Powering Knowledge Discovery

Insights from big data with Linguamatics I2E
The world now generates an overwhelming amount of data, most of it written in natural language, as unstructured text.

Hidden within all these data are business-critical facts and relationships — insights which could give your organization a significant competitive advantage.

I2E from Linguamatics will help you find those insights, fast.

I2E is an agile and interactive text mining platform for the extraction and analysis of information. Text mining goes beyond keyword search by extracting key facts and relationships. This means you don't have to read the whole document to find what you need. I2E lets you answer questions and discover knowledge that would be impossible to find using any conventional method.

It's hard to overestimate how valuable this can be. To take just one example: many drug development projects fail at a late stage due to safety concerns. It has been suggested that more than half of these failures could have been predicted, prior to development, had the literature been searched with I2E thereby potentially saving millions of dollars.

Gain actionable insights from unstructured data
Beyond R&D, the identification of key opinion leaders and analysis of public sentiment can be critical for the development of brand marketing strategies and the management of product life cycles.

Tasks like these are easy with I2E. Its powerful natural language processing (NLP) capabilities go way beyond keyword searches; instead it lets users mine large, complex collections of documents – unstructured and semi-structured sources like scientific literature, patents, electronic health records (EHRs), clinical trials data, news feeds, social media and proprietary content and more – to extract relevant facts, relationships and entities.

And it delivers those results in the right format to support your decision-making.

Information everywhere is growing exponentially; there are over 20+ million abstracts in the MEDLINE biomedical database and over 13 million registered patents in the US alone.

I2E is the most effective way to mine those resources for critical information and hidden connections.

Linguamatics I2E is a flexible, scalable knowledge discovery platform. It delivers high quality structured results, fast. I2E can take input from documents and unstructured text, apply domain knowledge and the metadata for those data sources, and output extracted facts in table or graphical formats, for actionable decision support.

High value insights from text through:

◆ Rapid, flexible text analytics to support complex, high value problem solving
◆ Semantic information extraction
◆ Systematic and reproducible analysis
◆ Innovative federated text mining architecture to enable querying across data silos
I2E was developed by a team with specific expertise in information extraction from content-rich environments, from life science research and business intelligence.

I2E is particularly strong in its ability to answer a wide range of questions, from apparently simple open queries to questions that need advanced linguistic analytics. For example:

- What treatments are being used for breast cancer?
- What positive comments have been made by a group of opinion leaders about my product?
- What's the smoking category of my patients?
- Do I have freedom to operate for my chemical lead series?

By providing interfaces for both occasional and expert users, I2E addresses the needs of people across the enterprise. Optimized queries can be saved and published for use by less experienced users, thus leveraging expert knowledge. Validated queries can then be applied repeatedly; for example, to run regular queries against a growing database of literature.

Searching in context

It’s easy to plug domain-specific knowledge into I2E using thesauri, taxonomies or ontologies which describe the concepts and relationships in any given field.

The system is equally able to answer the questions that come naturally to a life scientist or clinical researcher – interested, perhaps, in biological targets, novel chemistry, biomarkers, pathways and diseases – as it is those of the business analyst, who wants to know the trends and influencers in a particular market.

Versatility

I2E offers a spectrum of capabilities that can be applied flexibly to address different text search and mining challenges. Strategies range from simple document retrieval using keyword search to advanced information extraction, including the retrieval of facts, relationships and entities using linguistic analysis. Users match the approach to their specific task and combine it with complementary techniques to get the answers they need.

Get right to the answers

I2E gets straight to high value, relevant, re-usable knowledge. Tabular reports summarize and structure the extracted information in a compact, clustered format with easy drill-down to supporting evidence, including direct links to source documents. Results can be visualized by tables or graphs enabling rapid analysis; or exported directly into other visualization analytics tools such as Microsoft Excel, Spotfire or Cytoscape. Thus, researchers and information professionals get the flexibility and analytical rigor that they need.
Connected Data Technology

Linguamatics Connected Data Technology uses an innovative federated text mining architecture, integrated into the I2E text mining platform. The technology allows a text mining query to be run simultaneously over multiple data sources, whether they are located locally or on the cloud. The results from the different data sources are integrated together as a single result set that is sorted and clustered for faster analysis and better decision support.

Flexible and easy to use

I2E is highly user-centric. The I2E Express interface is easy to use and similar in appearance to a conventional search engine. For more advanced searches, I2E Pro lets the user view, construct, and manage sophisticated queries using an intuitive drag and drop graphical interface.

Quality and accuracy

I2E has powerful linguistic capabilities. Searching with concepts and classes (e.g. genes, proteins) provides flexibility and wildcard features allow users to ask open questions and search for entities, verbs or unknown relationships. Searching using document position information can identify stronger relationships by finding terms that appear together: for example, within a proximity of one or more words; or in the same sentence; or in a particular region of a document.

I2E can filter results based on the broader context of the document, not just on the keywords. More distant and indirect relationships can also be detected by combining result sets or visualizing information networks, to find relationships across multiple documents and synthesize new knowledge.

Performance and robustness

I2E is scalable to millions of documents and terabytes of data – the whole of MEDLINE, for example – and works with large documents such as patents.

I2E is also able to combine large numbers of search terms together into concepts. For example, classes of genes, proteins, diseases or compounds, perhaps containing up to tens of thousands of terms, can all be encapsulated within a single “class” search term.

Search performance is fast. Customers report that I2E has been shown to be at least ten times faster than conventional keyword search-based approaches for the discovery and analysis of relevant information.

Control and interactivity

I2E makes NLP-based text search and text mining intuitive, but also interactive. Its flexibility, transparency, speed, and the way it delivers structured results all mean that users can refine queries on-the-fly.

I2E puts power into the hands of the users.

Integration

I2E can play a key role when embedding information extraction into more complex workflows. The platform incorporates a comprehensive RESTful Web Services API (WSAPI) that gives access to the majority of I2E functionality. The WSAPI makes it much easier to create your own applications with features like custom interfaces, visualization of results and ad hoc indexing. I2E can easily be integrated with workflow software (such as Pipeline Pilot or KNIME) and results can be exported in many data formats that are compatible with standard enterprise software systems.
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

To fully evaluate the unique and compelling benefits that I2E can bring to your organization, please contact your local Linguamatics representative or email us at enquiries@linguamatics.com.