IdiomSearch : automatic extraction of formulaic language and phrases

KEYWORDS

- Reading English
- Computer-aided translation (CAT)
- Proficiency level

Market description

Language learners & translators, automatic translation, computer-aided translation

All learners of English (and other foreign languages) are encouraged to read texts in order to develop their mastery of formulae and phrases (fixed expressions). Their English (or other foreign language) is supposed to benefit from the reading of newspaper articles and other documents, but they **largely fail to detect the relevant pieces of information in texts**, because these are nowhere to be found in dictionaries or other databases.

This tool enables users to receive an **almost immediate feedback on the presence of formulae and phrases** in any text, as well as an identification of their frequency and linguistic category (partly fixed, fixed, very fixed).

It can also be used to **check any source text** or **any translation** (whether machine or human translation) for the presence of formulae and phrases, **thereby revealing** many of its linguistic qualities. Finally, it is an additional method for assessing automatically the **proficiency level** of language learners by checking the percentage of formulae they use in their linguistic production.

Key features & benefits

This project is the result of an innovative algorithm that makes it possible to **automatically extract all types of phrases (fixed expressions in the broad sense) from huge linguistic corpora**.

For the very first time, a huge database of over 700,000 phrases (containing structures ranging from 2 to 8 words) has been compiled for English, and is under way for French, Dutch, Spanish, German and (Mandarin) Chinese.

Software status

English, French: ready to use

Dutch, Spanish, German: beta

Chinese, Italian: in progress

➔ http://idiomsearch.lsti.ucl.ac.be ⬅

Preferred partnership

Joint developments, licensing opportunities

Screenshot of IdiomSearch web interface

Interested to develop and / or commercialize this software?

Please contact

Sébastien ADAM
Technology Transfer Advisor
+32 10 47 24 43
sebastien.adam@uclouvain.be
www.ltto.com

UCL Université catholique de Louvain

Louvain Technology Transfer Office