprender com a Internet. Desde 2010, o NELL vem tentando fazer duas tarefas por dia: Primeiramente, ele tenta ler (extrair) fatos de textos encontrados em milhões de páginas da web e, em seguida, procura melhorar sua habilidade de leitura, tal que futuramente ele consiga extrair mais fatos da Web, de forma mais precisa. o NELL possui mais de 50 milhões de crenças candidatas (a estarem corretas), que ficam divididas em diferentes níveis. Ele possui uma alta confiança em cerca de 3,565,291 dessas crenças.

Diferentes de outros sistemas, NELL aprende de forma autômata. Porém, como erros podem ocorrer, a cada duas semanas, a equipe do NELL procura por erros em sua base de dados e coloca o NELL para "ler a web" novamente.

Em Madureira et al. (2017), A base de dados do NELL é representada como uma estrutura baseada em ontologias, caracterizadas por categorias, relações e suas instâncias. Como a base de conhecimento do NELL expande todos os dias, ela não contém todas as instâncias de todas as categoriais nem todas as instâncias de cada relação descritas na ontologia.

Referências

- Madureira, A. M., Abraham, A., Gamboa, D. & Novais, P. (2017), *Intelligent Systems Design and Applications: 16th International Conference on Intelligent Systems Design and Applications (ISDA 2016) held in Porto, Portugal, December 16-18, 2016*, Vol. 557, Springer.
- Miller, E. (1998), 'An introduction to the resource description framework', *Bulletin of the American Society for Information Science and Technology* **25**(1), 15–19.
- University, P. (2017), 'Wordnet, a lexical database for english'.
URL: <http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>
- YAGO, *A High-Quality Knowledge Base* (n.d.), <http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>. Acessado em 08-06-2017.