



LINGEA

Translations and localizations
Language technologies
Products





We cannot do without language technologies

Language, translation and speech technologies represent essential elements in efficient communication with computers today, in particular when working with textual and multimedia data. Whether you use office programs, search and analyse large amounts of textual data or use content management systems or multimedia, you always encounter language technologies that are meant to make your work easier and faster.

Potential for improvement

Although some technologies are already included in applications that we use now on a daily basis, there is still huge potential for their improvement. This is clear not only to large multinational companies that invest considerable amounts of money in further research and development, but also to potential customers who may save lots of time and resources as well as increase their work efficiency using these technologies.

Future

The future development of these technologies will be driven by growing demands for user-friendly software. Given that quality simulation of natural human language is not yet in sight, researchers have several short-term and very realistic goals that are supposed to improve the user experience: flawless grammar checkers for text editors, intelligent e-mail sorting and the automated creation of replies, programs for categorization and summarization of documents or the extraction of important information from large texts.

About us

Lingea has been developing language technologies and applications since 1997 and has also been publishing printed titles using its own language data.

We are a leading supplier of language technologies and a partner of reputable world companies such as HarperCollins Publishers, Pearson Education, Oxford University Press, Microsoft, Adobe, Autodesk, Siemens, Sony and others. Our applications and data are used by customers in Europe, USA, Brazil, Mexico, China or Korea. Currently we have our subsidiaries in the Czech Republic, Hungary, Poland, Romania, Serbia and Slovakia.



TOOLS AND TECHNOLOGIES

/06

Lingea covers a broad spectrum of language, translation as well as speech-oriented technologies for a whole range of European and world languages.

TRANSLATIONS AND LOCALIZATIONS

/24

Among other things, Lingea also specializes in the localization of software, information systems, web sites and large translation projects.

PRODUCTS AND DEVELOPMENT

/26

Lingea develops both printed and electronic dictionaries as well as mobile applications. These products can also be customized.



Microsoft®



Technologies

Our technologies combine our language data, know-how and software for processing one of the most complex systems of our world: the human language. We encounter human language in both written and spoken forms and our technologies are meant to help you orient yourself in it better, search it, structure it efficiently or otherwise use it. We are able to prepare outputs for different platforms or programming languages.

Our technologies are just as suitable for various search engines, such as Solr or Elasticsearch, using the Lucene library.



LANGUAGE TECHNOLOGIES /08

We support over 20 languages in terms of full text search, spell checking, hyphenation or language recognition. These are suitable for search engines, text editors, online shops or information, content management and library systems.

TRANSLATION TECHNOLOGIES /16

Automated translations for multilingual search and translation support. Technologies based both on language description and large corpora, combining linguistic as well as statistical approaches depending on the purpose.

SPEECH TECHNOLOGIES /20

Technologies for voice communication with computers or mobile devices, searching audio and video files or, as the case may be, the enhancement of language learning..

OUTPUTS WE ARE ABLE TO SUPPLY

- processed and raw data
- printed publications
- e-books
- mobile applications
- customized software, turnkey solutions
- stand-alone tools
- integrated tools
- search technologies
- translation tools

PLATFORMS WE USE AND SUPPORT

PC	MOBILE	ON-LINE
Windows	Android	Applications
Linux	Apple iOS	Web sites
Mac OS X	Windows Phone	

PLATFORMS



COMPATIBILITY



SEARCH ENGINES



Language technologies

The term "*language technologies*" represents sophisticated software solutions based on the knowledge of natural language. This includes useful components expanding the possibilities of search tools and enhancing the quality of applications such as search engines, online shops, advertising tools, text editors, content management systems, OCR systems, library systems, etc.

You surely know, for example, the spell checker or the synonym dictionary used in the Microsoft Office product line, which have been supplied by Lingea since 1997.

THE FOLLOWING PAGES
CONTAIN AN OVERVIEW OF OUR
TECHNOLOGIES AND TOOLS.



Spell checker

Spell checker is probably well-known and commonly used by virtually everyone. And as it is the Lingea solution that is integrated in, for example, MS Word, Excel, PowerPoint, Adobe InDesign and Photoshop, you too may have long been one of the satisfied users of our products. Thanks to the efficient compression algorithm and the overall compilation of the dictionary, the program interface is simple and easy to use in virtually any software product.

Among other things, Spell checker, checks the correctness of capitalization in acronyms (USA, IBM) and the final dots in abbreviations (i.e., etc.).

AVAILABLE FUNCTIONS

- Mistyping suggestions: will generate and offer all words which may be the correct form of the mistyped word
- Use and management of user-defined or special dictionaries for automated replacement of misspelled words.
- Setting of parameters, for example to ignore acronyms, one-letter words, words containing digits, etc.
- Currently, the spell checker is available for a whole range of languages (see table on page 29).

Full text search

For reliable full text search, the “lemmatiser” is useful – it is a component that enables you to look up words regardless of the grammatical form they have in the searched text. When searching for “playing guitar”, you would probably also appreciate finding articles containing the phrase “how to play guitar,” etc.

LANGUAGE PART

The basis of the entire solution, as well as of the Spell checker, lies in formal description of the morphology, enriched by other information. That includes especially parts of speech and grammatical categories.

PROGRAM SOLUTION

The program solution is very economical.

AVAILABLE FUNCTIONS

- Return the basic form of a word
- Return all morphologically related forms of a given word (lemma).
- User-defined dictionary
- Heuristics for the automated lemmatisation of unknown words.

Tip: If you want to make the search more user-friendly, it is recommended to use another of our components – the synonym dictionary. It can be further combined with our translation technologies for multilingual search and with speech technologies to search audio and video files.

Automatic hyphenation

The module for automatic hyphenation at the end of lines is absolutely essential, for example in DTP systems and text editors. It has also been used recently in more advanced e-book readers.

PROGRAM SOLUTION

The hyphenation program is designed to maximally cover all possible hyphenations for a given word. Unlike simple hyphenation algorithms, it is thus not limited to some safe determination of hyphenation possibilities. The program solution includes a very efficient algorithm to look up paradigms and relevant hyphenation information.

▶ AVAILABLE FUNCTIONS

- Return all possible hyphenation points in a word
- Currently, the hyphenation module is available for a whole range of languages and platforms (see table on page 29).

Thesaurus

The synonym dictionary is a useful aid both for users and other programs, such as search engines. A human can make use of the dictionary when looking for a more suitable term or trying to avoid repeating the same word. Programs can use this component to search for semantically interchangeable words.

LANGUAGE PART

The dictionary of synonyms contains synonymous terms (i.e., terms having the same or similar meaning) for each respective headword or its meaning. The synonyms are arranged in synonym rings according to the meanings of a given word. The first word in the ring is usually the most common one. The dictionary mostly contains neutral words. If informal, expressive or obsolete words are mentioned, such words are accompanied by relevant style notes. The dictionaries may also include information on antonyms.

▶ AVAILABLE FUNCTIONS

- Look up a corresponding headword for a given entered word.
- A list of all **synonyms** and **antonyms** together with possible style notes.
- Offer the most likely mistyping suggestions.

When also using the full text search module, synonyms will be found regardless of the tense, grammatical case or mode of the entered word. We offer the synonym dictionary for languages shown in the overview of language technologies (see table on page 29).

Language recognition

When searching multilingual texts on the Internet and databases, it is good to know what particular language is used in a given section of the text. Automatic language recognition is important for further work with the text such as indexing, lemmatization, tagging, searching, etc. A search engine can then easily recommend suitable tools for further processing. Language detection is the first step when processing texts for companies and institutions using multiple languages, searching large textual or audio data sets, processing them and, as the case may be, further analysing them.

LANGUAGE PART

Some languages can also be recognized by their typical script or characters specific to the given language. However, mostly it is morphological information that is used to recognize the language. In total, this component correctly detects over 40 world languages.

🔍 AVAILABLE FUNCTIONS

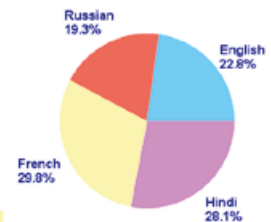
- Recognition of the main language used in the text
- Recognition of all detected languages appearing in the text and their tagging.

PROGRAM SOLUTION

The language recognition module works with text segments ranging from several words to entire documents, because the longer the evaluation text is, the more reliably the program can detect the main language. Take, for example, the phrase “je mine”, this is correct in French, Czech and Slovak, too.

Identify more than twenty languages and their scripts within large volumes of text.

Как пояснили в пресс-службе администрации Владивостока, для всех пенителтей рок-музыки на. Le traitement automatique du langage naturel est une discipline à la frontière de la linguistique, de l'informatique. सभी मनुष्यों को गौरव और अधिकारों के मामले में जन्मजात स्वतन्त्रता और समानता प्राप्त है ।



Automatic completion of diacritics

It can happen that you need to complete the text in a program solution by adding diacritics that are used in many languages. This can be also useful when processing requests and queries.

This problem is addressed by Lingea language technologies on two levels:

- A module that finds all relevant words containing diacritics for a word without diacritics. This is suitable when pre-processing a query in a search engine, for example.
- Fully automated solution using a statistical language model capable of converting the entire text without diacritics (hacek, acute accents) into a correct text with diacritics.

LANGUAGE PART

It is based on a formal description of morphology, thanks to which we are able to assign all correct words to a word without diacritics. There can be more than one such word, for example "mur" and "mûr" in French. Therefore, if we want to achieve fully-automated solutions, we also have to apply statistical methods based on a large corpus of correctly written texts. Thanks to this, we are able to find the most likely possibility.

PROGRAM SOLUTION

The first level consists of a function which returns all possible variants of words with diacritics for any entered word. It only uses a morphological dictionary for a given language that is about 1 MB in size. The second level is more demanding in terms of memory. It uses a language model the size of multiple gigabytes. The input can even be html text, and the module will preserve all html tags unchanged, only adding its own tags indicating the words which were changed. You can try this function at - www.nechybujte.cz.

Addressing letters

In some languages, the vocative (a form differing from the basic word form) is used to address people. If you want to gain a new customer and address a letter or an e-mail to him/her, it is essential that you don't make an error in addressing that person. Otherwise, it could easily happen that your message will end up in a spam filter or even in the trash. This component is a must when ensuring quality e-mail marketing in countries where the form of surnames is inflected.

LANGUAGE PART

Addressing a person may not always be that simple, and there is no unified rule that you can use. Some names are just extended by one letter, but some by multiple letters. In some cases the names do not change at all, but there is also a whole range of specific inflexions. Our tools are also able to identify if a given person is male or female, or we may, if needed, also generate the vocative of other words such as job titles.

AVAILABLE FUNCTIONS

Find the correct vocative form for any noun or adjective or some numerals. Identify the first name, surname or gender of a person. Return the correct male form for a female surname or a family name from the male surname, such as Dufek/Dufková/Dufkovi. The function also works the other way around. We offer this module for languages shown in the overview of technologies (see table on page 29).

Translation technologies

Over recent years, our own development has advanced to such a degree that we are also able to offer advanced solutions in the area of translation.





Multilingual search

This technology enables a search query in one language to obtain search results from relevant documents written in other languages. The user thus can concentrate on the information sought rather than on the language used to describe it. This technology is especially crucial in environments where multiple languages are used, such as in multinational organizations or companies having foreign business partners or ownership.

LANGUAGE PART

Among other things, this technology uses morphological tools, bilingual dictionaries and translation tools for a given language pair. Our large bilingual dictionaries containing about 100 000 entries are offered and marketed for major European languages (English, German, Russian, French, Spanish and Italian) as well as for Czech and Slovak, and smaller dictionaries are available for about 30 languages. Also available for selected language combinations are specialized terminology dictionaries for economics, technology and medicine, featuring up to 50 000 terms. If required, we are also able to create specific dictionaries for any specialized field.

PROGRAM SOLUTION

Given that this component does not use extensive corpora or translation models yet, its use is quite easy. Data of about 20-30 MB in size is sufficient for work with two languages. No high-performance processors or large data storage are necessary, so it can be integrated even in applications for mobile phones and tablets.

Translator

Informative translation deals with the rough translating of entire sentences and articles from one language into another. The translation is not perfect and will not be for some time to come, but it allows the reader to understand on a general level what a given article or web page is about. Currently, we offer this option for translations from English, German or Slovak into Czech. The quality of the translation results is comparable to the Google Translate or Microsoft Bing projects. Moreover, when translating from German into Czech, English is not used as a reference language.


LANGUAGE PART

If we want to achieve truly perfect translations, we cannot just rely on dictionaries and morphology. Some fundamental problems need to be addressed:

1. Selection of the proper meaning in cases of polysemous words (words with multiple meanings)
2. The word order in sentences
3. Use of the correct forms of the words in the target language
4. Idiomaticity and other atypicalities of particular languages

We address these obstacles in our company as well, trying to achieve the best results possible given current limitations.



anglicky ► česky		přeložit	smazat	kopírovat	least little ¹ little ² few, little little, small
More than a fifth of men and nearly a quarter of women were taking at least three prescriptions.		Více než pětina mužů a téměř čtvrtina žen užívali minimálně tři recepty.		least [li:st]  <i>adv</i> 1. the least nejmenší, minimální 2. sebensí (takřka žádný) <i>adv</i> nejméně (ze všeho) <i>phr</i> at least nejméně, alespoň, přinejmenším (minimálně) at least aspoň (když nic jiného) not in the least ani v nejmenším, vůbec ne, ani trochu last but not least v neposlední řadě (se stejnou důležitostí)	
				překladový výkladový synonymický	

PROGRAM SOLUTION

In the area of statistical translation, we now combine our own data and technologies with the Moses system, a result of many years' development by several European universities, headed by the University of Edinburgh. Together with the very same university, we are now participating in research within another European project focused on machine translation.

The size of models and their configuration determines the hardware demands. Given the big differences in hardware demands, depending on the differing quality and types of outputs, we customize our translation language models according to the types of texts to be translated. The narrower the domain and the larger the available translation and language corpora are, the higher the quality of translation.

When preparing models we also use our own corpora, bilingual dictionaries, terminology databases, morphologies and their combinations, which will enable us to achieve better results with smaller models.

Speech technologies

Speech technologies have recently been used when solving problems that only humans could have handled before, which means a huge reduction in labour costs. However, they are also often used in cases where their application only brings increased user comfort, and thus represent a certain competitive advantage. Futuristic ideas are turned into useful helpers, for example when using and controlling mobile devices, but also in teaching and learning, processing voice recordings or searching multimedia content. Together with its partners, Lingea also offers solutions in this area.

We focus especially on learning and efficient text processing using speech technologies. More details relating to respective solutions can be found in other sections.



Voice search

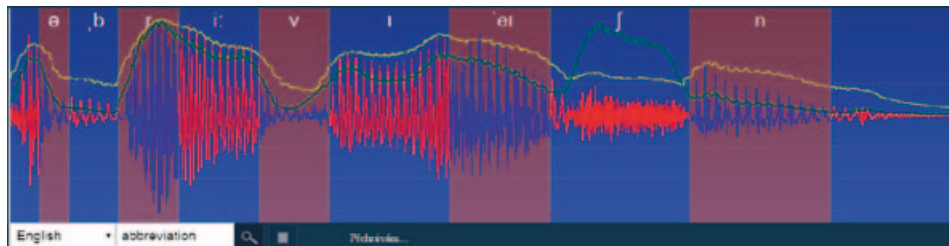
The voice search component was created to be used in Lingea's electronic dictionaries, but in combination with other language technologies it becomes an efficient tool for searching databases and texts under conditions where voice input is a more comfortable option than typewriting. This technology can also be combined with multimedia content search, which creates a system that is not only controlled by speech, but that also searches speech.

LANGUAGE PART

It combines a speech recognizer with technologies for full text search and, if necessary, the synonym dictionary or even translator. The result is a simple but powerful interface that is easy to use and yet delivers results comparable to advanced searches using typewritten queries.

PROGRAM SOLUTION

An online recognition server is typically used for processing, and demanding calculations are handled thanks to sufficiently dimensioned infrastructure. The application can thus be used on virtually any device, including those having weaker processors or limited memory capacity (such as mobile phones).



Pronunciation trainer

Pronunciation practise had been an area where a language teacher was absolutely indispensable to demonstrate correct pronunciation to the learner. It was speech recognition with graphic audio display and highlighted boundaries for individual phonemes that enabled teacherless learners to obtain some feedback. The recording of a native speaker's voice is used as a model, again, but this time it is not only replayed, but also displayed as a graph with a cursor pointing to the exact point being replayed. A similar graph is generated from the learner's pronunciation, allowing comparison of mutually corresponding parts of the audio graph and finding deviations in the learner's pronunciation. During the next replay, the learner can focus on a given spot to better realize the differences between his/her own pronunciation and the model pronunciation.

SPEECH PART

The core of the technology is a speech recognizer that attempts to interpret a certain sound according to a given transcription. It searches for the parts that are the most similar to the phonemes from the transcription and marks their boundaries.

PROGRAM SOLUTION

The recognizer is the most hardware-demanding part of the technology, and it usually runs on a dedicated server. Integrated in the application are graphic displays with highlighted phoneme boundaries, replay animations and communication with the recognition server.

▶ AVAILABLE FUNCTIONS

- The recognition server gets a sound input together with its transcription and outputs the positions of the boundaries of individual phonemes.
- The technology can evaluate the quality of pronunciation.

APPLICATION PART

It passes on to the server the audio as well as its transcription, displays the graph of audio values with highlighted sound intervals of individual phonemes, and animates the cursor in the graph when replaying the sound.



Searching multimedia content

Thanks to the recognition of recorded texts and their indexing, you can gain quick access to information contained in a recording without having to listen to the whole recording. That results in huge time-saving in cases where you are working with larger volumes of audio recordings. If you have available a recording archive processed in this way, you can even find information occurring only marginally in the recording, when manually entered queries would probably not be effective in finding it.

LANGUAGE PART

This technology combines the speech recognizer with full text search complemented by language technologies for morphological search and, if necessary, the synonym dictionary or translator. It can also be combined with voice search, which results in a system that not only searches spoken texts but is also controlled by speech.

PROGRAM SOLUTION

As in all fast search systems, indexing first takes place in the given recordings, and subsequently it is possible to search them efficiently. The key component of this technology is the speech recognizer, converting audio to text. After that, the text is processed, prior to indexing (creating an index), by processes such as lemmatization (converting words to their basic forms) for morphological search or, as the case may be, translation into the index language; and finally the index itself is introduced, which is subsequently searched by the program according to the queries made. The queries are also processed and, for example, lemmatization or expansion of the forms or of the synonyms or translation into the index language may take place.

Translations and localizations

To achieve the maximum quality and speed of translation, we use our own software technologies that we have been developing for more than 15 years. The description of these technologies represents the main part of this text.

Among other things, we also use our own constantly updated bilingual and monolingual dictionaries, CAT tools and naturally specialized teams of high-quality translators, proofreaders and native speakers.

WE PROVIDE:

- translations within regular and express deadlines
- grammar checks and corrections
- proofreading
- DTP and graphic layout
- application of language technologies
- terminology creation and maintenance
- checking of consistency of terminology
- post-editing of machine translations
- linguistic testing
- creation and maintenance of translation memory
- publishing of both printed and electronic titles





Localization

As we are software developers ourselves, we are able to carry out quality localizations of software, information systems, applications and similar projects.

We specialize in larger contracts and are able to take care of the entire process from assignment, localization and testing to final handover of the project.

LOCALIZATION OF WEB SITES AND WEB APPLICATIONS

We will translate your web presentation, including web applications, their control and user-related elements, or your online shop, and adapt them to the standards of the desired client target group, including the implementation of our language search technologies.

Tip: Do you have your own program or application and want to offer it to foreign clients? You can rely on us to translate the texts, controls, strings, helps and other textual components of your applications, and we will adapt them to the customs and standards of the target market and test their functionality.

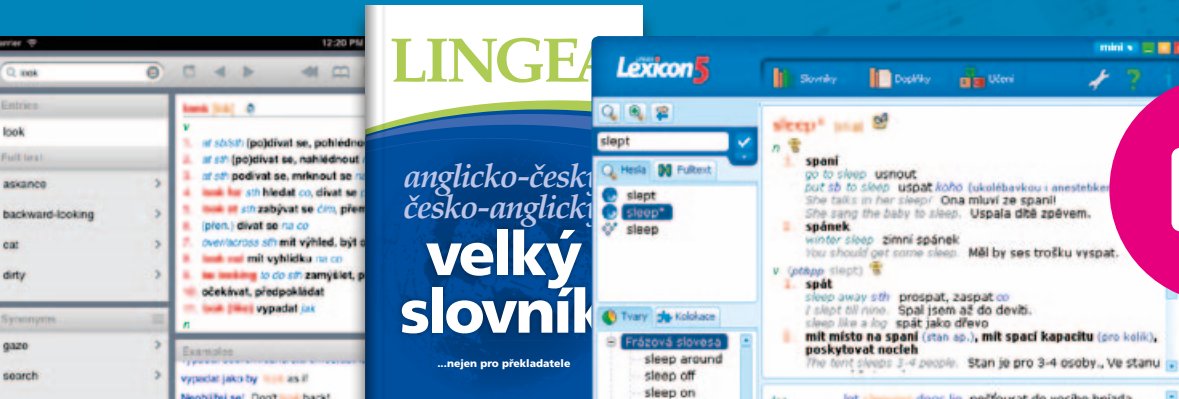
Products

The exceptionality of our printed and electronic dictionaries consists in the unique data that we have been developing for years, constantly extending and updating it. Our well-known applications include Lexicon, with its extensive general, technical, economic and medical dictionaries, and HandyLex for mobile phones and tablets with Apple iOS, Android and Windows Phone systems. We have published over 400 printed titles of our own, ranging from pocket dictionaries to the largest dictionaries for translators, and other language publications such as phrase books, "speaktionary", grammar overviews and synonym dictionaries - all of the above for over 50 languages in various combinations.

DICTIONARIES AND
OTHER APPLICATIONS
FOR MOBILE DEVICES

BOOKS MAPPING IN
OVER 50 LANGUAGES

LEXICON 5 – THE LARGEST AND MOST
SOPHISTICATED DICTIONARIES FOR
PROFESSIONAL AND DOMESTIC USE.



Development and data

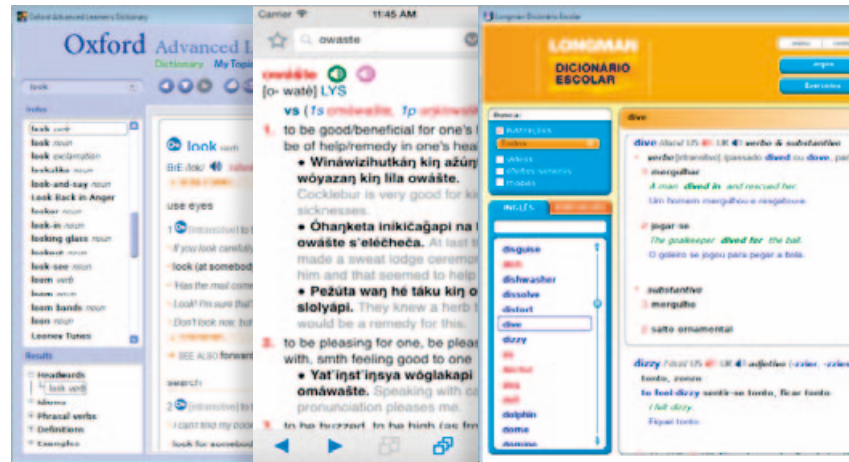
Lingea has extensive experience in the development of desktop, web and mobile applications for clients worldwide. We can create learning software platforms, internal communication solutions for companies or virtually any customized mobile application. We can efficiently create books, e-books and applications using the same source data, which helps reduce costs and prevents needless errors.

DATA

We grant licenses for our products, especially for dictionary data that have a broad scope of use. In addition, it is possible to license our phrase books, "speaktionary" or grammar overviews. Above all, we provide customized data according to the client's needs, and therefore we place a great emphasis on the maximum utility that can be derived from product development (for example, specialized phrase books or dictionaries for less frequent language combinations for language courses, etc.).

DEVELOPMENT POSSIBILITIES

- mobile applications
- software for PC and Mac
- e-books
- printed publications
- development of components for larger wholes
- development of customized products
- on-line applications



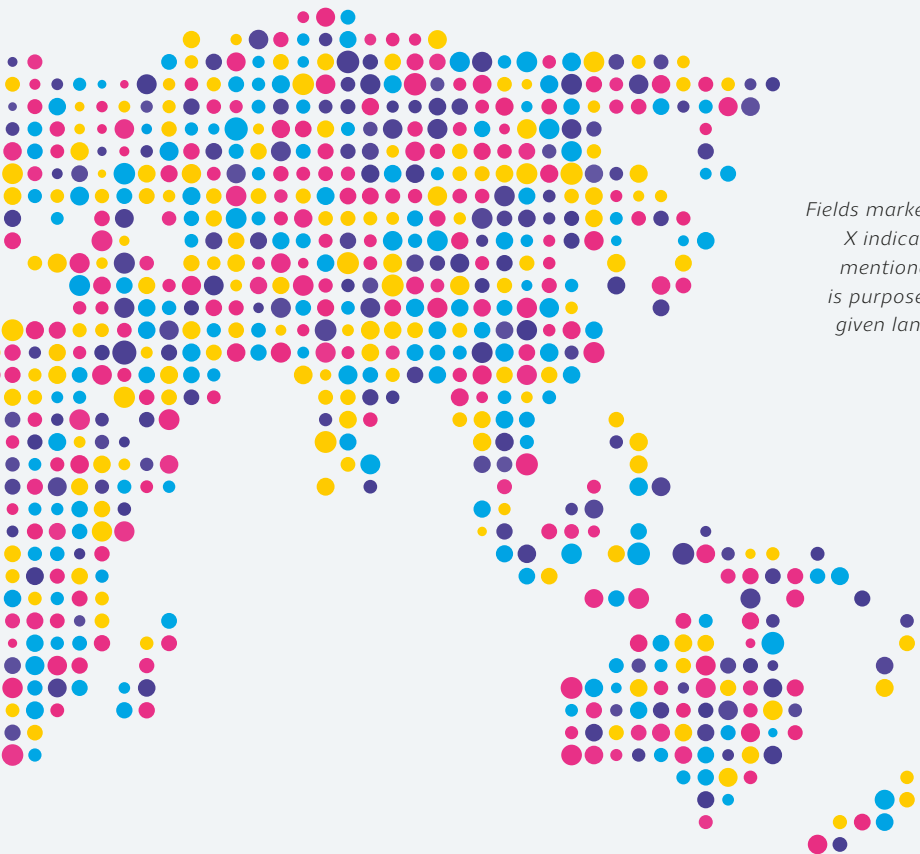
Supported languages

AFRIKAANS
ALBANIAN
ARABIC
ARMENIAN
BULGARIAN
CATALAN
CHINESE
CROATIAN
CZECH
DANISH
DUTCH
ENGLISH
ESTONIAN
FILIPINO
FINNISH
FRENCH

GEORGIAN
GERMAN
GREEK
HEBREW
HINDI
HUNGARIAN
ICELANDIC
INDONESIAN
ITALIAN
JAPANESE
KOREAN
LATIN
LATVIAN
LITHUANIAN
MACEDONIAN
MALAY

MONGOLIAN
NORWEGIAN
POLISH
PORTUGUESE
ROMANIAN
RUSSIAN
SERBIAN
SLOVAK
SLOVENE
SPANISH
SWAHILI
SWEDISH
THAI
TURKISH
UKRAINIAN
VIETNAMESE





Fields marked with
X indicate that
mentioned tool
is purposeless in
given language.

	Spellchecker	Full text search	Automatic hyphenation	Thesaurus	Automatic completion of diacritics	Addressing letters	Language recognition
Bulgarian	✓	✓	✓		x	✓	✓
Catalan	✓	✓	✓		✓	x	✓
Croatian	✓	✓	✓		✓	x	✓
Czech	✓	✓	✓	✓	✓	✓	✓
Dutch	✓	✓	✓		✓	x	✓
English	✓	✓	✓	✓	x	x	✓
Estonian	✓	✓	✓		✓	x	✓
Finnish	✓	✓	✓		✓	x	✓
French	✓	✓	✓		✓	x	✓
German	✓	✓	✓	✓	✓	x	✓
Hungarian	✓	✓	✓		✓	x	✓
Italian	✓	✓	✓		✓	x	✓
Japanese	✓	✓	✓			x	✓
Latin	✓	✓	✓		✓	x	✓
Norwegian	✓	✓	✓		✓	x	✓
Polish	✓	✓	✓		✓	✓	✓
Portuguese	✓	✓	✓		✓	x	✓
Romanian	✓	✓	✓		✓	✓	✓
Russian	✓	✓	✓		✓	x	✓
Serbian	✓	✓	✓		✓	x	✓
Slovakian	✓	✓	✓	✓	✓	x	✓
Spanish	✓	✓	✓		✓	x	✓
Turkish	✓	✓	✓		✓	x	✓

Customized solutions

In the above pages we have briefly introduced our offering of technologies, components and translation and localization services, which have been used by our customers worldwide for years. If you have any specific requirements, or are not knowledgeable about certain issues, not sure about what solutions or combination thereof would be the most convenient or economical for your needs, do not hesitate to contact us. Our experts will carry out the necessary analyses, suggest a customized solution for you, and perform comprehensive implementation. All Lingea applications are modular, so you can easily extend your system at any point in the future by adding other components or languages.



LINGEA

Lingea s.r.o.
Vackova 9, 612 00 Brno
Czech Republic
+420 541 233 160

info@lingea.com
www.lingea.com

www.dict.com

mobile.lingea.eu