

## Contributo per il workshop AI per l'industria presso Ital-IA 2022

**Nome dell'azienda:** Babelscape

**Titolo:** SemanticPaths: Multilingual Knowledge Graphs and their application to head hunting, the food industry, and monitoring AI

**Persone coinvolte:** prof. Roberto Navigli, Stefan Bejgu, Francesco Cecconi, Alessandro Scirè

**Temi di ricerca:** Multilingual Natural Language Processing, Knowledge Graphs, Semantic Search

### Application:

Babelscape (<https://babelscape.com>) is a deep tech company, born as a Sapienza university spin-off company, committed to providing strong innovation in AI at the intersection of multilingual Natural Language Processing, knowledge graphs and text analytics. Starting from cutting-edge research carried out in the Sapienza NLP Group, the company created **WordAtlas**, the largest multilingual knowledge graph, covering more than 20 million concepts and named entities, tens of different domains and 500 languages. As a showcase of the usefulness of large knowledge graphs and information extraction technology, Babelscape recently launched a new software and service called **SemanticPaths**.

SemanticPaths brings together neuro-symbolic multilingual Natural Language Processing, which includes state-of-the-art Word Sense Disambiguation and Entity Linking, concept extraction and personalized knowledge graph creation to:

1. create a **customized intelligent search engine** along several custom dimensions (e.g., for a research body of work, scientific papers, researchers, institutions, concepts, macroareas; for a company's document base, documents, companies and people involved, concepts, etc.),
2. **monitor the information** within the knowledge graph and the underlying document base in order to see how it evolves over time,
3. **discover connections and correlations** between people, documents, articles, concepts, companies, entities and much more,
4. **share the connections** discovered through semantic searches.

Semantic Paths is **fully personalizable** with specific dimensions and categories of interest to highlight in the search results, seamlessly **scales to multiple languages** and needs only a document base to set up a custom semantic search engine. Importantly, once the search engine is set up, thanks to the underlying multilingual knowledge graph, search can be input in arbitrary languages even though the retrieved documents are written only in a given language.

**Related articles**

- SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining**  
Andrea Esuli, Fabrizio Sebastiani  
Opinion mining (OM) is a recent subdiscipline at the crossroads of information retrieval and computational linguistics which is concerned not with the topic a document is about, but with the opinion it expresses. OM has a rich set of applications, ranging ...  
Cited by 3326
- SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining**  
Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani  
In this work we present SENTIWORDNET 3.0, a lexical resource explicitly devised for supporting sentiment classification and opinion mining applications. SENTIWORDNET 3.0 is an improved version of SENTIWORDNET 1.0, a lexical resource publicly available for ...  
Cited by 3297
- Corpus-based and Knowledge-based Measures of Text Semantic Similarity**  
Rada Mihalcea, Courtney Corley, Carlo Strapparava  
This paper presents a method for measuring the semantic similarity of texts, using corpus-based and knowledge-based measures of similarity. Previous work on this problem has focused mainly on either large documents (e.g. text classification, information re...  
Cited by 1185
- Logic programming and databases**  
Stefano Ceri, Georg Gottlob, Letizia Tanca  
The topic of logic programming and databases. has gained in creasing interest

**Related researchers**

- Roberto Navigli**  
Università degli Studi di Roma La Sapienza  
H-index 58 | [ORCID](#) | [SCHOLAR](#)
- Roberto Basili**  
Università degli studi di Roma "Tor Vergata"
- Alessandro Moschitti**  
UniTrento  
H-index 44
- Bernardo Magnini**  
Istituto Trentino di Cultura  
H-index 44 | [SCHOLAR](#)
- Giovanni Semeraro**  
Università degli Studi di Bari - "Aldo Moro"  
H-index 43 | [ORCID](#) | [SCHOLAR](#)
- Alberto Lavelli**  
Istituto Trentino di Cultura  
H-index 28 | [ORCID](#) | [SCHOLAR](#)
- Pierpaolo Basile**  
Università degli Studi di Bari - "Aldo Moro"

**Related concepts**

- machine translation**  
The use of computers to translate from one language to another
- computational linguistics**  
The use of computers for linguistic research and applications
- Neural machine translation**  
Neural machine translation is an approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.
- expert system**  
In artificial intelligence, an expert system is a computer system emulating the decision-making ability of a human expert.
- Information extraction**  
Information extraction is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents and other electronically represented sources.
- natural language understanding**  
Natural-language understanding or natural-language interpretation is a subtopic of natural-language processing in artificial intelligence

Once it is up and running, as shown in the above picture, one can just:

- **Type** someone's name, any term matching a concept or area in the multilingual knowledge graph, a company name etc.
- **Check** and **monitor** the resulting information connections across dimensions
- Further **refine** the search
- **Share** the results of the search
- In the case of retrieved documents, it is possible to **explore their content** and the most relevant concepts and entities extracted from them

We demonstrate SemanticPaths in two different use cases:

- a real-world application in which the user can search for **scientific articles**, **experts**, **research bodies** and **concepts** in two areas: the broader area of **Artificial Intelligence** for the Italian community (developed for the CINI AIIS lab) and the area of **Natural Language Processing**, specifically created for the ACL and EMNLP 2021 conferences. This application enables not only the monitoring of the research area, but also head hunting, that is, finding people relevant for a given job with the desired skills and expertise.

- A scenario in the **food industry** in which, starting from ingredients, the user aims to retrieve relevant dish recipes to cook and, vice versa, given certain dishes and/or recipes which ingredients and dishes are most relevant to them.

### **Challenges/perspectives**

Importantly, Babelscape's multilingual semantic search engine and monitor can be customized in any domain or area by just starting from a document base, which can be retrieved online or provided offline. However, two key challenges are ahead:

- 1) **Updating the knowledge graph continuously** based on updates of the underlying document base (e.g. new papers are published in the area of interest; new jobs are posted; new "heads" have to be indexed, etc.), and as a result the semantic search engine;
- 2) Integrate **new ways of showing the data and follow original paths** throughout (and while exploring) the data.

The first challenge is currently under development at Babelscape, thanks to continuous indexing and enrichment of the document sources. The second challenge is still in a preliminary phase and will see important developments in the upcoming months.

### **Links to papers and Web resources**

- The WordAtlas Web page: <https://babelscape.com/wordatlas>
- The SemanticPaths Web page: <https://babelscape.com/semantic-paths>
- SemanticPaths for NLP (ACL-IJCNLP 2021 conference):  
<https://acl2021.semanticpaths.org/>
- SemanticPaths for NLP (EMNLP 2021 conference):  
<https://emnlp2021.semanticpaths.org/>