
Advancing Search in the Algemeen Nederlands Woordenboek

Carole Tiberius, Jan Niestadt, Lut Colman, Boudewijn van den Berg
Institute of Dutch Lexicology, belberg
{carole.tiberius,jan.niestadt,lut.colman}@inl.nl, boudewijn@belberg.eu

Abstract

In this paper, we will take a closer look at the advanced search option in online dictionaries from the perspective of the user. In the context of dictionaries, the advanced search allows users to retrieve sets of words which match a particular description, for example ‘archaic compounds consisting of three syllables’. However, the possible sets of words that could be retrieved are endless, and the challenge is how to present all these options to the user in a way that he can grasp and understand. Studies on dictionary use and log file analyses suggest that existing solutions offered by online dictionaries are not very successful as users do not really seem to use this search option. We will discuss different types of advanced search that can be distinguished and we will present a new, more user-friendly approach to advanced search in dictionaries using the *Algemeen Nederlands Woordenboek* as our test case.

Keywords: advanced search; online dictionaries; user-friendly.

1 Introduction

Comprehensive scholarly dictionaries contain a wealth of information. They do not only provide information on the meaning of a word, but they also contain information on morphology, pronunciation, etymology, pragmatics etc. With the online medium it is theoretically possible to let the user search on all the information available in the dictionary database. Most online dictionaries attempt to do this by offering an advanced search option allowing users to retrieve sets of words matching a particular description, for example ‘the set of Dutch nouns that can have more than one plural ending’ or ‘archaic compounds consisting of three syllables’. However, the sets of words that could possibly be retrieved are endless, and the challenge is how to present all these options to the user in a way that he can understand. Studies on dictionary use suggest that users do not really use the advanced search option. This is also the case for the *Algemeen Nederlands Woordenboek* (ANW) where log files show that the advanced search only accounts for 3% of all searches in the dictionary (Tiberius & Niestadt, To Appear). There may be different reasons for this: a) users do not use the advanced search as they are not familiar with such a search option in the context of dictionaries; b) users do not understand the interface that is used for the advanced search; or c) a mixture of these.

In this paper we will present an approach for a more user-friendly advanced search option for the ANW. First we define what advanced search is, then we will discuss different types of advanced search and finally we will present a new approach to advanced search in dictionaries.

2 Advanced search in electronic dictionaries: a definition

In order to define what is meant by advanced search and what user requirements it fulfils, we have looked at what different dictionaries have to say about their advanced search. The **Oxford English Dictionary** (OED) defines its advanced search as follows:

Advanced search is a full search of the entire dictionary text. It finds your term wherever it occurs in the dictionary. This could be in the form of an entry name, part of another word's definition, in a quotation, etc. An advanced search also allows you to search for words that occur near one another, such as bread before butter.¹

The German **exlexiko** dictionary offers an advanced search which allows users to search for words on the basis of specific criteria such as orthography, grammar and word family.

Die erweiterte Stichwortsuche in exlexiko erlaubt dem Benutzer, Stichwörter mit bestimmten Kriterien aus den Bereichen Orthografie, Wortartzugehörigkeit, Grammatik, Sinnverwandtschaft oder Zugehörigkeit zu einer semantischen Klasse zu suchen.²

The **Trésor de la Langue Française Informatisé** (TLFi) offers what it calls an assisted search ('recherche assistée') and a complex search (recherche complexe'). The assisted search allows the user to search the dictionary articles on the basis of a number of criteria:

Permet de rechercher **à travers tout le TLF** les articles correspondant à **plusieurs critères**.

Quelques possibilités:

Quels sont les mots d'origine espagnole ?

Quels sont les exemples de Zola illustrant un sens ironique?

Quels sont les verbes utilisés dans la marine pour la manoeuvre des voiles?

Quelles sont les expressions contenant le mot singe?³

The complex search in the TLFi is similar to the assisted search, but is described as being even more powerful.

The **Algemeen Nederlands Woordenboek** does not employ the term advanced search, but offers four types of search of which the option to search from features to words is the most advanced.

1 <http://www.oed.com/public/advancedsearching/advanced-search/> [10/04/2014]

2 <http://www.owid.de/erweitert.jsp> [10/04/2014]

3 <http://atilf.atilf.fr/dendien/scripts/tlfiv5/showp.exe?13;s=2657178285;p=aide.htm> [10/04/2014]

This search option is the most advanced way of searching the ANW. Through different kinds of information that is stored in the dictionary articles you can search for words, idioms and proverbs. The possibilities to search for information are almost infinite. You can search for information which can occur anywhere in the article or you can search for information in a specific field.[...]

This is a search option which requires a certain amount of creativity from the user and works better the more one gets familiar with the system. We suggest that you take some time to try out this search option.⁴ The last sentence shows clearly that we were rather idealistic when we first developed this search option for the ANW back in 2009.

Summarising, advanced search can be described as a powerful and complex search option that allows users to examine the dictionary using many different criteria.

3 Types of Advanced Search

Advanced search can be realised in different ways. We identify four types, i.e.

- **Classical/traditional advanced search:** where boxes and dropdown lists together form the search query;
- **Faceted search:** a step by step search where the user gradually refines the query by adding criteria (e.g. shopping websites);
- **Wizard search:** a step by step search where the user is given a sequence of questions and no intermediate results are shown (e.g. Foreign Labor Certification⁵);
- **Query language:** single search box which offers many possibilities, but which is hard to learn and remember.

A fifth type could be identified, i.e. natural language queries. However, we do not consider this option here, as we do not believe that the current state of technology is advanced enough for this to be a viable candidate in the context of online scholarly dictionaries.

From these four types, the classical advanced search is the most popular among scholarly e-dictionaries at the moment (cf. the ANW, the OED, *lexiko* and the complex search in the TLFi). The TLFi also offers a wizard-like search ('recherche assistée') by showing a sequence of questions which are to guide the user to an answer.⁶

The DWDS (das *Digitale Wörterbuch der deutschen Sprache*) is an example of a dictionary that does not offer a separate advanced search option but offers a query language to give the user more flexibility. For instance, the query "Stein with \$p=NN" searches for occurrences of the lemma *Stein* 'stone' as a noun.

For the ANW too, a custom query language, called FunQy ('functional query language'), was developed to power its traditional advanced search option (Niestadt et al. 2009). As an undocumented feature, users can play with this query language themselves. At one time we planned for this query language to

4 <http://anw.inl.nl/show?page=help#zoek3> [10/04/2014]

5 <http://www.flcdatacenter.com/OESWizardStart.aspx> [10/04/2014]

6 <http://atilf.atilf.fr/dendien/scripts/tlfiv5/showp.exe?30;s=1771472760;p=assiste.htm> [10/04/2014]

be used internally at the Institute of Dutch Lexicology (INL), but this never materialised. The FunQy query language may be too complex even for language professionals to use.

4 A new way of advanced search

We now turn to our approach to realise a more user-friendly advanced search option for scholarly dictionaries. We believe that it is a mistake to think that one single advanced search option can appeal to all users equally. A full-featured search may be convenient for frequent users, but new users will most likely be put off by its complexity. However, if you do your best to capture new users with a friendly, step-by-step approach, experts will get annoyed by how much clicking is required to perform common searches. This means there is no single best approach; you have to compromise, or develop separate interfaces for different users. We decided to clearly identify the target users of the ANW to be linguists and academics more generally, who want powerful search features, but are intimidated by our current interface (see Figure 1).

The features that can be searched for are presented in a tree structure on the left of the screen. This tree structure is the same as the one used to structure the dictionary articles (as seen on the article screen). It starts with syntactic category and then spelling and pronunciation, etc. The user starts with an empty query screen and is asked to select criteria from this tree structure on the left. As soon as the user selects a criterion, a query appears on the right-hand side of the screen. By default, the user searches for words, but it is also possible to search for proverbs or idioms. This will result in a tree structure with different criteria as only a subset of the criteria that can occur in a query for words apply to idioms and proverbs.

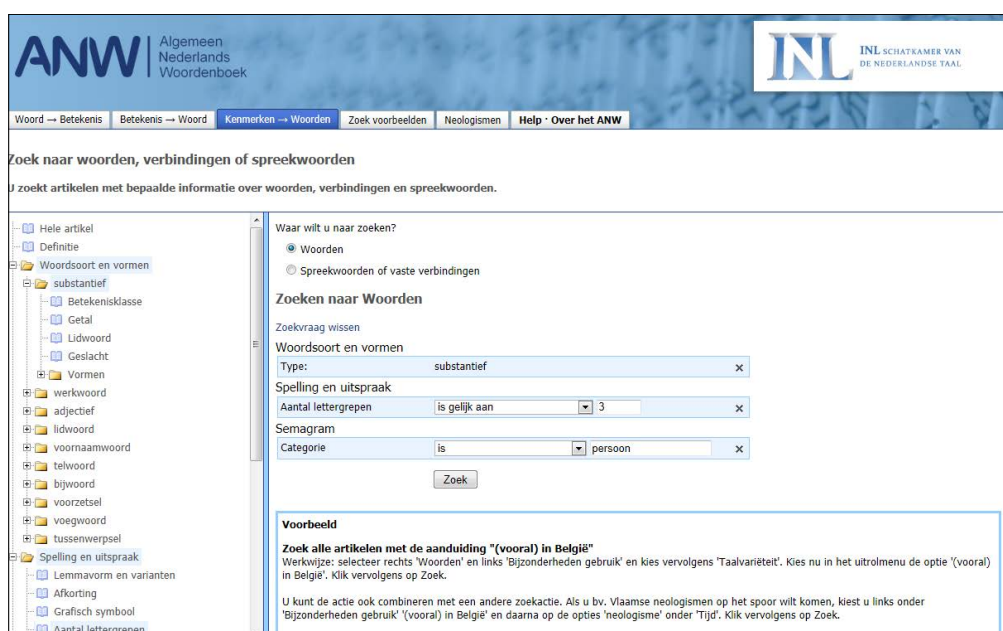


Figure 1: Screenshot Features → Words in the ANW.

Our aim was to combine the best parts of the different types of advanced search discussed in Section 3 in our solution. A very preliminary version of our approach has been presented at the eLex 2013 conference.⁷ We are now in the process of developing a full-working prototype which will be accessible through the ANW website.

The opening page of the ANW remains as it is with a simple search box in the center of the screen for searching a word (or multi word expression). However, below the search box a link will be added for users who are looking for something else. After clicking on this link, users are asked what they would like to search for. They can choose a question or criterion from a scrollable list, or they can type in a word to filter the list. The current list covers spelling, pronunciation, combinations and pragmatics, and has been defined on the basis of the criteria offered in the current search interface of the ANW web application. In future, this list will be further expanded and refined.

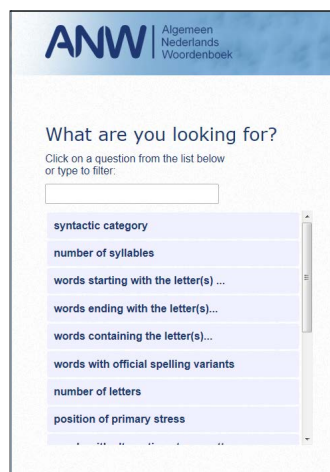


Figure 2: Prototype of list of questions users may like to ask.

For convenience, the most frequently used questions will appear at the top of the list so that users see these first. To make the search as effective and user-friendly as possible (Lew 2012) functionalities such as autosuggest, fuzzy matching and smart filtering will be integrated. Thus, the filter box will also respond to terms that do not literally occur in the visible descriptions because hidden tags specifying the category to which questions belong have been added (e.g. typing ‘morphology’ also shows questions such as ‘words derived from ...’). Clicking on a question opens up the search form, with a single input box for that question. As soon as the user types something in this search box, the results are shown on the right-hand side of the screen together with the relevant information from the dictionary entry: for instance, if the user was searching for words consisting of four syllables, the syllable structure of the resulting words is shown. This direct feedback makes this ‘advanced’ search faster and less intimidating as the user has direct access to the information he is interested in.

7 [http://eki.ee/elex2013/\[10/04/2014\]](http://eki.ee/elex2013/[10/04/2014])

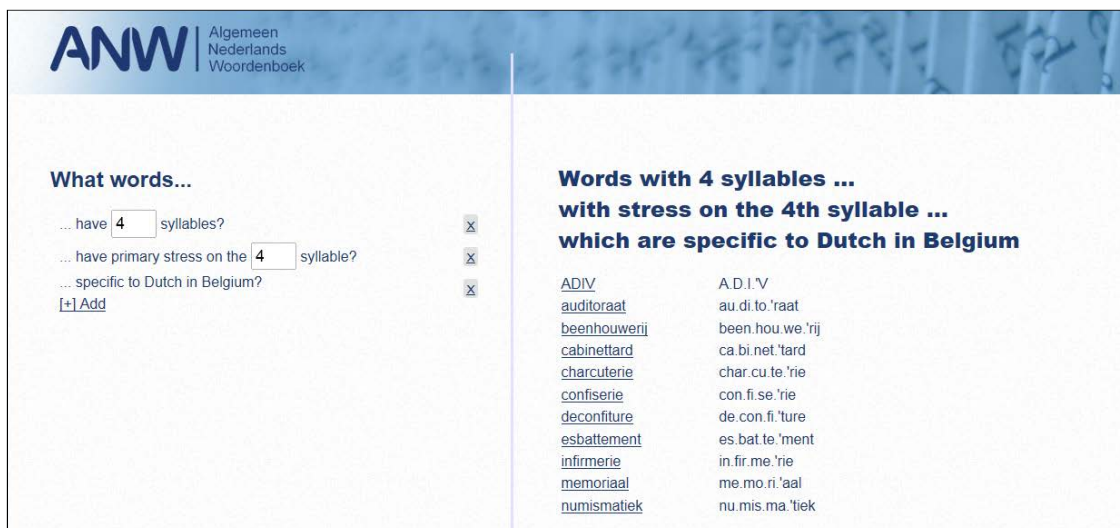


Figure 3: Prototype of what a search query plus results may look like.

If the user wants to refine his query, he simply clicks 'Add' and goes again to the list of search questions where he can select another question to add to his query. As a bonus, an expert who regularly uses the same criteria with different values can bookmark the page in its current state, so he can avoid the ten clicks required to get here.

5 Conclusion

We have presented an approach for a more user-friendly advanced search option in the ANW. It combines the best aspects of the advanced search types we have discussed in Section 3. It is like a wizard, because the user is guided through constructing a query. It also includes the advantages of a faceted search because there is direct feedback when you add or change something. The possibility to bookmark a search query makes it again very similar to the classical advanced search.

So far user research has not been carried out, but this is planned for the near future when the prototype is up and running.

6 References

- Algemeen Nederlands Woordenboek*. Accessed at: <http://anw.inl.nl> [10/04/2014].
- das Digitale Wörterbuch der deutschen Sprache*. Accessed at: <http://www.dwds.de> [10/04/2014].
- elexiko*. Accessed at: <http://www.owid.de/wb/elexiko/start.html> [10/04/2014].
- Lew, R. (2012) How can we make electronic dictionaries more effective? In Sylviane Granger and Magali Paquot (eds.) *Electronic lexicography*. Oxford. 343-361.

Niestadt, Jan, Carole Tiberius & Fons Moerdijk (2009). Searching the ANW dictionary. Poster presented at eLexicography in the 21st century. Louvain-la-Neuve.

Oxford English Dictionary. Accessed at: <http://oed.com> [10/04/2014].

Tiberius, C and J. Niestadt (To Appear) Dictionary Use: A Case Study of the ANW Dictionary. In: Tiberius, C. and C. Müller-Spitzer (eds.) *Wörterbuchbenutzungsforschung* 5. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. Mannheim: Institut für Deutsche Sprache. (OPAL – Online publizierte Arbeiten zur Linguistik).

le Trésor de la langue française informatisé. Accessed at: <http://atilf.atilf.fr/> [10/04/2014].

