OntoChem GmbH, Blücherstr. 24, D-06120 Halle (Saale), Germany

OntoChem Press Release

OntoChem launches SciWalker Studio Relationship Extraction software

OntoChem announces the release of **SciWalker Studio**, a new semantic software that allows its users to annotate named entities and to extract complex semantic relationships from any text-based document such as PDF, Office or html documents. The software can be installed locally on a PC or in the cloud, and is geared to extract data points of interest for life and material science. Data extraction can be fine-tuned by a relationship extraction editor using a simple to use browser-based interface.

The use of OntoChem's advanced context sensitive named entity recognition engine, **OC|processor** together with its 35+ comprehensive domain knowledge ontologies have enabled our customers to identify relevant hits in any scientific or legal documents. For example, **Google Patents** text and images have been annotated with our advanced chemistry annotation modules, enabling a comprehensive structure- and substructure-based search over patents and non-patent literature.

Identifying the relationship between named entities is key to a deeper semantic understanding of science and products on the market and in development. Thus, OntoChem has developed a **Relationship Ontology** that extracts more than 900 complex relationships (N-tuple relationships) between two or more named entities, going beyond currently used sentence based relationship extraction methods and knowledge triples as used currently by competitors. Examples of those extracted relationships are:

- compound-target relations
- target-disease relations
- compound-disease relations
- chemical reactions,
- drug combination treatments,
- biological activities of compounds,
- nested composition relationships of materials,
- and adverse drug reactions.

Ready made relationships can be downloaded as RDF files for upload into graph databases or columns-based databases such as Google's Big Query.

However, going beyond these predefined relationships, the user can use SciWalker Studio to create and edit new relationships of interest for his particular area of interest. A browser-based GUI is available that allows to define and test such new extraction rules. Ontologies in standard format can easily be added or removed. If the extraction rules are optimized, the **Process Manager** of SciWalker Studio can be used to extract from single documents or whole document collections on the user's local computer. A variety of output formats are available such as RDF, JSON or CSV - allowing to use the output of SciWalker Studio as input data sets to subsequent machine learning or data analytics tasks.