



DAIL Software, SL. SimpleExtractor Versió 1.0. Disseny preliminar del Grupo de Investigación en Validación y Aplicaciones Industriales de la Universidad Politécnica de Madrid [recurs en línia].
<<http://www.dail-software.com/es/5-extractor-terminol%C3%B3gico-simpleextractor.html>>.
[Consulta: 20 setembre 2013]

Simple Extractor versió 1.0, de Dail Software (<http://www.dail-software.com/es>), es presenta com un extractor terminològic orientat a col·lectius professionals que treballen amb la llengua i amb textos d'especialitat. En concret es presenta com una eina útil particularment per a terminòlegs i bibliotecaris. El programa inclou un fitxer de paraules buides en castellà, i permet crear-ne de nous per a altres llengües, modificar-los i adaptar-los.

La interfície i les opcions de treball resulten molt simples. L'extracció es realitza en tot just tres passos. Primerament, cal seleccionar els arxius que es buidaran. Permet la selecció múltiple d'arxius i també la selecció de directoris. A més, pot buidar arxius en qualsevol format habitual, des de PDF fins a TXT. En cas de buidar arxius HTML pot presentar problemes en discriminar entre algunes etiquetes i les cadenes de caràcters dels candidats a termes. A continuació, un cop fixat el corpus de buidatge, les opcions de buidatge són molt intuïtives. Només cal determinar el rang de freqüència d'aparició mínima i màxima dels termes simples (d'una sola paraula) i dels termes compostos (de més d'una paraula fins a un màxim de divuit). Finalment, s'ha d'indicar l'arxiu de paraules buides a partir del qual realitzarà l'extracció. Es pot utilitzar l'arxiu que el programa inclou per al castellà o bé fer-ne un de nou. El programa inclou un editor per modificar o crear arxius de paraules buides en diferents idiomes. L'arxiu de paraules buides agrupa les paraules segons la seva categoria. L'arxiu que acompanya el programa inclou les categories bàsiques de paraules buides: adverbis, preposicions, conjuncions, determinants, les formes conjugades dels verbs auxiliars o més habituals, i caràcters. A més, permet crear categories noves per incloure altres paraules o cadenes de caràcters que l'usuari vulgui ignorar durant el buidatge.

El resultat és una llista de candidats a termes acompanyats del seu context. Permet visualitzar els candidats a terme segons la freqüència o bé per ordre alfabètic; tots

els termes alhora o bé les seqüències (n-grams) d'una, dues, tres, quatre o cinc paraules. Permet fer cerques de cadenes de caràcters i cerques amb comodins i truncació. L'extracció inclou qualsevol paraula o conjunt de paraules que no siguin a la llista de paraules buides ni comenci per cap d'elles. També permet editar qualsevol de les entrades directament. L'usuari selecciona o marca les opcions que desitgi i, finalment, genera un informe que inclogui la llista de termes amb el seu context.

La llista resultant d'una extracció no és necessàriament una llista de candidats a terme. En realitat, depèn de la naturalesa i el volum del corpus que es buida. Si es tracta d'un volum de text relativament petit i d'un gènere textual relativament poc rígid, les repeticions de paraules o agrupacions de paraules efectivament coincidiran amb els termes o les paraules clau del document. Per contra, si es tracta d'un volum de text relativament gran i d'un gènere textual molt rígid, com ara podria ser un manual d'instruccions, els candidats a terme poden incloure paraules o agrupacions de paraules que res tenen a veure pròpiament amb la terminologia del document.

Resulta, per tant, un programari molt intuïtiu. La precisió dels resultats dependrà de l'objectiu de l'usuari. La simplicitat del seu ús i el seu curt temps de resposta conviden a fer-lo servir per obtenir una primera llista d'unitats lèxiques que descriu el contingut d'una quantitat controlada de text, per exemple per descriure'l temàticament o per obtenir una primera visió del seu lèxic característic, especialitzat o no. Tanmateix, perd precisió quan es tracta de grans volums de text d'un gènere textual determinat. A partir de desembre de 2013, amb l'aparició de la versió 1.1, els usuaris podran obtenir una versió d'avaluació a través del web del programa.

PILAR SÁNCHEZ-GIJÓN
 Departament de Traducció i Interpretació
 Universitat Autònoma de Barcelona