

ie

Table of contents

1 NLP Tools for Information Extraction.....	2
2 Commercial Tools.....	2
3 Academic Tools.....	3

1. NLP Tools for Information Extraction

2. Commercial Tools

- PubGene
 - [PubGene Inc.](#)
 - [PubGene Org](#)
- [Ariadne Genomics](#)
 - [MedScan](#) extracts functional associations between proteins, cell processes and small molecules, recognizes types of regulatory mechanisms involved, the effects of regulation and more. Captured data is presented as a datasheet, an XML file or a pathway diagram
 - [Pubmed Search by Protein Name](#)
- [LSGraph](#) (IT.Omics) : bioanalysis tool that mines and integrates functional relationships between biological entities from the worldwide scientific literature - the bibliomics
- Biovista: support research-oriented analyses of scientific articles helping you understand complex relationships between entities such as diseases, pathways, drugs, experimental methods, scientists involved and reagents used.
 - [Journal Research Analyst](#)
 - [Biolab Experiment Assistant](#)
- OmniViz:
 - [Document Analysis](#)
 - [Pathways](#)
 - [RefViz](#) Explore Research Literature Visually
- [ClearForest Products](#)

ClearForest's advanced text-driven business intelligence solutions apply intelligent mark-up to key entities such as person, organization, location, as well as detailed facts or events embedded within free-form text such as news articles, web surveys and HTML documents. Once structured, this information can be used to drive stand-alone analytics applications or be fed into a company's existing data marts and combined with structured data to provide more comprehensive business intelligence.

- [BiblioSphere](#) (Genomatix)

BiblioSphere is a data-mining solution for extracting and studying gene relationships from literature databases and genome-wide promoter analysis. BiblioSphere contains literature

data-mining strategies using more than 350,000 quality checked gene names, synonyms and Genomatix proprietary semantic relation concepts. Based on PubMed, BiblioSphere currently searches over 14 million abstracts. The unique data-mining strategy allows to find direct gene-gene co-citations and even yet unknown gene relations via interlinks. BiblioSphere data is displayed as 3D interactive view of gene relationships. Results can be classified by tissue, Gene Ontology and MeSH. Statistical rating by z-scores indicate over- and under-representation of genes in the referring biological categories.

- [QUOSA](#)
 - commercial, launched late 2002
 - establish local paper collection by downloading
 - prioritizes full-text papers during search
 - available to hundreds of researchers in two US hospitals

3. Academic Tools

- [Chilibot](#) (University of Tennessee Health Science Center)
- [microGENIE](#) (Vrije Universiteit Amsterdam)
- [iProLINK](#) -- Integrated Protein Literature, Information and Knowledge
- [XplorMed](#) : allows you to explore a set of abstracts derived from a MEDLINE search. The system gives you the main associations between the words in groups of abstracts. Then, you can select a subset of your abstracts based on selected groups of related words and iterate your analysis on them. XplorMed is recommended for cases in which you do not know exactly what are you expecting to find. Your interests may be modified by the results obtained, or you may want to enquire new questions as the analysis develops. Also, the results may suggest you additional words that should be used to expand your query in MEDLINE (e.g., unexpected abbreviations of a protein name, or synonyms of a disease).
- [AbXtract server](#): the system extracts information from the MEDLINE abstracts. Relevant keywords are selected according to the difference between their frequency in the family object of analysis and their frequency in other unrelated protein families.
- [Kinase Pathway Database](#)
- [BioRAT](#): search engine and information extraction tool for biological research
- [MedLee](#) : a tool to extract, structure, and encode clinical information in textual patient reports
- [Genescene](#) : tool that allows users to navigate the relations extracted from abstracts.

- [Suiseki](#) : system for the extraction of protein-protein interactions from large collections of scientific text. It combines the statistical analysis of protein interactions, the analysis of the syntactical structure of the phrase, and a frame-based module dedicated to the detection of protein and gene names.
- [MedMiner](#) filters will extract and organize relevant sentences in the literature based on a gene, gene-gene or gene-drug query. This tool combines the GeneCards and PubMed search engines with user input and automated server-side scripts in an integrated text filtering system