
The SWEGOT Lexical Resources Framework

(Software Demonstration)

Monica von Martens, Louise Holmer

Dept. of Swedish, University of Gothenburg

e-mail: monica.martens@gu.se, louise.holmer@svenska.gu.se

Abstract

The SWEGOT lexical resources framework consists of a database model, sample SQL scripts for database update, configuration text files, and program files written in the PHP programming language providing an easy-to-use web interface for searching, comparing and accessing collections of lexical data, such as glossaries. The database model is flexible and additional resources can easily be added with limited configuration needs.

The user interface provides several methods for limiting the search results and the presentation grid as suitable for the user. An advanced query interface facilitates comparisons between different lexical resources such as ‘entries found in BookA and BookC, but not in BookB’. Actual access to each lexical resource is provided via image files or links to external resources, as defined separately for each ‘book’.

Keywords: Lexical resources; Glossary; Database

1 Background

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 (Holmer, Malmgren & Martens forthcoming). The framework provides an easy-to-use tool for accessing and comparing lexical resources bridging differences in spelling, entry structure, etc.

2 Software Design

2.1 Basic approach

- Designed for the internet
- Low cost in development and maintenance
- Device independent
- Scalable solution
- Same interface for experts and public

2.2 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 and JavaScript must be enabled.

2.2.1 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 shared 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 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.

```

<?php
    $bokar = array('ENG1','ENG2','ENG3' );
    $skortrubrik = array(
        'ENG1' => 'Gloss1',
        'ENG2' => 'Gloss2',
        'ENG3' => 'Gloss3',
    );
    $langrubrik = array(
        'ENG1' => 'old English gloss (17xx-20xx)',
        'ENG2' => 'Contemporary Glossary',
        'ENG3' => 'online lexicon'
    );
    $boktitel = array('ENG1' => 'xx',
        'ENG2' => 'klicka för att bläddra i boken',
        'ENG3' => 'index pages NA'
    );
    $shrefpopup = array(
        'ENG1' => "javascript:visaPopup('demo/ENG1ref.html','ENG1');",
        'ENG2' => "javascript:visaPopup('demo/ENG2ref.html','ENG2');",
        'ENG3' => "javascript:visaPopup('https://simple.wikipedia.org/w
    );
    $stbellnamn = array('ENG1' => 'ENG1wpref',
        'ENG2' => 'ENG2wpref',
        'ENG3' => 'ENG3wpref'
    );
    $vmeny = '<span id="visa_intro"><a href="intro_eng.html" onclick="visaPopup(thi
    $vmeny = $vmeny . ' &nbsp;<a href="index.phtml">welcome page</a>';
  
```

Figure 1: Sample configuration file content

3 User Interface

The user can choose which (or all) of the “books” to include in presentation and searches. Searches are always performed on both the shared 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 id’s

can be shown.

3.1 Sample searches

Figure 2 shows the result matrix for a query ‘words ending in –light, included in one of two specified books, missing from a third, showing POS’ for a non-public database at the University of Gothenburg, used by editors and research staff.

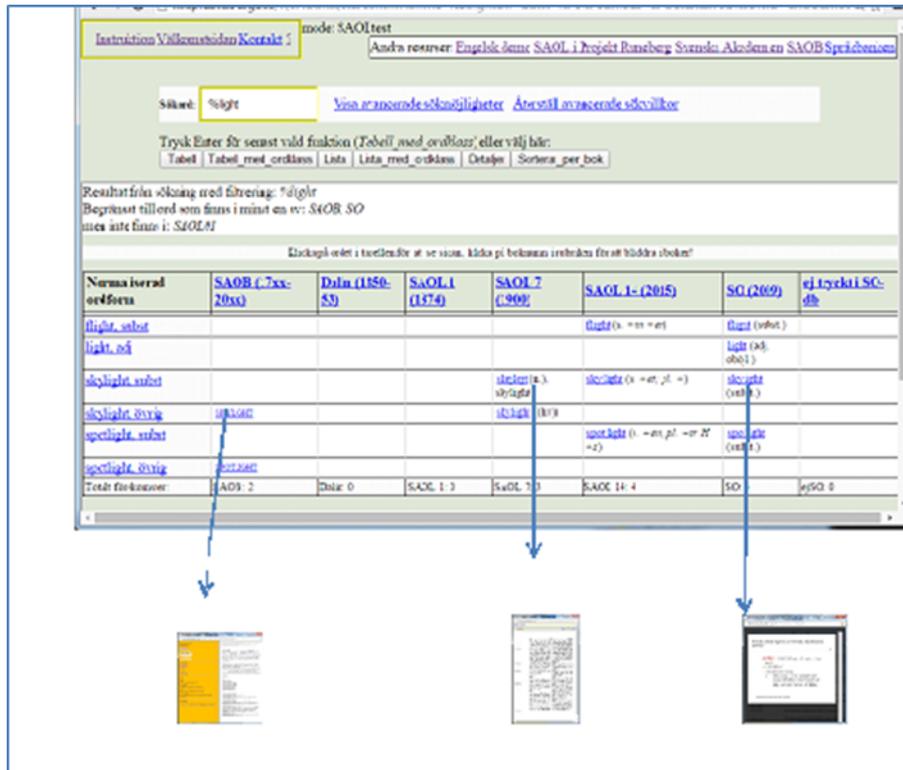


Figure 2: Example Search

3.2 Limiting the result grid

Figure 3 shows the start page where the user can pick which books to include.

Välkommen till SAOLhist

SAOLhist-databasen innehåller material från utgåvorna 1, 6, 7, 8, 9, 10, 11, 12 och 13 av Svenska Akademiens ordlista över svenska språket samt ur *Ordbok öfver svenska språket af A. F. Dalin (1850–1853)*.

Hemsidan vidareutvecklas kontinuerligt och materialet är endast delvis korrekturläst, störande fel får gärna rapporteras och kommer att rättas lopande.

Välj vilka böcker du vill söka i:

- Dalin (1850-53)
- SAOL 1 (1874)
- SAOL 6 (1889)
- SAOL 7 (1900)
- SAOL 8 (1923)
- SAOL 9 (1950)
- SAOL 10 (1973)
- SAOL 11 (1986)
- SAOL 12 (1998)
- SAOL 13 (2006)

Gå till söksidan

Om du endast är intresserad av att titta i en enskild bok, klicka på boknamnet för direktaccess.
För att söka i en eller flera böcker, kryssa för de utgåvor du är intresserad av och tryck på knappen **Gå till söksidan**.

Läs mer om SAOL
(nedladdning 9Mb startas automatiskt):

För synpunkter och felrapporter, klicka här.

SAOLhist-projektet har finansierats av Svenska Akademien och Göteborgs universitet.
Systemdesign och innehåll: Lexikaliska institutet i samarbete med Språkbanken och Projekt Runeberg
Licens CC-BY gäller för sökgränssnitt och index

Figure 3: Picking the books to be included

4 Implementations (publicly available)

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/>, described in Diderichsen et al (2015).

English demo site: <http://demo.spraakdata.gu.se/svemmm/swegot/>

5 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, Dorte, Eva Skafte Jensen & Ole Ravnholt (eds.). *Rette ord. Festskrift til Sabine Kirschmeier-Andersen i anledning af 60-årsdagen*. Dansk Sprognævn. Copenhagen, pp. 97-107.
- 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.