Linguamatics

This application note describes how I2E has been used to capture valuable information from the Life Sciences literature, saving time, and increasing productivity.

I2E Application Note
Text Search & Mining for Safety/Tox

Using I2E to Identify Potential Safety Issues for Drugs at Specific Dosages

Challenge

Early identification of potential issues relating to safety is of crucial importance for pharma/biotech companies needing to optimize investment in R&D and avoid late stage failures. In this application note we show how a set of compounds can be comprehensively evaluated for side effects, adverse events and toxicity. In particular, these side effects need to be matched against specific tissues and the dosages at which they occur. The challenge is to mine available literature sources, both internal and external, to find and extract relevant information in a timely manner.

Linguamatics Approach

Linguamatics I2E can query and extract dosages, drug names, tissues and safety indicators from large document collections. Sources can include publicly available documents such as Medline, web or intranet pages, scientific articles, regulatory documents, study reports, news feeds, patent documents, etc. Queries can be defined using keywords and linguistic expressions. By plugging in ontologies, queries will automatically find synonyms or search for entire classes of items. Pre-defined smart queries can also be used; these are templates that hide complexity from the user by only exposing specific pre-defined options. In addition, queries can be combined to answer a set of questions simultaneously, e.g. for providing systematic profiles of compounds.

I2E presents the structured results in a choice of formats. These include web pages with results classified by drug, dosage and adverse event, Microsoft Excel spreadsheets, XML files, or as network graphs to allow the user to visualize direct and indirect relationships between entities. Results can also be presented in formats suitable for export to databases.

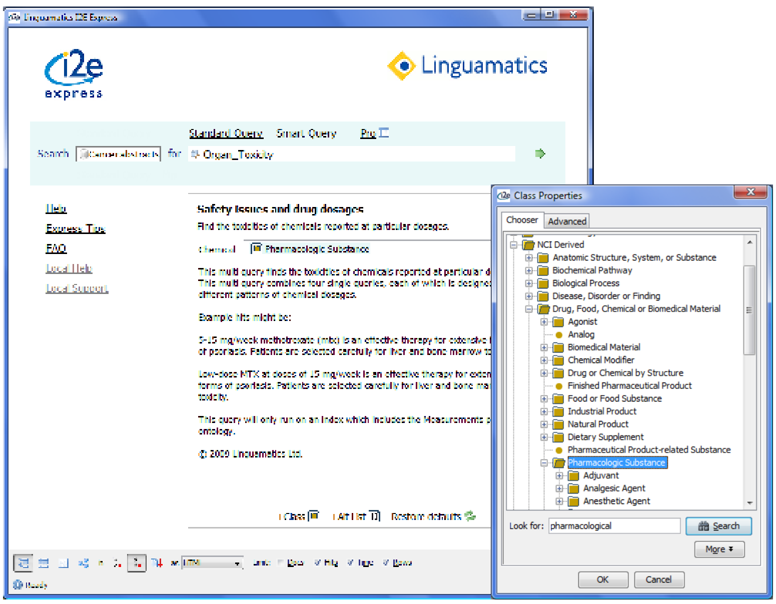

Figure 1 shows a search strategy can be built in the query editor and then presented as a smart query. Selecting a compound for the query is as simple as picking an item from a tree, or the user can just highlight one of the folders to select an entire compound class.

Method

In this example, a smart query has been written to address this specific problem (Figure 1). This query allows the user to either specify a particular drug or a class of compounds using the class chooser, or to provide a set of keywords. The user can also control the number of results to be returned, along with their preferred results format.

A more experienced information scientist may wish to explore and if necessary modify the components that make up the smart query. In this way, they can customize the query to extract additional information that is tailored to their own specific requirements.

The individual queries contain precise search terms used to identify key information. For example, to find information about adverse events in the liver, there are two sets of terms: one set looks for liver-specific terms; the other looks for toxicity-specific terms. These queries can return the same document hit(s), for example, "hepatoxicity" will match both terms. The term "hepatotoxicity" was matched against two wildcard matches: "hepato*" (one of the liver-related terms; others included "liver" and "hepatic") and "*toxicity" (one of the adverse event related terms; others included "side-effects", "adverse events", "sensitivity").

Results

Figure 2 shows the typical output from this smart query as a web page containing evidence linking adverse events to compounds at particular dosages. It is also possible to cluster the results by compound, dosage and tissues of interest to provide a rapid overview prior to more exhaustive analysis. Links are provided to highlighted evidence in relevant documents.

As can be seen, standardized names are presented for the compounds; these allow the user to systematically see information that covers different synonyms for a particular drug (one example is Cyclosporine: additional names that will match this drug include ciclosporin, CsA, Neoral, among others). The data can be re-ordered on the fly, so that the results are ordered by frequency (in Figure 2: the most common assertions are reported first), alphabetically or on a document-by-document basis.

The dosage column provides dosage information extracted in context. I2E recognizes different expressions for dosage, e.g. "10mg/kg/day", "5mg bid", "5mg rofecoxib per day", "100mg every day", and others.

The information has been extracted and converted from free text in an abstract into structured data in a table. This structure allows the results to be exported into other software packages or processes. For example, from within I2E it is possible to redisplay results directly in Excel and Cytoscape, a network graph viewer.

Linguamatics

Pharmacologic Substance	Safety issues	Tissue	Dosage		Doc		Hit
Cyclosporine	Safety issues	Kidney	15 mg/kg/day	6	17497475	11	... of hyperbaric oxygen on cyclosporine-induced nephrotoxicity and oxidative stress in rats... : a control group, a cyclosporine group (15 mg/kg/day intraperitoneally for 14 days)...
Mycophenolate Mofetil	Safety issues	Kidney	2 g/day	2	16041270	3	... /day of everolimus or 2 g/day of MMF, plus full-dose cyclosporine (... Concerns over nephrotoxicity led to a protocol amendment ...
Amphotericin B	Safety issues	Kidney	0.4 mg/kg/day AMB	2	15761070	8	Low nephrotoxicity of an effective amphotericin B ... for 10 consecutive days with 0.4 mg/kg/day AMB in the form of traditional ...
Everolimus	Safety issues	Kidney	everolimus 1.5 mg/day	3	16041270	3	Concerns over nephrotoxicity led to a protocol amendment ... loss to follow-up) were everolimus 1.5 mg/day, 33.7% (65/...
Lamivudine	Safety issues	Liver	100 mg/day	5	17283489	1	Lamivudine (100 mg/day) was continued throughout the ... Liver Transplantation adverse effects
Alcohol	Safety issues	Liver	20 g/day	1	15553597	7	Insulin sensitivity and hepatic steatosis in obese subjects with ... analyzed 86 obese patients whose alcohol intake was less than 20 g/day and who showed no signs ...
Gentamicin	Safety issues	Kidney	100 mg/kg/day	2	14748758	6	... 4-hydroxy tempo) on gentamicin-induced nephrotoxicity in rats. ... The rats were given gentamicin (100 mg/kg/day, i.p., once a ...
Sirolimus	Safety issues	Kidney	1 mg/day	2	16364861	2	... primary immunosuppressant in calcineurin inhibitor-induced nephrotoxicity. ... Sirolimus was started at 1 mg/day with titration over 2 weeks ...
		Liver	sirolimus 2 mg/day	1	15899725	2	... a 6-mg loading dose of sirolimus, participants received sirolimus 2 mg/day for at least 7 days... dropped out because of trimethoprim-sulfamethoxazole-related hepatotoxicity.
Indinavir	Safety issues	Kidney	IDV/r 800/100 mg bid	1	17263634	4	... started zidovudine plus lamivudine plus IDV/r 800/100 mg bid. ... was a surprising lack of nephrological side effects during the 6 months of ...
Atorvastatin	Safety issues	Liver	10 mg/day	2	17473378	3	... a potential role in statin-related adverse events, and withdrawal of ... in patients developing myotoxicity or liver toxicity. ... Twenty-six patients with hypercholesterolemia received atorvastatin at 10 mg/day for 3 months.
			atorvastatin 10 mg/day	1	16731999	1	... benefit beyond that afforded by atorvastatin 10 mg/day in patients with stable coronary ... in the rates of treatment-related adverse events and persistent elevations in liver enzymes.
			atorvastatin 80 mg/day	1	16731999	1	... that intensive lipid-lowering therapy with atorvastatin 80 mg/day provides significant clinical benefit beyond ... in the rates of treatment-related adverse events and persistent elevations in liver enzymes.
Acetaminophen	Safety issues	Liver	1000 mg/kg/day	1	15027815	7	... inhibitors in the prevention of hepatotoxicity after paracetamol overdose in rats... The appropriate doses of paracetamol (1000 mg/kg/day) and the 'inhibitor ...

Figure 2 shows example results from this smart query. The search time, documents searched and number of results are all reported together with evidence from document hits and links to each document.

Summary

A systematic query strategy as described above gives confidence that decisions are being supported by high quality extracted information. The speed of search enables the user to have rapid access to new insights much faster than by traditional methods. Querying is highly flexible and queries are easily modified and shared, for example to compare results from alternative strategies.

Dosage information, expressed in many ways, is extracted directly into the results table. This also applies to other numerical data, such as concentrations, temperatures or even amino acid numbering. This last example is very useful for linking toxic effects to particular genotypes and protein mutations.

Smart queries can be created by expert information scientists as part of a standard search strategy. Once built, they enable less expert users to rapidly execute sophisticated searches, tailored to their specific requirements, on a variety of literature sources.

Linguamatics

About Linguamatics

Linguamatics is the world leader in deploying innovative natural language processing (NLP)-based text mining for high-value knowledge discovery and decision support. Linguamatics I2E is used by top commercial, academic and government organizations, including 17 of the top 20 global pharmaceutical companies, the US Food and Drug Administration (FDA) and leading US healthcare providers. I2E can be used to mine a wide variety of text resources, such as scientific literature, patents, Electronic Health Records (EHRs), clinical trials data, news feeds, social media and proprietary content.

Linguamatics is committed to excellence in healthcare informatics and is a corporate member of AMIA and HIMSS. The company operates globally, with headquarters in Cambridge, UK, and a U.S. office in Westborough, MA.

Linguamatics is a winner of the Queen's Award for Enterprise 2014 for International Trade.

For further information, visit **www.linguamatics.com**

About I2E

I2E is an agile, scalable, high performance text mining system that facilitates discovery and knowledge synthesis from unstructured text in large document collections.

I2E has a proven track record in delivering best of breed text mining capabilities across a broad range of application areas. Its agile nature allows tuning of query strategies to deliver the precision and recall needed for specific tasks, but at enterprise scale.

There is a choice of ways in which you can connect to I2E's unique capabilities: either by deploying I2E Enterprise in-house, or via I2E OnDemand, our Software-as-a-Service (SaaS) version of I2E.

For more information, visit **www.linguamatics.com** or **www.whatistextmining.com**

For more information
or a demonstration,
please contact us
+44 (0) 1223 651910
enquiries@linguamatics.com

THE QUEEN'S AWARDS FOR ENTERPRISE: 2014

Linguamatics, 324 Cambridge Science Park,
Milton Road, Cambridge, CB4 0WG, UK
Tel: +44 (0)1223 651910

Linguamatics, 1900 West Park Drive, Suite 280
Westborough, MA, 01581, USA
Tel: +1 617 674 3256

www.linguamatics.com