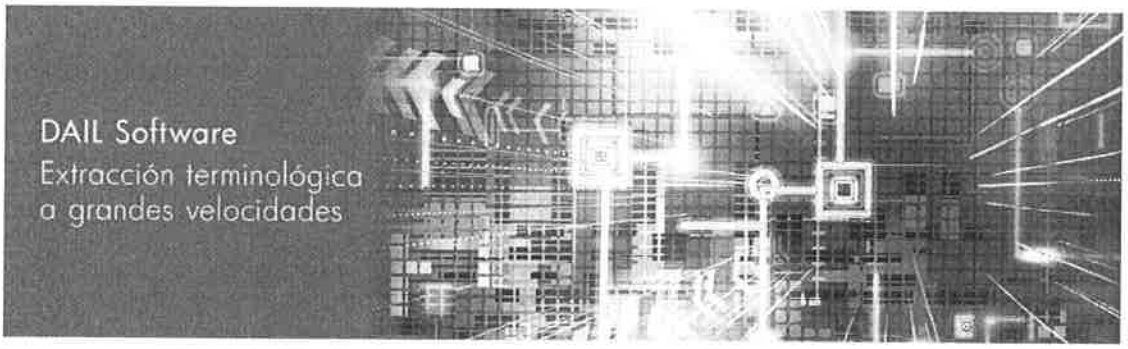
6 Consideraciones finales

Los propósitos de esta aplicación son:

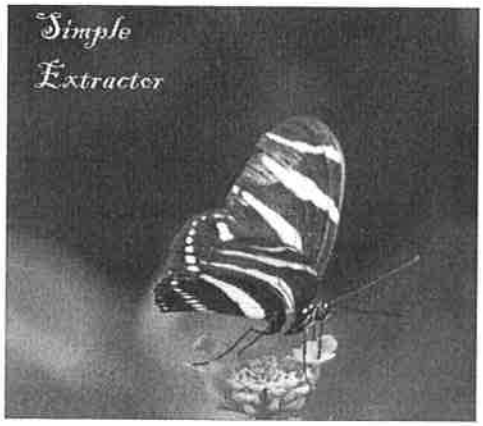
- **Versatilidad**. Es útil para diferentes propósitos
- **Alta capacidad**: Maneja grandes volúmenes de textos sin problemas.
- **Altas prestaciones**: La velocidad de la extracción terminológica es muy alta.
- **Fácil de instalar**: Únicamente se precisa la instalación de JAVA 7.0 (gratuita).
- **Fácil manejo:** Una completa ayuda y ayudas contextuales hacen su manejo muy intuitivo y sencillo.
- **Fácil mantenimiento**: No requiere más mantenimiento que aumentar a petición del usuario la lista de palabras vacías.
- **Fácil adquisición**: Una simple descarga desde la página web (htpp://www.dail-software.com/ y pago por tarjeta de crédito o PayPal).
- **Precio asequible**: Desde la primera licencia con descuentos importantes según cantidades o licencias corporativas.

La empresa DAIL Software S.L. (Desarrollo de Aplicaciones de Ingeniería Lingüística) aprovecha el equipo humano y experiencia nacional e internacional y tecnologías desarrolladas durante años en el seño del Grupo de Investigación en Validación y Aplicaciones Industriales (1) de la Universidad Politécnica de Madrid, habiéndose constituido a primeros de 2013 como *spin-off* de la misma.



[1] Grupo de Validación y Aplicaciones Industriales: http://www.vai.dia.fi.upm.es

AIL Software



"SIMPLE EXTRACTOR"

APLICAÇÃO INFORMÁTICA PARA EXTRAÇÃO E GESTÃO TERMINOLÓGICA

RESUMO

O *Simple Extractor* é uma aplicação informática orientada para a extração de terminologia a partir de textos. De entre as suas vantagens destacam-se o carácter intuitivo das interfaces e a simplicidade de utilização. Trata-se de uma ferramenta que permite configurar a extração e exportar ficheiros.

Palavras-chave: extração terminológica, palavras a ignorar, terminologia, lexicografia, glossários, *e-learning*, ensino de línguas, ferramentas para tradutores/intérpretes/documentalistas.

Índice

1 INTRODUÇÃO

O *Simple Extractor* é um extrator terminológico muito fácil de utilizar, que concilia grandes potencialidades com uma gestão feita prioritariamente pelo utilizador. Este pode configurar a capacidade de extração, o número de frequências e a seleção dos termos que considera válidos – de uma a sete palavras -, e eliminar e gerir as listas de palavras a ignorar (*stop words*). Os termos podem ser extraídos de ficheiros TXT, PDF e Word (.doc e .docx).

No âmbito de diversos projetos e de outras atividades de desenvolvimento levados a cabo no passado pelo Grupo de Investigação VAI da Universidade Politécnica de Madrid, verificou-se que, enquanto o uso quotidiano deste tipo de aplicações se centra num número reduzido de operações, as ferramentas existentes são muito complexas – tornando a aprendizagem um tanto desencorajadora – ou não são utilizáveis nas nossas línguas. Outras há que, integrando ferramentas de maior potencial, são muito caras. Por outro lado, algumas são difíceis de adquirir – muitas vezes por terem sido desenvolvidas numa perspetiva académica, com pouca vocação comercial.

No que respeita à investigação e ao desenvolvimento em Linguística, tais ferramentas oferecem diversas funcionalidades de grande importância (concordâncias, n-gramas, comparações estatísticas, medidas de similaridade), mas não dispõem de interfaces intuitivas nem apresentam uma visão integrada do processo de extração. O desenho do *Simple Extractor* pode ser visto como orientado para o utilizador, primando pela facilidade e pela simplicidade de utilização (com interfaces intuitivas e um fluxo de trabalho claro, sabendo-se sempre qual é o próximo passo), a que se junta ainda a acessibilidade do preço. O *Simple Extractor* foi concebido como uma aplicação de ambiente de trabalho, de forma a garantir a confidencialidade dos textos e dos materiais de trabalho do utilizador. Uma vez que a única componente que varia de língua para língua é o filtro das palavras a ignorar, a sua adaptação a novas línguas é simples – apenas é necessário adaptar o ficheiro de palavras a ignorar. Além disso, o *Simple Extractor* possibilita a introdução de texto codificado em UTF-8, permitindo assim a extração em todas as línguas românicas e em línguas como o Russo, o Árabe, o Chinês ou o Japonês.

Do ponto de vista das funcionalidades, o *Simple Extractor* caracteriza-se pela sua alta velocidade e capacidade de extração – no mínimo, 10 000 palavras por segundo e uma capacidade para tratar volumes de mais de 5 milhões de palavras numa só extração –, bem como pela sua robustez e fiabilidade.

Os potenciais utilizadores desta aplicação informática (utilizaremos também o termo "ferramenta", por razões de simplificação e por ser um termo bastante utilizado em projetos e trabalhos diversos) são vários: documentalistas, para seleção de termos para *thesaurus*; tradutores, para elaboração de glossários; lexicógrafos, para elaboração de listas de termos para fins específicos, bem como extração dos contextos em que estes ocorrem; professores de línguas, para preparação de materiais para os alunos e estudo dos diferentes contextos de utilização de um termo; terminólogos, para elaboração de listas de termos em função da sua morfologia e do seu uso específico num texto;

especialistas em estudos literários, para observação da composição e da frequência de uso das diferentes palavras numa determinada obra ou por um determinado autor; investigadores das línguas em geral ou de um autor em particular; investigadores no âmbito da Linguística de *corpus*.

O desenho preliminar desta ferramenta foi levado a cabo pelo Grupo de Investigação em Validação e aplicações Industriais da Universidade Politécnica de Madrid [www.vai.dia.fi.upm.es] e uma vez criada a empresa *spin-off* "Desarrollo de Aplicaciones de Ingeniería Lingüística" (DAIL Software SL), procedeu-se à conversão do que foi um protótipo de laboratório num produto suscetível de ser oferecido à comunidade internacional a um preço muito acessível, tendo em conta as suas elevadas potencialidades de desempenho, com fiabilidade e robustez.

2 DESCARREGAR E INSTALAR O *SIMPLE EXTRACTOR*

O *Simple Extractor* funciona em Windows, Linux e Mac. Antes de instalar a ferramenta é necessário instalar no seu computador a versão 7, ou superior, do *software* JAVA. A máquina virtual Java pode ser instalada gratuitamente a partir de http://www.java.com/pt_BR/download/manual.jsp.

Descarregue a ferramenta a partir de www.dail-software.com. Depois de realizado o processo de compra, irá receber um e-mail com o assunto **Descarregar o *Simple Extractor***, contendo a ligação para um ficheiro PDF (Figura 1):

Figura 1. E-mail para descarregar o Simple Extractor

Este ficheiro permite descarregar o *Simple Extractor* para Windows, Mac e Linux. A licença da ferramenta é válida para qualquer destes sistemas operativos (Figura 2).

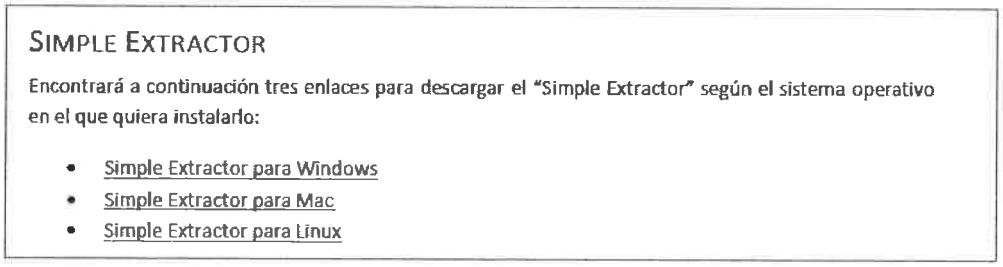


Figura 2. Ligações para descarregar o ficheiro de instalação para os diferentes sistemas operativos

2.1 Instalação em Windows

Descarregue o ficheiro de instalação para Windows (SimpleExtractorInstaller.exe) e execute-o. O processo de instalação é simples e basta seguir as instruções do programa de instalação.

2.1.1 Avisos de segurança

Dependendo da versão do Windows e das definições de segurança do computador, durante a instalação do *Simple Extractor* poderão surgir alguns avisos de segurança. O *Simple Extractor* é um *software* seguro, por isso deve ignorar os avisos de segurança do Windows durante a sua execução. Contudo, é possível que apareçam algumas das mensagens abaixo.

A Figura 3 mostra um aviso de segurança no **Windows 8**. Para instalar corretamente o *Simple Extractor* deve permitir as alterações.

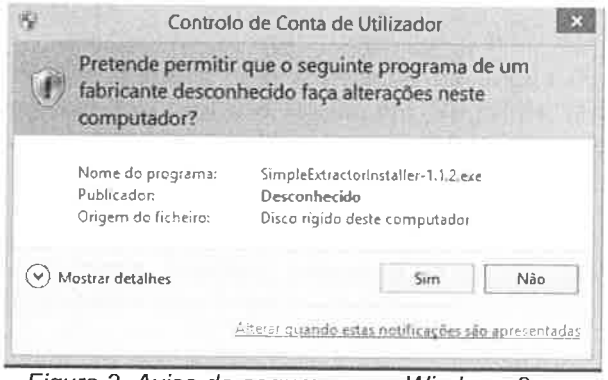


Figura 3. Aviso de segurança no Windows 8

No sistema operativo **Windows 7**, pode aparecer o seguinte aviso quando descarregar a aplicação (Figura 4):

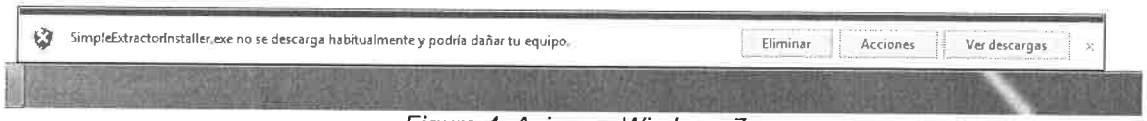


Figura 4. Aviso no Windows 7

Clique em **Acções** e aparecerá uma janela (Figura 5), onde deve selecionar **Mais Opções**. De seguida, clique em "**Executar de qualquer forma**" (Figura 6), para iniciar o processo de instalação do *Simple Extractor*.

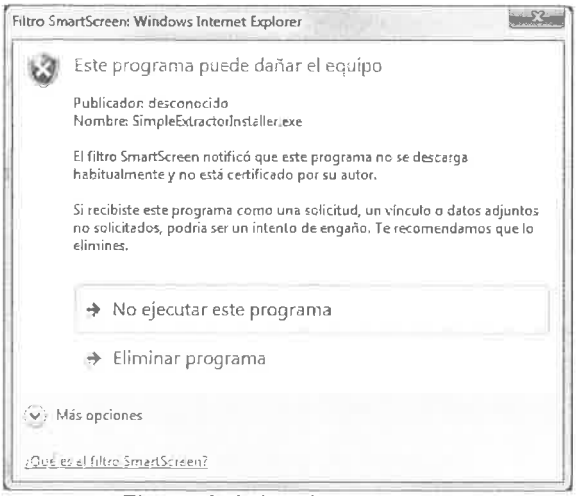


Figura 5. Aviso de segurança

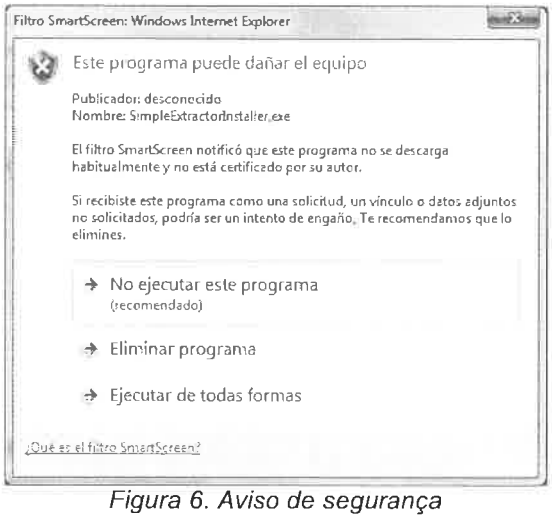


Figura 6. Aviso de segurança

2.1.2 Processo de instalação

O processo de instalação começa com a seleção da língua do Acordo da licença (Figura 7). Aceite os termos da licença e clique em **Seguinte** (Figura 8). Escolha o caminho pré-definido pelo sistema ou se desejar, selecione outro e clique em **Seguinte** (Figura 9). Na janela seguinte (Figura 10) clique em **Seguinte** quando o botão estiver ativo. A configuração dos atalhos que vão ser criados no computador do utilizador efetua-se na janela correspondente à Figura 11; pode aceitar as opções pré-definidas pelo sistema e continuar - desta forma, finaliza o processo de instalação do *Simple Extractor* em Windows (Figura 12).

Figura 7. Seleção do idioma

Figura 8. Aceitar o Acordo de licença