

## Practical applications of linguistic technology

compiled by Dagobert Soergel

**Reading 3** (required). Note: This TOC also provided as a separate file to view side-by-side. Look over these pages to get a general idea and dig deeper if something piques your interest.

This is a compilation of materials on the LUXID system (and its forerunner Insight [not to be confused with Inxight]) produced by TEMIS Corp., and on the SAP BusinessSystems ThingFinder system. ThingFinder originated at Inxight corporation, a spinoff from Xerox PARC (Palo Alto Research Center of mouse and the MAC and Windows interface fame); Inxight was bought by BusinessSystems, which was bought by SAP, a major business software vendor.

Both systems extract information from text. They process text to extract entities and statements that connect entities through relationships. They also assign metadata to documents, in other words, they assist in (or take over) cataloging. The marketing hype says that these systems convert unstructured data (text, better called data with complex structure) into structured data, entity-relationship statements often represented in tables and graphs, structures that are simpler and easier to search than text.

The compilation includes old documents from Inxight because they give better examples and a much better explanation of the process.

<b>Overall framework and broad applications</b>	
• Luxid 6 Content Enrichment Platform Brochure (Luxid 7 not available)	3 - 8
• Luxid 7 press notice	9 - 11
• OECD chooses TEMIS to semantically structure knowledge	13 - 14
<b>How it works. A first glimpse</b>	
• Luxid. From text to meaning. One slide	15
• Inxight LinguistX platform	17 - 18
• Inxight. Linguistics: Adding value to e-publishing and e-content P. 17 - 20 is general hype, p. 21 - 24 is useful technical information	19 - 27
<b>A little more detail on how it works</b>	
• Inxight SmartDiscovery Extraction Server	29 - 32
• Inxight ThingFinder Server	33 - 34
• Fact sheet. TEMIS Skill Cartridge Biological Entity Relationships	35 - 36
• Statistical Translation WashPost20110222Health&ScienceP1	37 - 40
• Google switches to its own [statistical] translation system	41

**Reading 3a** (optional) gives more detail. Only for those really interested in this topic

<ul style="list-style-type: none"> <li>• Open source NLP software can help. A representative sampling of open source software in a nice NLP flowchart</li> </ul>	45 - 50
<ul style="list-style-type: none"> <li>• Transforming Unstructured Text into Actionable Data - slides, esp. p. 56 - 59</li> </ul>	51 - 68
<ul style="list-style-type: none"> <li>• ThingFinder Concepts Guide. A detailed description how ThingFinder works and a good introduction to natural language processing (NLP) in general. Start at p. 47 The XeLDA paper (next) may be even better.</li> </ul>	69 - 106
<ul style="list-style-type: none"> <li>• XeLDA: integrate a linguistic engine in your applications Another good introduction to natural language processing (NLP) with many illustrations. Perhaps even better than the ThingFinder paper</li> </ul>	107 - 141
<ul style="list-style-type: none"> <li>• Jaime Carbonell. Natural Language Processing 33 slides with very nice examples. Good overview</li> </ul>	143 - 175
<ul style="list-style-type: none"> <li>• Libby, Elizabeth. Natural Language Processing. Article in the Encyclopedia of Library and Information Science</li> </ul>	177 - 190
<ul style="list-style-type: none"> <li>• Inlight SmartDiscovery Awareness Server Interesting product that uses linguistic technology to post-process Google search results; for examples, extracts information from text into a relational database. Not sure this product still exists</li> </ul>	191 - 194
<ul style="list-style-type: none"> <li>• Going beyond Google. Another document on the same system</li> </ul>	195 - 199
<ul style="list-style-type: none"> <li>• <b>Several TEMIS products (from the Luxid or the earlier Insight suites) for illustration of applications</b></li> </ul>	
<ul style="list-style-type: none"> <li>• Luxid in the medical domain. 18 slides</li> </ul>	201 - 218
<ul style="list-style-type: none"> <li>• Analyzing patent literature to gain competitive insight with Luxid (medical domain) 42 slides</li> </ul>	219 - 260
<ul style="list-style-type: none"> <li>• Insight Discoverer Extractor short</li> </ul>	261 - 262
<ul style="list-style-type: none"> <li>• Insight Discoverer Extractor long</li> </ul>	263 - 266
<ul style="list-style-type: none"> <li>• Insight Discoverer Categorizer short</li> </ul>	267 - 268
<ul style="list-style-type: none"> <li>• Insight Discoverer Categorizer long.</li> </ul>	269 - 273