R&D Solutions for PHARMA & LIFE SCIENCES

Rapid Access to Relevant Answers
Elsevier Text Mining
Finding detailed information from unstructured content

Every year, researchers worldwide produce and publish a huge amount of scientific information in articles, conference proceedings, textbooks and other media. In every field relevant for the life sciences and pharmaceutical research, thousands of items appear annually, and any one of them could contain the information crucial to a life sciences project.

Finding answers within that mass of content is particularly difficult because the content is not homogenous. In order to find detailed information in that unstructured data, researchers need the support of a dedicated tool. That’s where Elsevier Text Mining comes in.
Designed to support all life sciences researchers who need to rapidly retrieve answers from unstructured data, Elsevier Text Mining is an easy-to-use, cloud-based solution. It facilitates the discovery of novel insights and relationships within the huge volume of scientific content that is published every year, including abstracts, full-text articles, conference proceedings and grants. With Elsevier Text Mining, professional search specialists and researchers can work more rapidly and more efficiently - it no longer takes hours to find the data for informed decisions. Streamlined and optimized software takes the user from a query to result in seconds.

What is Elsevier Text Mining?

The Benefits of Elsevier Text Mining

Superior retrieval of answers
- Full-text searching as opposed to searching on citations and abstracts only
- Comprehensive biomedical dictionary and taxonomy developed by life sciences experts
- Superior performance at extracting entities and relationships

Immediate access to information and dictionaries
- Rights to mine full-text Elsevier content included with product subscription fee
- No need for tedious dictionary creation or system training

User-friendly system
- Cloud-based solution, easily deployed enterprise-wide
- Easy-to-use interface - no informatics expertise required
- World-class support from Elsevier experts and consultants

Flexible system to suit users’ needs
- Customizable to meet a variety of research needs
- Integration of internal and third-party content possible
- Alternate taxonomies and dictionaries also available
Creating a single, searchable resource

Elsevier Text Mining is a tool that is easy to learn and easy to use. It takes the vast and complex wealth of information from all of the different relevant sources and organizes that content into one normalized and searchable resource. A comprehensive dictionary of terms enables the system to recognize different entities and their respective synonyms, while the system taxonomy categorizes these terms into important biomedical concepts such as genes and proteins, diseases, small molecules, clinical parameters and medical and research procedures. Elsevier Text Mining uses a taxonomy-assisted search to help users create simple queries that then leverage the taxonomy and comprehensive dictionary for thorough searching. Researchers benefit from Elsevier’s extensive experience with life science terminologies, indexing life science content and providing empowering information solutions to science, health and technology professionals.

Retrieving more targeted answers
Searches are further facilitated by Elsevier’s Natural Language Processing (NLP) engine. It is programmed to recognize terms as parts of speech, and users can also specify additional linguistic relationships within their searches. Several levels of term co-occurrences are also available, ranging from the precise semantic-level co-occurrence to less stringent sentence-level, paragraph-level or document-level co-occurrences. Some document sectioning also enables users to target their searches to find the results of interest. Several search options are available so both novice users and professional search specialists can take advantage of the system.

Integrating data into in-house systems and analysis tools
Elsevier Text Mining is not isolated from the working environment of researchers. Results are displayed via the web interface, but that is not the only option. Users can choose to export the results so they can be further analyzed with other analytical tools. Programming access is also possible to give users even more possibilities.

What are the applications of Elsevier Text Mining?
All life sciences researchers can benefit from Elsevier Text Mining. It supports applications at every stage of the drug development pipeline, from basic research through clinical trials to healthcare economics outcome research. It is equally suited to rapidly finding the relevant data to answer critical questions in early drug development and retrieving the information needed for systematic reviews and evidence-based medicine decisions.

Sample Applications for Elsevier Text Mining

1. **Biological pathway-based analysis**
   - Easily discover the entities and pathways involved in a particular disease.

2. **Biomarker identification**
   - Find all of the possible bio markers reported for a medical condition.

3. **Lead optimization**
   - Rapidly retrieve all of the available toxicity, efficacy, pharmacokinetic, safety and metabolic data for a drug candidate to be able to rapidly determine its potential.

4. **Drug repositioning studies**
   - Determine other indications for an existing drug, in particular for rare diseases where it might prove effective.

5. **Drug safety monitoring**
   - Identify adverse events in a class of compounds for a single or multiple indications.

6. **Clinical trial analyses**
   - Explore the other drugs being tested in the same pharmaceutical space and see the different clinical endpoints or adverse events that are reported for each one.

7. **Patent analyses**
   - Identify the leading therapeutic areas for a given compound or target and see what claims and patents have been made for them.

8. **Systematic reviews to answer patient-centric questions**
   - Thoroughly yet quickly research the behavior of patients to discover whether they are reported as complying with therapeutic instructions.
New content added weekly from public and commercial sources

- Abstracts
- Full-text articles
- Conference proceedings
- Grants
- Clinical Trials
- Patents
- Internal Data

Keywords and data are extracted from the content, including

- Drug names
- Measurements
- Disease names
- Research methods
- Treatment methods

Natural language queries find semantic relationships like:

- "Drugs that reduce blood pressure"
- "Synthesis of Thalidomide"
- "Pathogenesis of cancer"
- "Resistance to linezolid"

Output formats suit various user work flows

---

**Growing resources**

New content is added and indexed on a weekly basis. This ensures that pharmaceutical researchers, healthcare professionals, informaticians and other life science researchers always have up-to-date information for their critical decisions.

**Input**

Content (ca. 7 million items)

**Taxonomy**

Keywords and data extracted from the content, including:

- Drug names
- Measurements
- Disease names
- Research methods
- Treatment methods

**NLP Engine**

Natural language queries find semantic relationships:

- "Drugs that reduce blood pressure"
- "Synthesis of Thalidomide"
- "Pathogenesis of cancer"
- "Resistance to linezolid"

**Output**

Online interface

Data files

---

**Figure 1. The Elsevier Text Mining workflow**

**Figure 2. The Elsevier Text Mining user interface**
Why use Elsevier Text Mining?

Our extensive publishing experience has made us experts in classifying, indexing and working with life science information. We also produce professional search and workflow solutions for librarians, researchers and health professionals. These activities have allowed us to develop the expertise and technology needed to help with the extraction of information from scientific, regulatory and trade content.

We also have a team of highly skilled text mining consultants with extensive experience in life sciences and pharmaceutical drug development that can help address information challenges as well as scientific and business needs.

Sample content sources available with Elsevier Text Mining

<table>
<thead>
<tr>
<th>Content Source</th>
<th>Elsevier Text Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elsevier content</td>
<td>Full text articles from ScienceDirect Life and Health Sciences categories.</td>
</tr>
</tbody>
</table>
| Public content        | • Abstracts from all years of PubMed  
                         • Full-text articles from PubMed Central open access journals  
                         • Grants from all years of NIH Reporter                                    |
| Other licensed content| Additional content can be added upon request.Elsevier Professional Services will manage the integration of such content. |
| Internal content      |                                                                                      |
R&D Solutions for Pharma & Life Sciences

Elsevier's R&D Solutions is a portfolio of tools that integrate data, analytics and technology capabilities to help pharmaceutical & life sciences customers achieve pharmacovigilance pain relief with less pain and more efficiency.

For more information about R&D Solutions for Pharma & Life Sciences, visit elsevier.com/rd-solutions/pharma-and-life-sciences

Elsevier offices

ASIA AND AUSTRALIA
Tel: + 65 6349 0222

JAPAN
Tel: + 81 3 5561 5034

KOREA AND TAIWAN
Tel: +82 2 6714 3000

EUROPE, MIDDLE EAST AND AFRICA
Tel: +31 20 485 3767

NORTH AMERICA, CENTRAL AMERICA AND CANADA
Tel: +1 888 615 4500

SOUTH AMERICA
Tel: +55 21 3970 9300