The SWEGOT lexical resources framework

Summary

The ***SWEGOT lexical resources framework*** originates in the SAOLhist application developed in 2012 at the Department of Swedish at the University of Gothenburg.

SAOLhist serves as the retrieval interface for the digitalized lists of headwords in the old editions of *the Swedish Academy Glossary.* The basic functionality includes search and comparison of lexical data and was originally aimed primarily at fulfilling research needs, but the system proved to be of use for a variety of use cases, from debugging of scanned data to word game quick help. The framework provides an easy-to-use tool for accessing and comparing lexical resources bridging differences in spelling, entry structure, etc.

Approach

- Designed for the internet

- Low cost in development and maintenance

- Device independent

- Scalable solution

- Same interface for experts and public

Software requirements

Server side: webserver supporting php (eg apache) and a relational DBMS (eg MariaDB).

Client side: Any device having a browser supporting html5. No cookies are used, JavaScript must be enabled.

Basic Architecture

One software installation can serve several implementations. An implementation consists of a database holding reference tables for a number of "books" (printed or electronic resources) using common key word forms and a common user interface configuration. The basic reference table for a "book" holds information on key word form, original word form, part of speech or other grouping criteria, and page number (when applicable). Aggregated tables are created after loading or reloading basic reference tables.

The implementation configuration file defines such things as language, headlines, menu contents, index page location for facsimiles, external links etc.

Search output is presented as html pages in iframes by functions called using the GET-method. This means that SWEGOT function calls can easily be embedded in other html interfaces if you want a higher degree of control or a "prettier" user interface.

Functionality

The user can choose which (or all) of the 'books' to include in searches. Searches are always performed on both the common key word forms and the individual word entries for each book, i.e. you will rather get too many than too few matches. Every match is linked to either a facsimile image file or an on-line resource.

Advanced search includes support for typical queries such as *'words that only appear in X'* or *'words ending in -ing that appear in X or Y but not in Z'*

Results are displayed as matrices or lists, as chosen by the user. For debugging purposes internal ids can be shown.
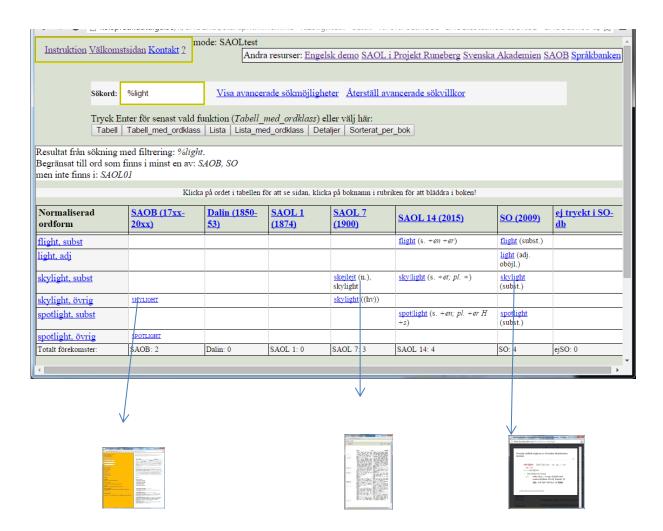
Publicly available implementations

The Swedish Academy Glossary historical site (in Swedish): http://saolhist.se

The official dictionary of Danish standard orthography historical site (imbedded

implementation, in Danish): http://rohist.dsn.dk/

Soon to come, English demo site: http://demo.spraakdata.gu.se/svemm/swegot/

Example search

Result showing words ending in –light, included in one of two specified books, missing from a third, showing POS

(from the internal database at the University of Gothenburg, used by editors and researchers)

References

Diderichsen, Philip, Anna Sofie Hartling, Anne Kjærgaard & Anna Kristiansen. 2015 I ulige linje fra *Linje* til *linje*. Og andre nedslag i Retskrivningsordbøger gennem historien. In: Duncker, Dorthe, Eva Skafte Jensen & Ole Ravnholt (eds.). *Rette ord. Festskrift til Sabine Kirschmeier-Andersen i anledning af 60-årsdagen.* Dansk Sprognævn. Copenhagen.

Holmer, Louise, Monica von Martens & Sven-Göran Malmgren (forthcoming). SAOLhist för allmänt och vetenskapligt bruk. In: Hovmark, Henrik & al. (eds.). *Nordiske Studier i Leksikografi 13*. Copenhagen.