Intellexer™ SDK

Processing information effectively
Faster development time with less costs
Successful expansion of the capabilities of existing systems
Independent utilization of necessary modules
Introduction

Ever-growing information flows from different sources, need for fast and accurate knowledge management, automatic document analysis. The list of the challenges the corporations and individuals are facing nowadays could be continuously enriched.

Intellexer Software Development Kit (SDK) is a flexible and easily-customizable platform that offers the audience a number of benefits:

▪ Answering your questions in natural language
▪ Intelligent search for a necessary document
▪ Extraction of relevant information from documents
▪ Summary of documents
▪ Comparison and grouping of documents by business domains, industries
▪ Recognition of named entities (names of people, organizations, locations, etc.) and relations between them

Intellexer SDK includes a set of tools that allow software engineers, natural language researches, or business professionals to successfully complete their objectives across diverse sectors, including:

▪ Enterprises with intensive document circulation
▪ Intellectual assets management departments
▪ Business Development/Marketing departments
▪ Research&Development departments
▪ Governmental/Local councils
▪ Consultancy agencies
▪ Online service providers
Intellexer SDK modules

Intellexer SDK is designed to deliver the maximum flexibility in terms of its use and integration into corporate business processes.

The modules of Intellexer SDK can be independently integrated into various business applications, for example:

- Document/knowledge management systems
- Enterprise content management systems
- Business intelligence solutions
- Intellectual property management systems
- Help desk applications

Along with the main modules, Intellexer SDK incorporates an additional module: **Intellexer™ Spellchecker**. It is used for spelling checking and correction of spelling mistakes.

Intellexer Spellchecker is a valuable addition to:

- Content management systems
- Corporate intranets
- Customer feedback and surveys
- Customer relationship management solutions
- Text editors
Intlexer Preformator

This module extracts text from documents of different formats and stores information about text layout along with extracted parts.

Intlexer Preformator easily determines:

▪ Document format
▪ Document structure
▪ Document topic
▪ Document section
▪ Text place
▪ Text fragment
▪ Importance of text fragments

Intlexer Preformator recognizes the most popular formats:

▪ Microsoft Word (DOC)
▪ Rich Text Format (RTF)
▪ Plain Text (TXT)
▪ XML
▪ Office Open XML (OOXML)
▪ HTML
▪ MHTML
▪ Microsoft PowerPoint (PPT)
▪ Microsoft Compiled HTML Help (CHM)
▪ MSG
Intellilexer Language Recognizer

This module easily recognizes the language and character encodings of the text under consideration.

Based on statistic and linguistic technologies, the module assures accurate recognition of about 15 languages covering the major European and Asian languages.

The intelligent tool can be effectively used in:

- Information extraction systems
- Automatic text summarization systems
- Question-answering systems
- Automatic document categorization systems
- Machine translation systems
- Document filtering systems

Benefits both for software engineers and end-users:

- Faster development time with less costs
- Expansion of the capabilities of existing systems
- Ease of integration
- High accuracy of language recognition
- Effectiveness
- Fast language identification
**Intlexer Linguistic Processor**

The core of Intlexer SDK is Linguistic Processor. It provides syntactical and semantic information for other modules (e.g. Summarizer, Comparator, Clusterizer, Categorizer, Question-Answering System).

Linguistic Processor is used to parse the input text and to extract multiple kinds of relations. For example, syntactic (noun phrases, verb phrases, adjectival and adverbial phrases, etc.), semantic (categorized relations within and between syntactic units of different complexity). The output of Linguistic Processor is a semantic tree with certain semantic types of relations assigned to the links between sentence elements.

Illustrated: Example of the semantic tree for the source sentence: *The company offers innovative computer-based testing and Internet-based testing solutions.*
Intelllexer Summarizer

This module automatically generates a summary of any English text.

The module can be easily integrated into:

▪ Document/knowledge management systems
▪ Scientific and technical knowledge bases
▪ Intellectual property management systems
▪ Desktop applications ¹ (Microsoft Word, Microsoft Outlook, Internet Explorer, etc.)

Those who can benefit from Summarizer:

▪ Company executives
▪ Patent lawyers
▪ News analysts
▪ Scientific researchers

Advantages:

▪ Identification of main ideas of the document ²
▪ Time saving
▪ Flexibility
▪ Ease of use
▪ Fast decision whether to read the whole document or not
▪ Generation of executive summary ³

¹ Summarizer as a desktop application is available at:

² Summaries are topic-oriented (e.g. politics, economics, law, health, information technologies) and structure-oriented (patent, scientific article) to take into consideration your specified interests and preferences. Other topics can be included to serve your needs better.

³ You can specify the number of sentences required in the summary output.
Example of summary

What is a Nanotechnology?

Despite unprecedented government funding and public interest in nanotechnology...

One of the most pressing issues facing nanoscientists and technologists today is...

As a result of decades of speculation, a number of myths have grown up...

... The exact size of the atoms is less important than communicating the fact that nanotechnology is dealing with the smallest parts of matter that we can manipulate.

While there is a commonly held belief that nanotechnology is a futuristic science with applications 25 years in the future and beyond, nanotechnology is anything but science fiction.

In the last 15 years over a dozen Nobel prizes have been awarded in...

According to CMP Científica, over 600 companies are currently active in...

Almost every university in the world has a nanotechnology department or will...

...

What is new about nanotechnology is our ability to not only see, and manipulate matter on the nanoscale, but our understanding of atomic scale interactions.

One of the defining moments in nanotechnology came in 1989 when Don Eigler...

For the first time we could put atoms exactly where we wanted them, even if...

...

Nanotechnology, like any other branch of science, is primarily concerned with understanding how nature works.

We have discussed how our efforts to produce devices and manipulate matter...

Nature has the ability to design highly energy efficient systems that operate...

We do not, although one day our understanding of nanoscale phenomena may...

The exact size of the atoms is less important than communicating the fact that nanotechnology is dealing with the smallest parts of matter that we can manipulate.

While there is a commonly held belief that nanotechnology is a futuristic science with applications 25 years in the future and beyond, nanotechnology is anything but science fiction.

What is new about nanotechnology is our ability to not only see, and manipulate matter on the nanoscale, but our understanding of atomic scale interactions.

Nanotechnology, like any other branch of science, is primarily concerned with understanding how nature works.
**Intellexer Comparator**

This module provides intelligent comparison of documents by finding and extracting information with similar meaning.

The module can be used in:
- Document/knowledge management systems
- Patent prior art search and analysis
- Protection against intellectual rights infringement
- Evaluation of new trends in business, science and technology
- Assessing potential competitors

Comparator collates texts and determines the degree of similarity between the ideas they express. The degree is calculated within the range of 0–100%, where “0” means "absolutely different texts" and “100” means "the same text".

Illustrated: Example of comparison
Intelllexer Clusterizer

This module extracts terms (noun phrases) from the analyzed text and presents a hierarchical listing of them, differentiating between terms with broader and narrower meanings.

Clusterizer enables you to:

▪ Organize, group extracted terms effectively
▪ Present results in a more logical, structured, and searchable way
▪ Browse and search information efficiently
▪ Overcome terminological confusion when developing Ontology

Clusterizer can be used for:

▪ Ontology building productivity, Ontology enrichment
▪ Ontology management to ensure that the Ontology remains relevant and usable over time
▪ Determining taxonomic relations
▪ Business classifications
▪ Categorization of products for sale and their features
▪ E-commerce applications

Illustrated: Example of a hierarchical listing of extracted terms
Intellyer Categorizer

This module automatically finds a pre-defined category for documents.

Categorizer is ideal for:

- Enterprise content/knowledge management
- Business information management
- Patent prior art search and analysis
- Protection against intellectual rights infringement
- Evaluation of new trends in business, science and technology
- Gathering intelligence about competitors’ activities

Main benefits of Categorizer:

- High accuracy
- Flexibility
- Use of semantic technologies
- Analysis of the entire text in a document
- Possibility to filter non-English text

Organize documents in an easy and intuitive way for smart search and retrieval

Illustrated: Example

---

1 Pre-defined categories can be: human resources, research and development, business, marketing, finance, banking, etc.

2 Categorizer as a desktop application is available at:

http://categorizer.intellexer.com/download.html
Intelllexer Question–Answering System

Fully automatic Question–Answering System retrieves direct, relevant answers to questions posed in natural language.

The right answer is determined by taking into consideration:

▪ Matched terms and pairs in the question and the answer
▪ Weights of matched terms and pairs in the question
▪ Total weight of pairs and terms in the question

Question–Answering System is a real benefit for:

▪ Casual questioneers who ask simple factual questions
▪ Employees who seek quick answers on company policy
▪ Consumers who look for specific product features or prices
▪ Researchers who collect information about the subject area they are interested in
▪ Knowledge engineers who create semantic databases

The module is extremely useful for:

▪ Help desk systems
▪ Enterprise document/knowledge management systems
▪ E-libraries/catalogues

Intelllexer Question–Answering System can be easily tailored to specific subject areas by building a searchable database from a set of documents provided by the user.
Intellexer Question–Comparison Tool

This is an effective module of question understanding and processing that recognizes equivalent questions/question paraphrases regardless of the speech act or of the words, syntactic inter–relations or idiomatic forms.

Question–Comparison tool effectively processes questions by:

▪ Grouping questions into fine–grained clusters
▪ Identifying paraphrases from all question pairs within each cluster
▪ Extracting patterns from the identified paraphrases

The core of the module:

▪ Question transformation patterns
▪ Syntactic and semantic similarity features
▪ Question focuses
▪ Concepts shared by questions
▪ Probabilistic framework

The effective matching process extracts basic syntactic and semantic constituents used for detecting the main topics of the compared questions:

▪ Question’s subject
▪ Main verb
▪ Direct object

Intellexer Question–Comparison tool can be used for a variety of purposes. As the example, it can be integrated into the system with a rich FAQs database. The tool helps to substantially reduce the amount of user efforts required to find the right answer.
Intelllexer Named Entity Recognizer

This module helps you properly identify and extract key information hidden within the text:

- People
- Position/occupation
- Nationality
- Organization
- Geographical location
- Date
- Event

The intelligent module can be effectively used in:

- Information extraction systems
- Question-answering systems
- Automatic text summarization systems
- Machine translation systems
- Automatic speech recognition systems
- Topic detection systems

Illustrated: Example of named entity recognition (named entities are highlighted with different colors)
Intellexer Event and Relation Extractor

This module effectively determines who did what to whom, when and where.

The main types of relations:

- Person-age
- Person-nationality
- Person-position
- Person-organization
- Organization-location

Intellexer Event and Relation Extractor facilitates to:

- Identify and track individuals across diverse sources
- Track activities of particular individuals across a range of organizations
- Track activities by time line and geographical space
- Learn more about a particular company
- Find out where and when an event happened, who was involved
- Detect semantic relations between named entities, which may be located in different sentences
- Integrate information from diverse sources
- Acquire domain relevant terms and their relations

Illustrated: relations between named entities in the source sentence: Eric Emerson Schmidt (born April 27, 1955 in Washington, D.C.) is Chairman and CEO of Google Inc.
Conclusion

Intellexer SDK, as a holistic linguistic platform for natural language processing, offers a set of advanced means for knowledge management and transformation with the aim of extracting information.

The linguistic platform is vitally important for:

- Document/knowledge management
- Intellectual property management
- Enterprise content management
- Business intelligence solutions
- Help desk applications
- Voice/data processing