

Linguamatics Spring Users Conference 2015

Hilton Doubletree Hotel, Cambridge, UK

13-15 April 2015

Welcome address

John M. Brimacombe, Executive Chairman, Linguamatics

Mr. Brimacombe will welcome delegates and open proceedings with a high level review of Linguamatics' corporate mission - with a focus on the direct impact of the use of I2E on revenue, costs and risk. He will also highlight the growing maturity of the I2E platform as a Big Data solution.

Semiautomatic extraction of pharmacologic dynamic parameters from literature full text

Martin Baron, Scientific Information Professional, Roche Diagnostics

Pharmacological and dynamical (pk/pd) parameters are essential to quantify and compare the mode of action of drugs. We developed a workflow for the extraction of such parameters using text mining on full text or abstracts. Our results show that the majority of pk/pd parameters can only be found in full text but not in abstracts. By comparing the results from text mining to a commercially available curated database we show that both information sources yield a large amount of non-overlapping data and are therefore complementary. Challenges in text mining using full text and the impact on result generation will be discussed.

What's new in I2E?

Guy Singh, Senior Manager, Product and Strategic Alliances, Linguamatics

This talk will provide a summary of the new features introduced in the latest release, I2E 4.3. This includes major new features such as Connected Data Technology to enable federated text mining, a new output editor, advanced table processing and a number of other enhancements. The talk will conclude with a summary of content available through I2E OnDemand and a preview of upcoming additions planned for 2015.

I2E in life sciences: recent developments and use cases

Jane Reed, Head of Life Science Strategy, Linguamatics

In this era of big data, life science organizations face the challenge of filtering ever-increasing volumes of text information to gain actionable insights for key decision-making. I2E's flexibility means it can be beneficial in many applications and use cases. This talk will provide an overview of some customer use cases from a range of different disciplines, and highlight some of the solution areas where significant benefit has been found.

Patent analysis workflow - fighting the information overflow

Thorsten Schweikardt, Senior Information Scientist, Boehringer Ingelheim

Scientists from the Therapeutic Areas and Medicinal Chemistry spend a lot of time separating patents of relevance to a specific question from non-significant ones. We are developing our patent analysis workflow to prioritise the patents being reviewed in order to save the scientist's time and cut costs. Besides a standardized set of parameters which is in every analysis, the prioritization criteria can be defined freely for each project, making the workflow very flexible to use. We describe the general workflow using Linguamatics I2E and KNIME, and take a look at caveats and future enhancements.

Future innovations: R&D update

David Milward, Chief Technology Officer, Linguamatics

I2E is being applied to more diverse data to answer an ever widening range of questions. To support this, the platform is being made more open and configurable. One example is the development of a new readable query language with the full expressivity of I2E Pro. This will allow further automation of I2E querying, and translation from other query languages. Another example is a language processing API to allow I2E to be quickly extended to new languages. Finally we will show new developments in visualizing I2E results built on an extension server.

Extracting conclusions and interpretations from internal preclinical safety reports using I2E

Wendy Cornell, former head of the Proprietary Information and Knowledge Management group, Merck

Safety assessment studies are a key element of the drug development process and the results of these experiments help determine whether a compound will progress in the pipeline. Much of the data generated in these studies are captured in structured format, however, a significant amount of unstructured information is captured in written reports that describe expert conclusions and interpretations. This unstructured information represents a rich body of knowledge which, in aggregate, has potential to identify capability gaps and evaluate individual findings on active pipeline compounds in the context of broad historical data. We describe the development of a natural language processing (NLP) workflow to extract conclusions and interpretations from Merck's large corpus of internal reports using the Linguamatics I2E software and the integration and analysis of the data using the ANZO platform from Cambridge Semantics.

Using UIMA and open source tagging software to enrich the content that feeds Linguamatics I2E

Raul Rodriguez-Esteban, Senior Scientist, Roche

In order to improve the recall and precision of Linguamatics I2E results, we have set up a UIMA pipeline that enriches the content that is indexed by Linguamatics I2E. This pipeline annotates chemical, disease and gene names using machine learning-based algorithms developed by academic laboratories and which are available as open source software. With such approach we have managed to incorporate state-of-the-art technologies in our existing workflows allowing us, for example, to be able to identify names of entities that are not available in standard terminologies.

Easy access to full text articles for text mining: a new service, a new era

Chris Hilbert, Sales Solution Engineer, Copyright Clearance Center, Guy Singh, Senior Manager, Product and Strategic Alliances, Linguamatics

Researchers struggle to gain access to full text articles for text mining. When they do get the full text they must contend with multiple formats and inconsistent license terms – all of which inhibit text mining efforts. To address these issues, Copyright Clearance Center (CCC), parent company of RightsDirect, has partnered with Linguamatics to make it easier for I2E end users to obtain and index full text XML articles from multiple scientific publishers. This collaboration has led to the introduction of a new service to make high value data sources available for information analysis through text mining. In this presentation we will talk about CCC's new XML for Mining service, the integration with Linguamatics I2E and how the combined solution improves the results of text and data mining queries, reduces costs and mitigates infringement risk.

Leveraging I2E software for text mining patents on antibody-drug conjugates

Julia Heinrich, Senior patent analyst, Bristol-Myers Squibb

Information in patent publications provides a wealth of opportunities for identifying prior art, licensing deals, and white space. However, the ever increasing number of documents in mostly unstructured format (>95%) brings on the challenges of how to effectively, efficiently and economically extract relevant information in order to make time-sensitive, actionable legal and business decisions. Antibody-drug conjugates (ADC)/ immunoconjugates are therapeutic modalities backed by a large and historical portfolio of patents and are fueled to continue to grow by the recent Food and Drug Administration (FDA) approvals of Adcetris and Kadcyla. We have used the I2E text mining software to develop strategies and queries for parsing information from ADC patents into relationship columns in a spreadsheet in order to generate more user-friendly, analysis-ready data outputs. The user case of text mining intellectual property that claim ADC technology for specific antigens will be discussed in this presentation.

Linguamatics I2E in healthcare: recent developments and use cases

James Cormack, NLP Resource Specialist, Linguamatics and Phil Hastings, Vice President, Sales and Marketing, Linguamatics

The rapid growth of electronic health records (EHRs) provides an abundant source of valuable data, with the potential to discover insights about patients and their response to treatment. However, with up to 80% of the richest information within the unstructured text, hospitals and medical researchers need better ways to leverage this vital information. Natural Language Processing (NLP) can be used to extract information from the huge range of medical documentation, identifying relevant scientific literature, match patients to clinical trials and power clinical risk models. This talk will look at these and other use cases that enable I2E to add value to EHRs, and how Linguamatics is supporting this vital area.