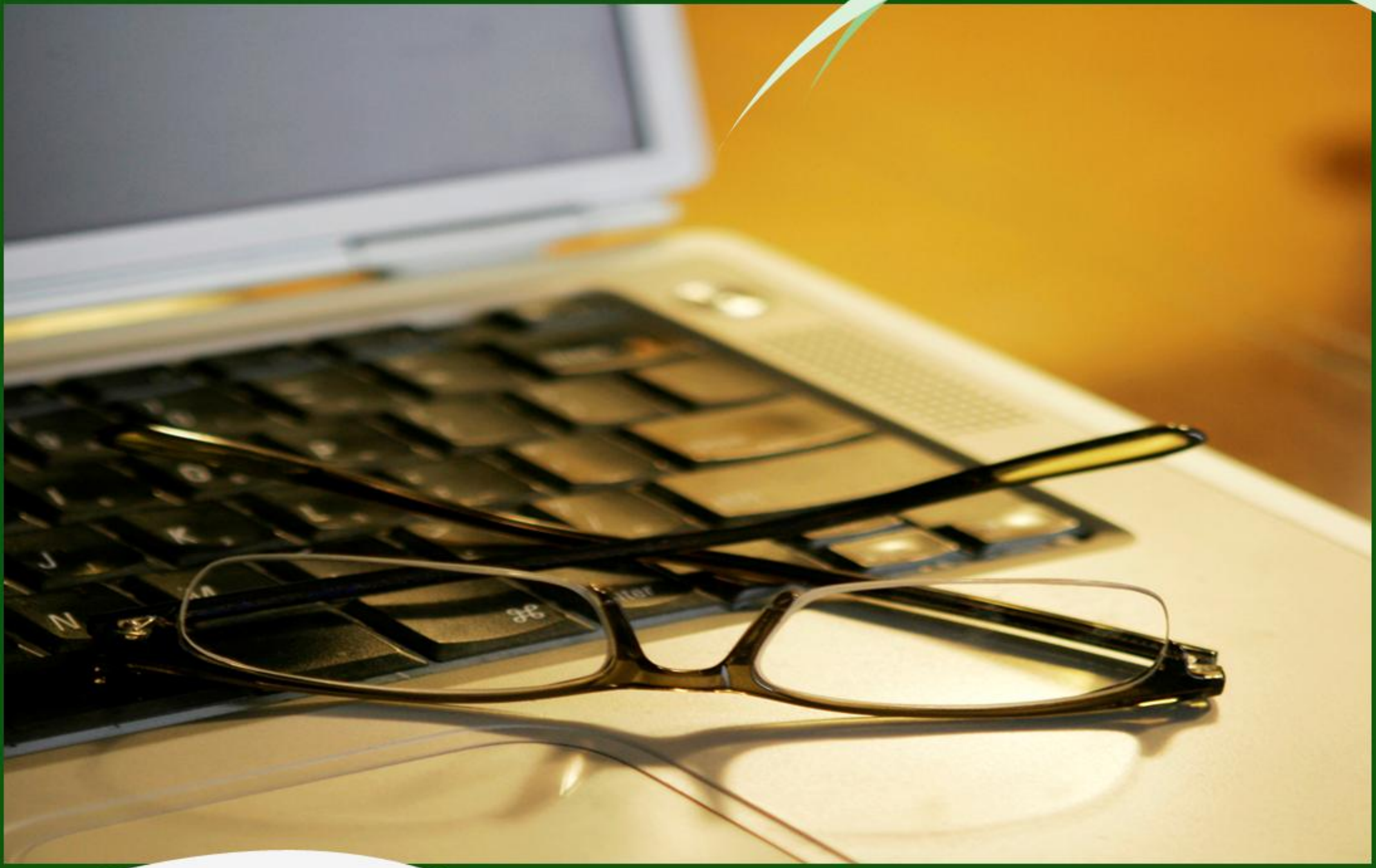
ProContent

We Translate Data Into Knowledge



- About Us
- Our Products
- Technology
- Knowledgebases
- Software
- Collaboration

Share your knowledge with the others!

Knowledge is one of the most valuable things in the world, and information flow, especially in life sciences, is overwhelming indeed.

To make your knowledge intelligible to the others it has to be properly organized.

It is a rather complex issue, but we offer a solution - the Knowledge-based format which will make your vision of a scientific problem assessable to everyone.

About Us

ProContent is a dynamic private research company with a primary focus on IT and bioinformatics.

We develop innovative custom-designed software tools for statistical analysis of the sparse data sources.

Data appears as a result of interaction of users and telecommunication devices.

In developing the software ProContent focuses primarily on personal profiles of the Web-users with their further implementation for informational mining within the World Wide Web.

The proprietary data-mining algorithms foster the knowledge retrieval in the course of Web-crawling and enable to present this knowledge in the clearly perceptive form of the BiblioBase.

With Neosemantic™ Technology, ProContent is committed to development and application of software and workflows for intelligible life science research.

Promoting the knowledge-driven paradigm shift in bioinformatics, the Company currently employs 15 associates.

OUR PRODUCTS

ProContent offers a series of knowledgebases and software for efficient data-mining on-line.

The associative search enables to unlock the unexpected ideas concealed under your everyday hard work.

The unique Neosemantic™ Technology enables on-line knowledge accumulation within precompiled or self-made bibliobase environment.

We provide programming interface to the most intensive data-producing software packages, available for genomic, proteomic and transcriptomic research.

The relevance between general research goals is graphically depicted in relation to the personal expertise of lab workers to steer the knowledge-driven discovery process.

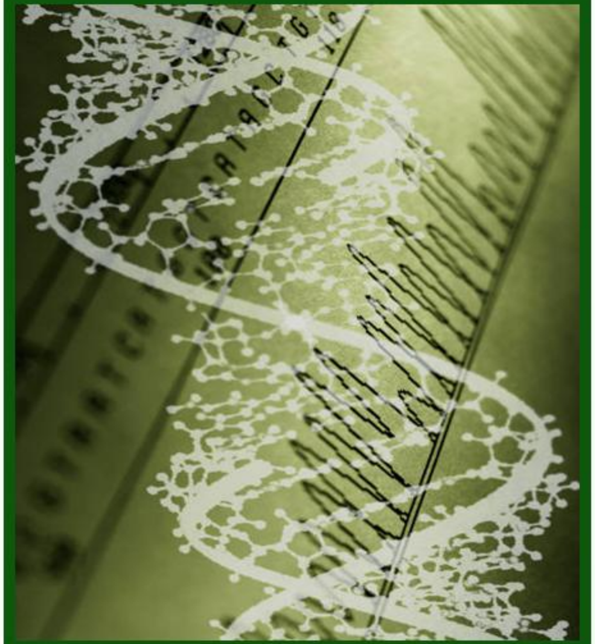
Technology

Neosemantic™ Technology comprises the paradigm shift in postgenomic research. It is committed to exploit that information, which is hidden behind the Web-access logs. By unlocking the value of Web-usage logs, the Technology enables translation of the wealth of experimental data into the knowledge-based format, which than becomes accessible to all the contributors.

Technology relies on two components:

- connectivity model, which algorithmically describes the procedure of establishing relations between any two information entities, encountered within the knowledgebase scope.
- associativity model, which ascribes the semantic relationships between entities by comparing the access patterns of Internet usage.

KNOWLEDGEBASES



The knowledgebase consists not only of the individual informational entities, but also comprises the relations between them.

The most important relations between entities are established based on the personal information.

The knowledgebase enables to discharge privacy in favor of intelligible knowledge sharing.

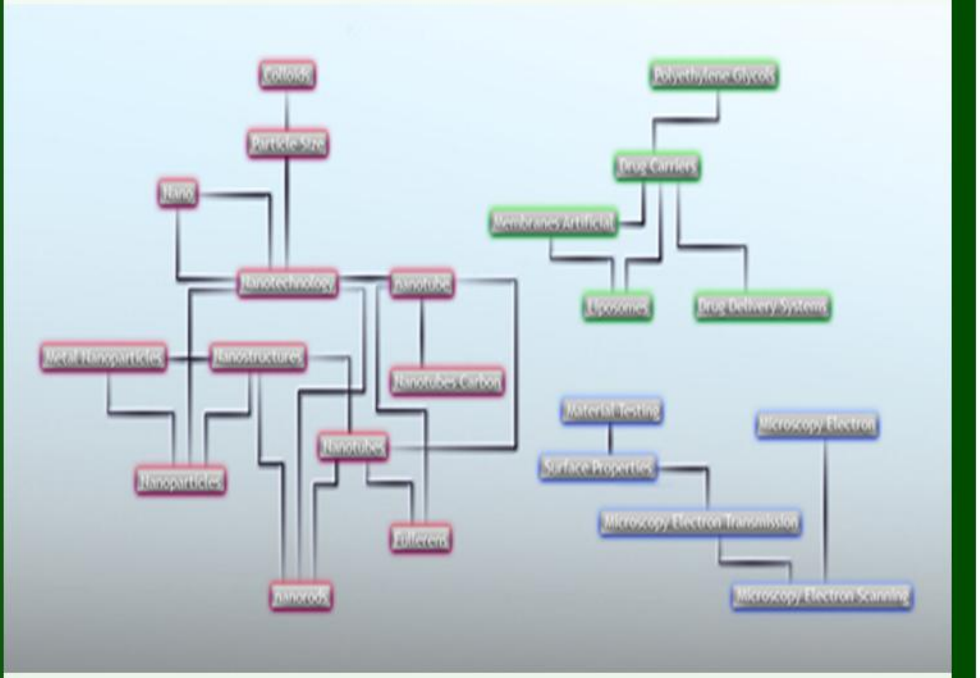
Nano BiblioBase

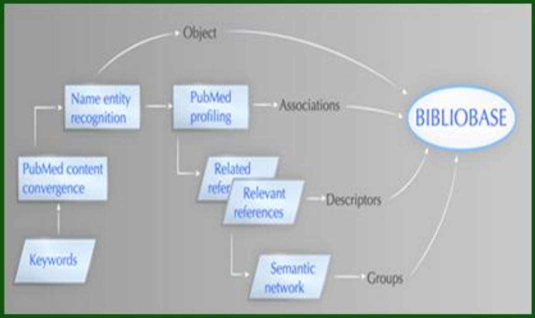
The «Nano» BiblioBase is a knowledge-based information retrieval system, with a special emphasis on nanobiotechnology. This is a comprehensive, integrated platform comprising all major R&D strategies that allow investigating the influence and applicability of nanomaterials for biological systems.

The core of the knowledgebase was automatically created by mining the PubMed for the following information:

- names of genes and/or proteins, involved into the respond to the exposure for nanoparticles;
- names of chemical substances, including drugs, which emerge in relation to the nanotechnology applications;
- different types of nanomaterials most extensively investigated for their therapeutic or pathological impact on animal models.

“Nano” BiblioBase at a glance





The development of more efficient sequels for the marketed drug is tightly linked with the research in pharmacology of drug transport.

The recent 25 years of the trial and errors in this direction comprised the compendium, called the «Drug Delivery» BiblioBase.

We assembled over 1 million of published papers to provide you with the efficient tool for navigating through this enormous amount of experimental reports.

Analysis of publications in drug delivery for the recent 15 years revealed that research trend in this field have reached saturation by 2005, and since than no significant increase in the number of publications has been observed.

Drug Delivery BiblioBase

The scope of **«Drug Delivery» BiblioBase** covers: the biotechnology to produce transportable compositions, including high pressure homogenization, sonication, extrusion, microfluidization, filtration etc. Mining PubMed both in automated and in expert-driven mode the assemblage of **2500** chemical substances was compiled, resembling any substance mentioned in the context of drug delivery.

The knowledge convergence coefficient (KCC) achieved the remarkable benchmark at the level of 0.56, making «Drug Delivery» BiblioBase the source of knowledge-driven hypotheses to steer the design of new therapeuticals.

Human Proteins & Genes BiblioBase

This bibliobase addresses the deliverables, coming from the large scale postgenomic projects:

- Human Genome Project;
- HapMap and ENCODE;
- Human Proteome Project (upcoming HUPO initiative).

The names of genes and proteins were collected from the UniProt knowledgebase and queried against:

PubMed for associating each protein/gene with a respective bibliometric profile;

On-line Mendelian Inheritance in Man and dbSNP for attributing polymorphic genes to diseases;

ProteinAtlas for mapping proteins to existing antibodies.

The semantic schema for the human protein and genes indicates the major types of tools used for high-throughput research: arrays, mass-spectrometry, blotting etc.

Semantic Schema of Human Proteins & Genes BiblioBase

The knowledgebase convergence coefficient was relatively low (0.34), as well-characterized genes comprise less then 25% of human genome.

Our Software

ProContent offers a broad range of software solutions built around the data-mining and text-mining systems.

Our software connects the user-specific research interests and achievements to public sources, comprising scientific papers, sequence homologues, transcribed genes, expressed proteins, diseases and pathways.

BiblioEngine

The BiblioEngine provides the set of tools to produce your own bibliobases using Neosemantic™ Technology. It is a package of programs and shells designed to perform necessary data- and text-mining.

Customizable solutions developed by our partners

The BiblioEngine Toolkit

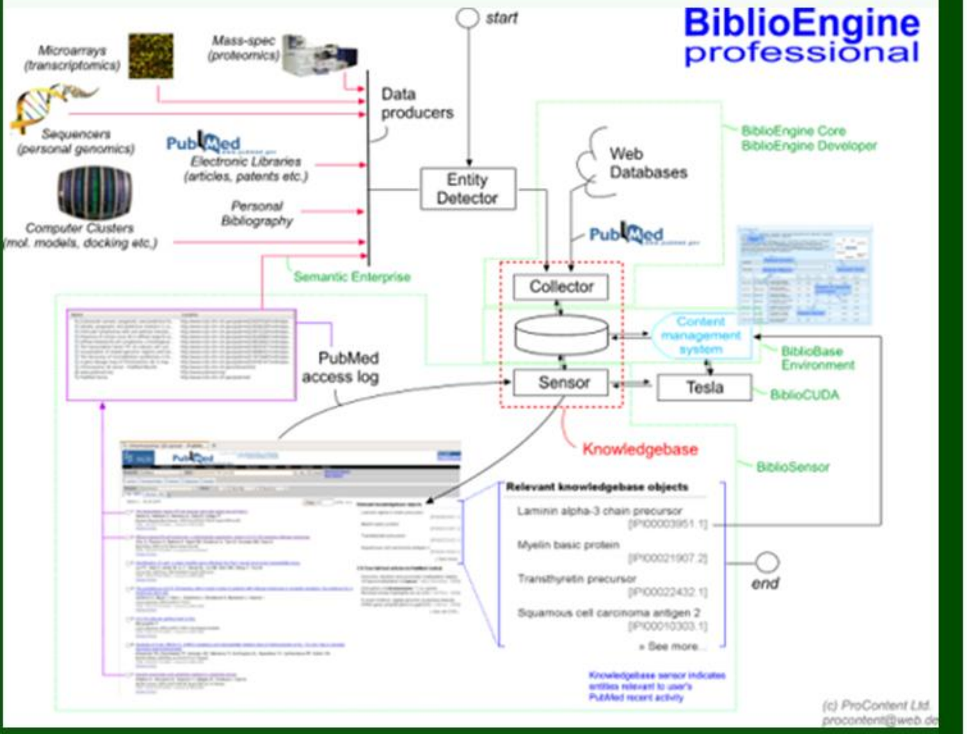
The most flexible solution supplied as a package of open-source Perl scripts to develop a proprietary pipeline for gathering the data into the knowledgebase.

The BiblioEngine Professional

Web-interface for convenient management of the knowledgebase creation process.

Proteome Organizer

Special design of the BiblioEngine for permanent uploading the data from the high-throughput proteomic project or from PRIDE / Peptide Altlas into the BibioBase CMS.



BiblioShelf

The BiblioShelf is a Web-based application to manage the collection of bibliobases.

The newly developed bibliobase is registered at the BiblioShelf server to become accessible to the Bio-Internet surfers via the BiblioSensor toolbar.

The BiblioShelf enables to organize the central knowledge sharing resource. The algorithm for embedded pattern recognition is used to perform the comparison of knowledgebases.

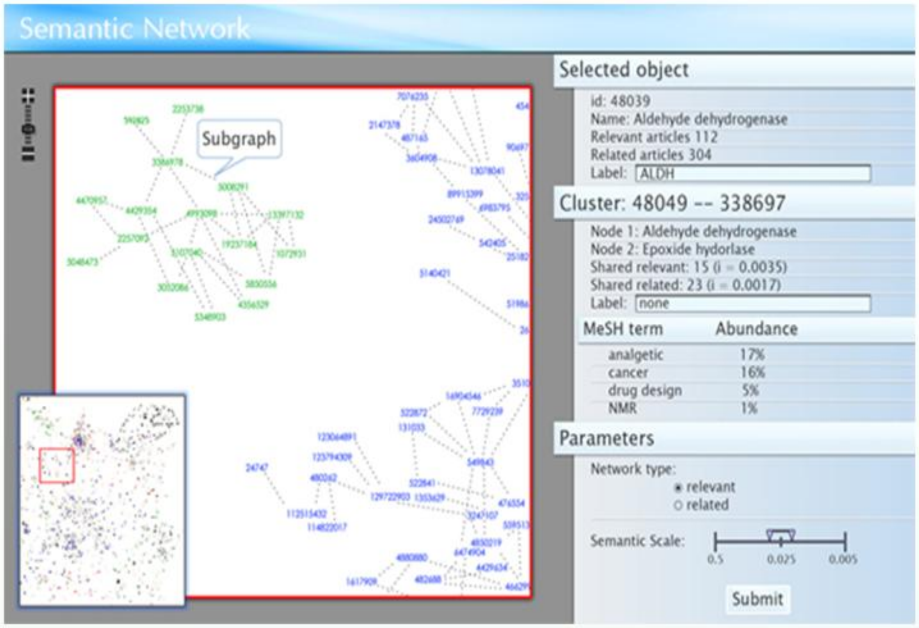
The comparison provides the unbiased insight into the relationship between different domains of knowledge.

Currently there are three bibliobases on the BiblioShelf:

«Nano» - current overview of toxicity issues, related to nanomaterials

«Drug Delivery» - a tool to conduct hypothesis-driven search for more efficient and safe drugs

«Human Proteins & Genes» - to be informed of the latest achievements in the postgenomic science.



Selected object

id: 48039
Name: Aldehyde dehydrogenase
Relevant articles 112
Related articles 304
Label: ALDH

Cluster: 48049 -- 338697

Node 1: Aldehyde dehydrogenase
Node 2: Epoxide hydorlase
Shared relevant: 15 (i = 0.0035)
Shared related: 23 (i = 0.0017)
Label: none

MeSH term	Abundance
analgetic	17%
cancer	16%
drug design	5%
NMR	1%

Parameters

Network type:
● relevant
○ related

Semantic Scale: 0.5 0.025 0.005

The BiblioBase Content Management System is build using Drupal developer's package to host the knowledgebase. Following features are supported by the BiblioBase CMS:

- uploading and Web-publishing of the archived text/data mining output as it comes from the BiblioEngine;
- calculation of the associations between informational entities according to the user-defined models, including sequence homology and shared papers;
- semantic search to capture the relation of the user’s query with the knowledgebase;
- tracing the hyperlinks to the external databases, including UniProt, PubChem, HapMap, OMIM, PRIDE, and ProteinAtlas; and more.

BiblioSensor

The BiblioSensor is a toolbar for the Web-browser (Internet Explorer). The toolbar enables to collect the personal information about Bio-Internet surfing over the following resources :

- Entrez, hosted by the National Institutes of Health
- UniProt; Expasy, hosted by the European Bioinformatics Institute

Mental Shield

The BiblioSensor supports the MentalShield option to switch to the privacy mode for Web-usage activity.

With the MentalShield option switched on, the information is not transferred to the knowledge-base. In addition, the Shield can be used to mask your IP address, to exclude your personal data from the statistical analysis of Internet portals.

NER Detectors

NER Detector is a special sub-program, which serves for Name Entity Recognition (NER). The detector is based upon the specific parsing rules and patterns that are used to perform the extensive data-mining. Each detector is optimally suited to match one of the external sources of information on the putative entities recruited for inclusion into the bibliobase.

Semantic Enterprise

Semantic Enterprise (SE) was developed as an integrated solution to exploit the Neosemantic™ Technology for creation of the in-house centralized knowledge warehouse.

The SE assignee obtains the middle-ware for gaining the information on the Bio-Internet surfing habits of employees. This information is processed on a regular basis to report the following:

- overall and daily activity;
- general research trends for the given time period, visualized as gene-centric and MeSH-centric semantic networks;
- cross-period comparison of research trends;
- most often accessed resources and papers;
- unusual, but statistically reliable novel research trends;
- matching the research interest between laboratories and groups.

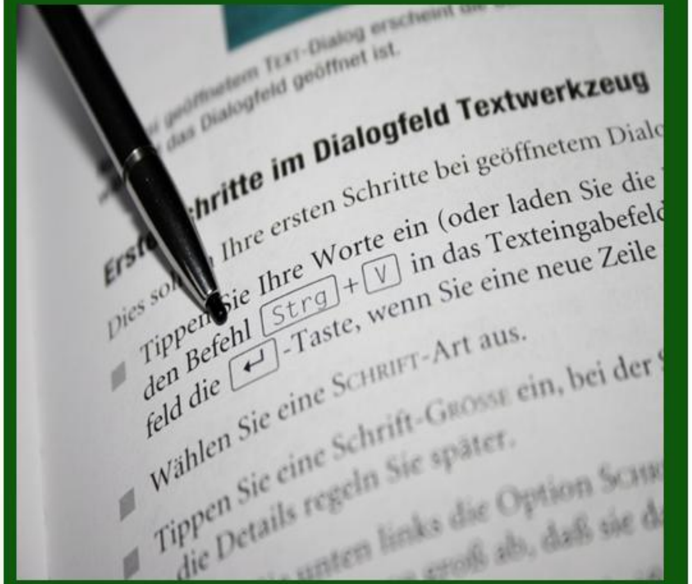
Choose peer-to-peer policy for SE in order to make statistical monitoring available to any member of your research team, and observe how scientific issues gain more popularity, than live journal.

Collaboration

ProContent seeks for the collaboration with research institutions and companies. The ultimate goal of collaboration is promotion of the Open Knowledge approaches for life science.

We invite collaborators to join the coalition of will to share the proprietary knowledge and obtains the necessary software prototypes to implement sharing:

- *Modules for* BiblioCMS;
- *Modules from* BiblioEngine Toolkit with BiblioSensor license;
- Technical support for customization of collaborator's Biblio products;
- Framework to develop intellectual "gear" between sensor and CMS.



The freely provided software enables collaborators to create proprietary knowledgebases, which can be incorporated to the ProContent business model.

ProContent Limited

Scheelestrasse 89
12209 Berlin

Tel +49 179 4631067
Fax +49 30 8015422
Email: procontent@web.de