Advanced graph-based searches in an Internet dictionary portal

Peter Meyer

Institut für Deutsche Sprache, Mannheim
E-mail: meyer@ids-mannheim.de

Abstract

The web portal *Lehnwortportal Deutsch* (lwp.ids-mannheim.de), developed at the Institute for the German Language (IDS), aims to provide unified access to existing and possibly new dictionaries of German loanwords in other languages. Internally, the lexicographical information is represented as a directed acyclic graph of relations between words. The graph abstracts from the idiosyncrasies of the individual component dictionaries. This paper explores two different strategies to make complex graph-based cross-dictionary queries in such a portal more accessible to users. The first strategy effectively hides the underlying graph structure, but allows users to assign *scopes* (internally defined in terms of the graph structure) to search criteria. A second type of search strategy directly formulates queries in terms of the relational graph structure. In this case, search results are not entries but n-tuples of words (metalemmata, loanwords, etyma); a query consists of specifying properties of these words and relations between them. A working prototype of an easy-to-use human-readable declarative query language is presented and ways to interactively construct queries are discussed.

Keywords: graph database; loanword lexicography; search technology

1. Introduction

The *Lehnwortportal Deutsch* (lwp.ids-mannheim.de) is a freely accessible online lexical information system, developed at the Institute for German Language (IDS), that provides unified access to dictionaries of German loanwords in other languages. As well as conventional access to the individual dictionaries, the portal offers complex cross-dictionary search functionality; in particular, it can be used as an "inverted loanword dictionary" to trace the way of German words into different recipient languages. The portal web software operates on a database that represents pertinent lexicographical information as a cross-dictionary network of relations – more technically, a directed acyclic graph (DAG; cf. Bang-Jensen & Gutin, 2012) – between word forms of all included dictionaries.

This paper focuses on the problem of making complex graph-based cross-dictionary searches in the portal accessible to a wide range of users. In section 2, the general architecture of the *Lehnwortportal Deutsch* is described from a user's point of view. The graph-based structure of the underlying unified data representation used for cross-dictionary searches is discussed in section 3. Section 4 shows how the web portal currently integrates some graph-related concepts in a unobtrusive way into fairly conventional HTML search forms suitable for average users. Section 5

concludes the discussion by outlining an alternative type of search strategy that provides advanced users with the opportunity to directly search the relational graph structure through an easy-to-learn, human-readable query language.

2. Basic access structure of the *Lehnwortportal Deutsch*

2.1 General information on the web portal

In its initial version, released in November 2012,[1] the web portal comprises three dictionaries on German loanwords in Standard Polish (de Vincenz & Hentschel, 2010), in the dialect of Polish spoken around the town of Cieszyn (Menzel & Hentschel, 2005), and in Slovene (Striedter-Temps, 1963). The two Polish dictionaries have previously been published electronically, whereas the Slovene dictionary was integrated through a combination of image digitization and manual extraction of relevant lexicographical information. The system is under active and continuous development and has a modular architecture that allows easy addition of new digital or digitized resources in XML format. In particular, a project is underway to integrate a newly-compiled dictionary of German loanwords in East Slavic languages that were mediated through Polish. There are long-term plans to incorporate a large number of further lexicographical resources on German loanwords in a wide range of other languages of the world.[2]

2.2 Accessing and navigating individual loanword dictionaries

The portal provides uniform access to the entries of all integrated loanword dictionaries. As a first step, a dictionary must be chosen from a menu on the right bar of the web page. In order to look up an entry in the dictionary, users may either type the beginning of a headword into an autocomplete text box or scroll through the alphabetical lemma list after selecting the initial letter in an alphabet bar (see Figure 1).

The microstructure of entries is entirely specific to the individual dictionaries. Due to considerable differences regarding intent, coverage and granularity, no attempt has been made to define a uniform one-size-fits-all entry structure (Meyer & Engelberg, 2010). There is, for each dictionary, a dedicated XML schema for its entry documents and, with the exception of those dictionaries where digitized images of print articles are shown, an accompanying XSLT stylesheet that transforms the XML source of its entries into HTML fragments.

[1] The web portal in its present form has been developed in a project funded by the Federal Government Commissioner for Culture and the Media upon a Decision of the German Bundestag.

[2] So far, there is little web traffic on the portal, possibly due to the limited number of available resources and the highly specialized targeted audience. On average, the number of page visits per day is still well below 100 and the advanced graph-based search options discussed in this paper are consulted less than twice a day.

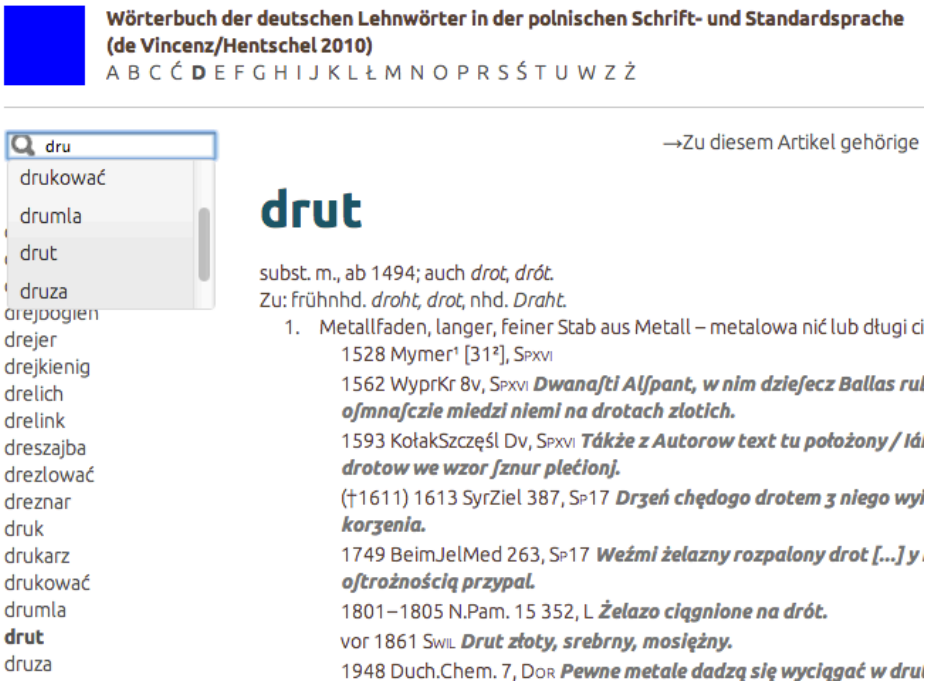

Figure 1: Navigational elements in a sample article

2.3 Etymological metalemmata and the inverted loanword dictionary

The *Lehnwortportal* features an 'inverted' loanword dictionary (Engelberg, 2010) that lemmatizes all words of the donor language, German, that have been borrowed into the recipient languages represented by the different loanword dictionaries included in the portal. The concept of an inverted loanword dictionary was proposed more than forty years ago by Karaulov (1979), but dictionaries of this type are virtually non-existent to this day, with the notable exception of van der Sijs (2010) for Dutch loanwords in the world's languages.

Setting up the inverted loanword dictionary for the *Lehnwortportal* is not a trivial task and cannot be performed automatically since any German etymon may appear in a variety of orthographical, diachronic, dialectal and other forms (henceforth referred to as 'variants' of the etymon) in different entries within and across loanword dictionaries. As an example, Standard Polish *lichtarz* is linked to a Middle High German etymon *liuhtaere* in de Vincenz & Hentschel (2010), whereas Slovene *lajhter* is related to New High German *Leuchter* and Middle High German *liuhtære* in Striedter-Temps (1963). Looking up the contemporary German word *Leuchter* 'candlestick' in the inverted loanword dictionary, the average user may reasonably expect to also be directed to entries that only list the corresponding Middle High German form of *Leuchter* in one of its orthographical variants *liuhtaere* or *liuhtære*. As a solution to this requirement, all German etymon word forms as they appear in the entries of the portal dictionaries were mapped to etymologically corresponding 'normalized' word forms, and wherever possible contemporary Standard German words. These normalized entries, henceforth *metalemmata*, are used as headwords

of the inverted loanword dictionary, whose entries, for the time being, mainly consist of hyperlinks to all loanword dictionary entries that list the metalemma or any of its diachronic, dialectal or other variants as an etymon. For each link, the corresponding German words in the target entry are given together with their definitions, if present.

Defining and mapping metalemmata involves many subtle philological and lexicographical problems and requires linguistically informed manual work. As the list of metalemmata grows rapidly with each newly included dictionary, and may require complex editing and correcting, using an administrative software tool for these tasks is indispensable. For the purposes of the initial version of the *Lehnwortportal*, a Java desktop application was developed that simply stores all information on metalemmata together with references to the exact places of corresponding etyma in the XML source documents in a separate file (henceforth 'metalemma file'). The metalemma administration tool is also used to edit the cross-references within the metalemma list; thus, it is possible to mark a metalemma as a morphological derivative or constituent of another metalemma. This kind of internal cross-referencing is a prerequisite for finding loanwords borrowed from compounds or derivatives of a given German word. In a more advanced multi-user setting, however, a database solution would be more appropriate than locally editing a file.

The presentation of each loanword dictionary entry in the portal is complemented by links to all German metalemmata that correspond to etyma appearing in the entry. This information is dynamically constructed from the information contained in the inverted loanword dictionary. There may be references to multiple metalemmata for a given entry in case the entry discusses borrowings from several different, possibly morphologically related, etyma.

3. Using a directed acyclic graph (DAG) for unified data representation across heterogeneous resources

One of the distinctive features of the *Lehnwortportal* is the possibility of powerful cross-dictionary searches. Apart from obvious performance considerations, there are two lexicographical obstacles to using the unaltered XML source documents of the various component portal dictionaries for portal-wide search processes (cf. Meyer, 2013 for details):

(i) As mentioned, the individual dictionaries differ widely with respect to the microstructure of their respective entries (as reflected in the dictionary-specific XML schemata). Put simply, information of a certain kind can usually not be found "at the same place" in XML documents belonging to different dictionaries.

(ii) The terminology, concepts and data formats for specifying, e.g., the time of borrowing, grammatical features, or dialect appurtenance may vary considerably between dictionaries.

As a consequence of (i), an additional layer of lexicographical data is needed that represents relevant information of all component dictionaries in a unified structural format amenable to fast and efficient database queries. The solution opted for in the *Lehnwortportal* is to represent this lexicographical information as a network of relations (such as 'is borrowed from' or 'is a derivative of') between word forms (metalemmata, etyma and loanwords as well as their respective variants, derivatives etc.). To overcome the problem stated in (ii), the words that form the vertices of this network are annotated with grammatical, diasystemic and other information that is extracted from the original lexicographic resource and translated into a uniform data format.

More formally, advanced searches in the portal operate on a directed acyclic graph (DAG) whose vertices are word forms and whose edges are relations between word forms.[3] At present, the following types of relations between two word forms *x* and *y* are used in the DAG:

- etymon *x* is mapped to metalemma *y*;
- loanword *x* is borrowed from etymon *y*;
- etymon or loanword *x* is an (orthographical, phonological, ...) variant of etymon/loanword *y*;
- *x* is a derivative of *y*;
- *x* is a compound of which *y* is a constituent;
- *x* is an etymologically related lexical parallel to *y* in another language (relevant for entries in Menzel & Hentschel, 2005).

In what follows, we will call *x* the 'child' and *y* the 'parent' of the relations enumerated above; in obvious graph-theoretical fashion, we will call the transitive generalizations of these terms 'descendant' and 'ancestor', respectively.

The DAG completely abstracts from the micro- and macrostructural idiosyncrasies of the individual component dictionaries; instead, it is generated in a fully automated process from parsing the underlying dictionary data and the metalemma file mentioned above. From the XML source of each dictionary entry in the portal (at least) one subgraph of the DAG – containing a loanword and its German etymon together with variants, derivatives etc., of either – is constructed in a dictionary-specific way. Roughly speaking, relations between word forms (edges in the DAG) are deduced from dictionary-specific structural relations between the corresponding XML elements or attributes.

[3] A DAG has also been employed in the construction of the *Wörterbuchnetz* (http://woerterbuchnetz.de/) by the Trier Center for Digital Humanities, but its vertices correspond to dictionary entries, not individual words within entries (cf. Burch & Rapp, 2007).

Information from the metalemma file is used to connect etymologically-related subgraphs extracted from different entries and/or dictionaries – whose sources (vertices with in degree 0) are German etyma – in order to create larger, possibly cross-dictionary subgraphs whose sources are metalemmata. The web portal offers interactive visualizations of these larger subgraphs on the entry pages for the respective metalemmata, thus making it possible to get a visual impression of borrowings from a German word (cf. Figure 2).

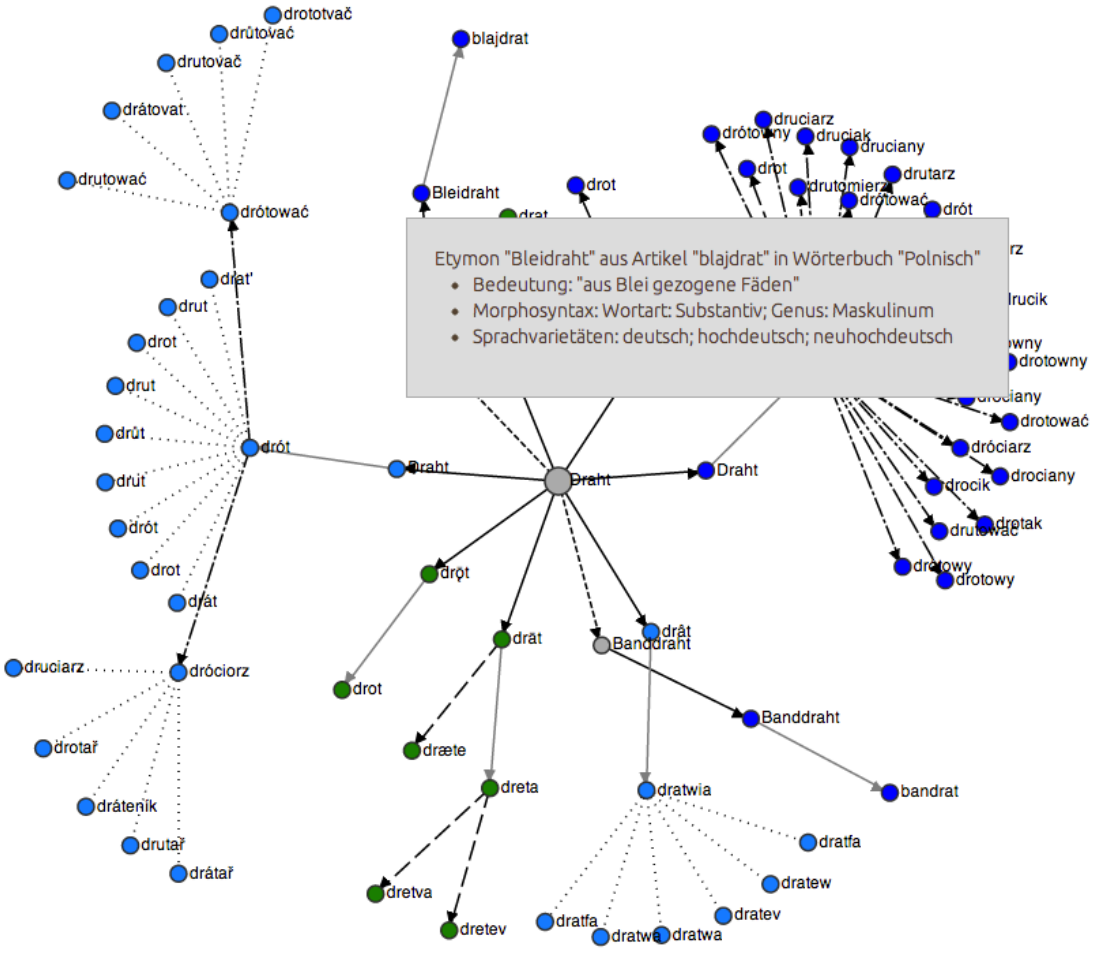

Figure 2: Interactive graphical representation of the subgraph related to the German metalemma *Draht* 'wire'

As stated above, all vertices (word forms) of the DAG are annotated with morphosyntactic, diasystemic and meaning information in a standardized cross-dictionary format. This implies that for each dictionary an automated procedure has to be defined that translates lexicographical specifications from the dictionary-specific format into the standardized one. The intricacies involved in this task will not be discussed here; just one example: the German language variety

(dialect, historical stage) of an etymon may be used as a search criterion in portal-wide queries; therefore, a unified taxonomy of such varieties was defined for the *Lehnwortportal*, with each dictionary-specific language/variety name (e.g., 'Silesian') being mapped onto a set of ever-narrower language categories that can be used in searches (e.g., *High German, Central German, East Central German, Silesian German*). As soon as 'fuzzy' categories such as date of borrowing come into play, the picture gets considerably more complicated: if the etymon of a loanword *w* is just tagged as 'Middle High German' in the original entry, a query for words borrowed from German between 1300 and 1700 should return *w*, if possible with a low rating or weight. One possible way to account for such cases would be the use of a fuzzy ontology (Sanchez & Yamanoi, 2006).

A major advantage of using a DAG in the context of loanword dictionaries is the ability to adequately handle chains of borrowings in forthcoming extensions of the web portal. Thus, the Polish loanword *drukarz* 'printer (profession)' was borrowed from German *Drucker* and served in turn as the etymon for Ukrainian *drukar*. The indirect borrowing relationship between the Ukrainian and the German word is neatly expressed by a path in the DAG: *Drucker* (German metalemma) → *Drucker* (German etymon) → *drukarz* (Polish loanword) → *drukarz* (Polish etymon) → *drukar* (Ukrainian loanword). Note how the Polish intermediate appears twice in this graph on account of its dual role: as a German loanword it is a headword in a Polish loanword dictionary, and as the etymon for a Ukrainian loanword it appears in an entry of a Ukrainian loanword dictionary. It is even possible that these two dictionary entries contain contradictory information on the lexeme in question. Identifying the two words through a relationship 'etymon *x* corresponds to loanword *y* in a borrowing chain' is therefore additional information that has to be added to the lexicographical database by an expert lexicographer.

Note that the DAG is not a standalone database resource; it has to be recreated each time one of the underlying resources (including the correspondence information just mentioned) is altered or a new resource is added to the portal.

At present, the DAG is stored in a standard relational database, basically using two tables, one for the vertices and their properties, and one for representing the directed edges (relations between words) as ordered pairs of vertex IDs. The database does not only store all direct relations (edges) between words as enumerated above, but also their transitive closure, i.e., all indirect ancestor-descendant relations are also stored, which improves lookup times for complex queries. There are plans to migrate to a dedicated graph database such as *Neo4j* in the near future.

The overall architecture of the portal as outlined above, with its combination of heterogeneous XML-based resources and a uniform cross-resource DAG representation of both micro- and mediostructural information, is obviously applicable to other projects where unified access and search structures for interlinked

heterogeneous lexicographical resources are required. From a technical point of view, however, creating a programmatic abstraction layer that separates the backend, database-related core technology from specific issues of the *Lehnwortportal*, such as the specific lexicographical toolchain and the particular web application framework used for the portal, is not a trivial task and has not been accomplished so far. Publishing such an abstraction layer as an open source Java library is a long-term goal of the *Lehnwortportal* project.

4. Graph-based searches for the layman: Hiding the complexity

Adding a DAG-based homogenized data layer to the *Lehnwortportal* opens up a range of new possibilities for advanced cross-dictionary queries, but also increases the complexity for the average user who might not wish for graph-based data modeling just for moderately complex searches. So the question naturally arises as to how to reconcile usability requirements with the inherent complexity of data representation. In this section, we discuss the strategy that is pursued in the present version of the portal, i.e. using a fairly standard form-based search interface that effectively hides the underlying graph structure from the user. The HTML form for advanced portal-wide searches (http://lwp.ids-mannheim.de/search/meta) is split into three sections. In the initial default view, the topmost section offers users four search options for German etyma, viz. (a) an input field for specifying the etymon word form or its initial, final or middle part; (b) an input field for specifying a search string within the definition of the etymon; (c) a drop-down list of German varieties (mostly dialects and language stages) the etymon might belong to; and (d) a drop-down list of possible grammatical and morphosyntactical characteristics (such as POS, gender) of the etymon. The middle section offers analogous search criteria for loanwords. The bottom section permits a choice between two different modes of presentation for search results: per default, all matching entries in all loanword dictionaries are shown in alphabetical order of their respective headwords; alternatively, the set of matching metalemmata from the inverted dictionary can be displayed.

A loanword dictionary entry is considered matching if and only if it contains both an etymon (including variants etc.) and an *associated* loanword (again including variants, derivatives etc.) that both match their respective search criteria. A loanword *L* is considered associated with an etymon *E* if and only if *E* and *L* have a German metalemma *M* as a common ancestor in the DAG. *M* is called a matching metalemma for the search. The requirement that *L* must be associated with *E* is not trivial since a dictionary entry might discuss several etymologically different loanwords with their respective etyma. The condition for being associated is certainly not the most obvious one (which would be to have *E* as an ancestor to *L* in the DAG) but has the advantage of being less sensitive to the exact structure of the DAG: if, for instance, *L*'s etymon is represented as a variant of *E* in the DAG, this does not necessarily imply that *E* itself

cannot also be called an etymon for *L*; a lot depends on the lexicographical practice and granularity of each individual loanword dictionary.

Internally, each query returns all matching etymon-loanword pairs together with their respective matching metalemmata. Depending on the selected presentation mode, either the entries corresponding to the etymon-loanword pairs or the metalemmata are shown. In the metalemma search mode, all matching etymon-loanword pairs, sorted by dictionary entry, can be displayed. Thus, the underlying search is formulated and executed in graph-related terms: the etyma-loanword-metalemma triples correspond to *subgraphs* of the DAG. From the user's point of view, however, only a simple conjunction of search criteria concerning etyma and/or related loanwords is specified as a query, the search result being a straightforward list of dictionary entries. As an example, Figure 3 shows a simple query for dictionary entries containing both a German etymon whose definition contains the word *Metall* 'metal' and an associated loanword that is a Polish noun. Neither the search form nor the search result (a list of links to dictionary entries) refers explicitly to graph-theoretical concepts, although they are implicit in the requirement that matching loanwords must somehow 'belong to' matching etyma.

For even more advanced queries, all eight search fields in the HTML form can be expanded to yield a conjunction of at most 16 search criteria altogether. Each criterion in turn can be a conjunction or a disjunction of two similar criteria (e.g., 'is a noun OR is a verb') and, more importantly, can be assigned what will be hereafter referred to as a *scope*. Apart from default scope (meaning that the criterion applies to the etymon or loanword in question) a user can assign *entry scope* or *portal scope* to any criterion. In this way, it is possible to additionally specify properties of *other* loanwords or etyma that are *associated* with the etymon-loanword pair in question and that appear either elsewhere within the entry (entry scope) or in any arbitrary dictionary entry of the portal (portal scope). Again, being associated is defined with respect to the DAG as having a common metalemma ancestor. A typical scenario for using a wider scope might be a search for loanwords that have derivatives or compounds with certain properties. Figure 4 presents a sample extension of the query shown in Figure 3 requiring that matching entries include an etymologically related word ending in *-owy* or *-owny* (both are typical denominal adjective suffixes in Polish). A reasonable example for a criterion with portal scope would be 'language: Slovene' in the loanword section; this amounts to the requirement that there be an etymologically-related loanword in Slovene.

The idea of 'annotating' search criteria could easily be extended to cover the problem of handling borrowing chains: users may wish to specify whether a certain criterion applies to intermediate or to terminal etyma or loanwords in a chain.

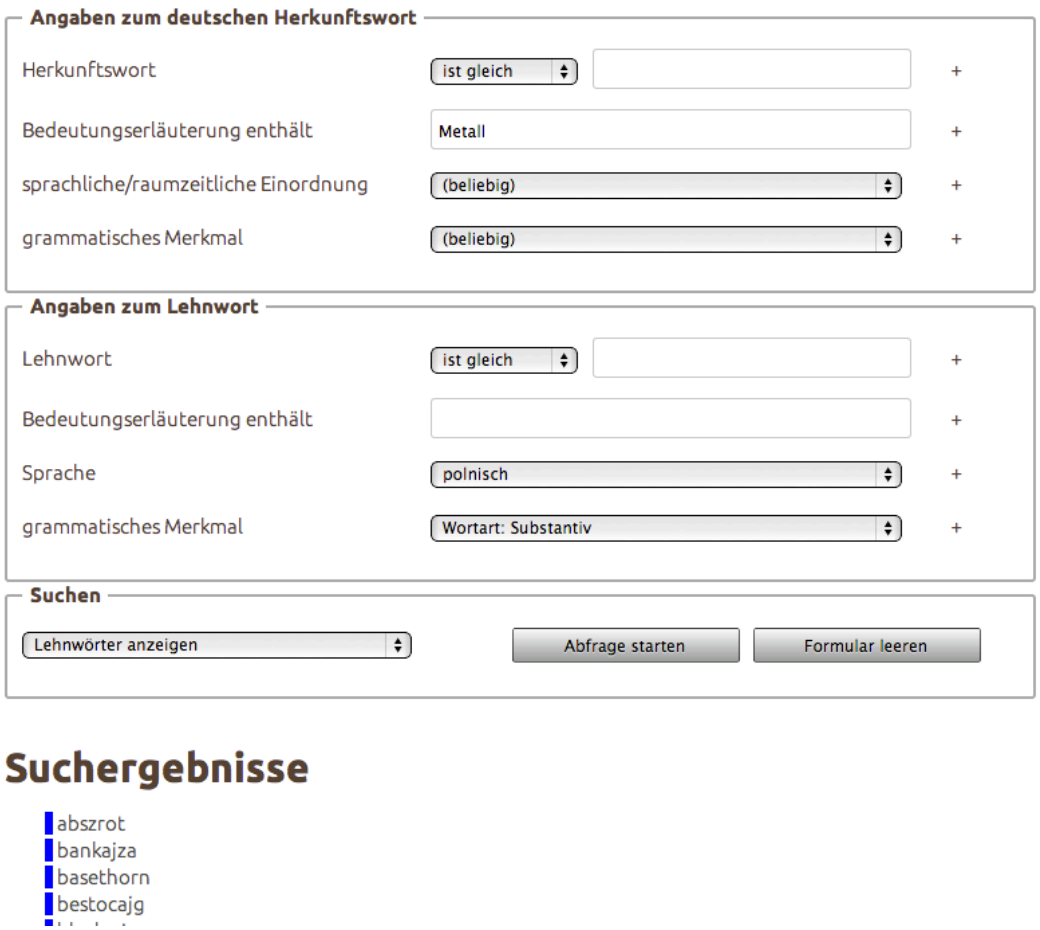

Figure 3: Example of an advanced cross-dictionary search query in the *Lehnwortportal*

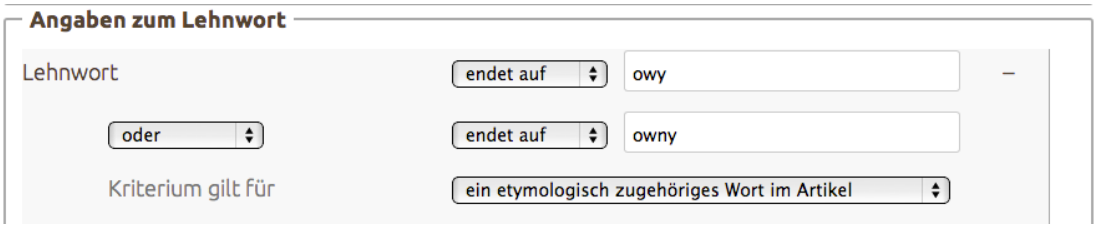

Figure 4: Assigning a scope to a search criterion

As a downside of this approach, queries might return surprisingly complex semantics. To really understand the results returned, the user has to be aware of the fact that the underlying query is formulated in terms of etymon-loanword pairs. Suppose, for instance, that only one criterion *C* is specified in the loanword section of the HTML form and that it happens to have entry scope. If at least one relevant loanword *L* in a dictionary entry complies with *C*, then the underlying result pairs every etymon *E* in this entry that matches the etymon-related search criteria, with *all* loanwords in the same entry that are associated with both *L* and *E*. This is in contrast to the case of

default scope of *C* where only those loanwords that fulfill *C* can be a component of the etymon-loanword pairs returned. Even more confusing is that the list of dictionary entries presented as the search result to the user is the same in both cases (default vs. entry scope of *C*); this is because in both cases the only loanword-related requirement is that matching entries contain at least one loanword fulfilling *C* and be associated with an etymon matching the other search criteria.[4]

Another restriction is that multiple criteria with extended scope cannot be made to refer to the same words. Thus, if a user assigns entry scope to two loanword-related criteria (such as 'language: Polish' and 'POS: adjective') this does not equate to the requirement that there be an etymologically-related Polish adjective in the entry; rather, it simply means that among the loanwords in the article there must be both an adjective and a (possibly identical) Polish word. Of course, it would be possible to refine the annotation scheme to cover at least the most useful relations between scoped criteria, but at the cost of reduced usability.

5. Graph-based searches for professionals: Using a declarative domain-specific query language

Under the hood, advanced searches in the *Lehnwortportal* as outlined above are all based on the graph-theoretical notion of a common ancestor of two or more nodes. To unleash the full range of structural search possibilities it is desirable to have the possibility of formulating queries directly in terms of arbitrary graph configurations.

For this kind of search technology to be accessible to interested professionals without IT background, an easy-to-use human-readable query language should be employed that allows the user to describe the properties of the subgraphs s/he is looking for. The language should be *declarative* in that the actual process of finding subgraphs with the desired properties in the DAG need not be defined by the user. The following remarks report on the results of some preliminary research work on a tailor-made query language for the *Lehnwortportal*.

Most currently used generic query languages (cf. Wood, 2012, for an overview) for graph databases are geared towards IT professionals, typically having an SQL-like syntax, like the *Cypher* language for the Neo4j database (see http://www.neo4j.org/learn/cypher; cf. Robinson et al., in press). The approach taken for the *Lehnwortportal* was to design a highly domain-specific language whose expressions are actually very close to human language; furthermore, complex queries should be expressible through an unordered list of short 'sentences' that can easily be adapted from some sample set. Here is how a query in such a language might appear for the search task that was used as an example above:

[4] As a convention in the *Lehnwortportal*, at least one criterion in an advanced query must have default scope because otherwise search results can easily get incomprehensible.

```
/* (1) Declare node variables: */
find metalemma metaLemma.
find etymon metalWord.
find loanword polishNoun.
find loanword polishAdj.
find loanword sloveneWord.

/* (2) Define relations between words: */
metaLemma is metalemma for metalWord.
polishNoun is descendant of metaLemma.
polishAdj is derivative of polishNoun.
sloveneWord is descendant of metaLemma.

/* (3) Express constraints on words: */
definition of metalWord contains 'Metall'.
language of polishNoun is Polish.
part of speech of polishNoun is noun.
part of speech of polishAdj is adjective.
polishAdj ends in 'owy'
or polishAdj ends in 'owny'.
language of sloveneWord is Slovene.

/* (4) Define how results are shown: */
show metalWord, polishNoun, polishAdj.
```

This query is obviously both more precise and semantically more perspicuous than its HTML form-based counterpart. Each query expression consists of an unordered list (a conjunction) of *clauses*, each ending with a period, that together specify a 'graph pattern' for subgraphs of the DAG. This is close to the syntax of the query language used for the NAGA search engine (Kasneci et al., 2008) with an additional layer of 'syntactic sugar' on top. Internally, the period-delimited clauses are just constituents of the query expression as defined in the context-free grammar for the query language. Strings enclosed between '/*' and '*/' are also constituents and are treated as comments. In (1), the nodes in the graph pattern (word forms) are labelled by user-defined node variables and simultaneously classified as metalemmata, etyma or loanwords. In (2), specific relations between these nodes are defined; edges between two vertices are specified by their type (e.g., 'is derivative of'), while indirect connections through paths of arbitrary length can be given in abstract graph-theoretical terms ('is descendant of'). Properties of vertices (words) are defined in (3). The clause in (4) controls how the search result is to be displayed. Formally, search results are ordered as n-tuples of words (metalemmata, loanwords, etyma) belonging to the appropriate vertices of matching subgraphs. In our example, all matching combinations of three of the five variables are to be shown, ordered alphabetically first by metalWord, then by polishNoun and finally by polishAdj.

The convoluted process of translating such query expressions into native database queries[5] creates a useful layer of domain-specific abstraction from implementation details. One advantage is ease of use: for each of the steps (1) to (4) demonstrated above, users can simply choose component clauses of their queries from a limited number of pre-defined clause templates and combine them, where necessary, with Boolean operators. It is straightforward to construct an interactive drag-and-drop user interface – similar to the Scratch programming environment (http://scratch.mit.edu/) – that guides users through the process of selecting templates and operators and constantly checks for errors such as misspelled variable names, illegal cycles in graph patterns etc.[6] As an additional benefit, it becomes almost trivial to create a multilingual version of the query language.

6. Conclusion: Making complex graph-based searches more accessible

The *Lehnwortportal Deutsch* offers an innovative and principled way of making a portal of heterogeneous lexicographical online resources more than the sum of its parts by providing a unified graph-based database representation of all lexicographical data. The benefits of this approach come at a price – not only on the lexicographical side, but also for the user who has to tackle increased complexity of search options. This paper has shown how the present version of the portal manages to shield users from direct exposure to the graph database, which, however, severely restricts and sometimes obscures the semantics of such queries. An alternative strategy has been outlined that tries to make it as easy as possible to use a graph-based query language. It must be emphasized, however, that both strategies address not casual users but experts who wish to use the portal as a research instrument. Integrating a graph database into a semantic-search system (such as Google Knowledge Graph or Wolfram Alpha) that is suitable for use by laypeople is a much more difficult task.

[5] On a technical note, a parser combinator library is used to construct an Abstract Syntax Tree (AST) from the query expression; the AST is then traversed and processed recursively to generate the underlying database query, at present a SQL query. For each node of the AST, an instance of a certain Java class is created that represents the different parts of the SQL query (select/from/where/order by) as they are partially determined by this node. The object corresponding to the root node of the AST is used to produce the SQL string.

[6] A further step would be the use of a visual version of the query language, comparable to *qGraph* (cf. Blau et al., 2002). Users could then literally draw the query subgraphs using a pointing device and a keyboard.

7. References

Bang-Jensen, J., Gutin, G. Z. (2009). *Digraphs: theory, algorithms and applications*. London: Springer.

Blau, H., Immerman, N. & Jensen, D. (2002). A Visual Language for Querying and Updating Graphs. Technical Report 2002-037, University of Massachusetts, Amherst.

Burch, T., Rapp, A. (2007). Das Wörterbuch-Netz: Verfahren – Methoden – Perspektiven. In D. Burckhardt, R. Hohls & C. Prinz (eds.) *Geschichte im Netz: Praxis, Chancen, Visionen. Beiträge der Tagung .hist 2006* (= Historisches Forum 10/I). Berlin, pp. 607-627. Accessed at: http://edoc.hu-berlin.de/histfor/10_I/PHP/Woerterbuecher_2007-10-I.php#007001

de Vincenz, A., Hentschel, G. (2010). *Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache. Von den Anfängen des polnischen Schrifttums bis in die Mitte des 20. Jahrhunderts*. (= Studia slavica Oldenburgensia, vol. 20). Oldenburg: BIS-Verlag. Online version: http://www.bis.uni-oldenburg.de/bis-verlag/wdlp.

Engelberg, S. (2010). An inverted loanword dictionary of German loanwords in the languages of the South Pacific. In A. Dykstra, T. Schoonheim (eds.) *Proceedings of the XIV EURALEX International Congress (Leeuwarden, 6-10 July 2010)*. Ljouwert (Leeuwarden): Fryske Akademy, pp. 639-647.

Karaulov, J. N. (1979). Obratnyj slovar' zaimstvovanij kak sposob isučenija lingvoėkologii. *Izvestija Akademii Nauk SSSR. Serija Literatury i Jazyka*, 38/6, pp. 552-562.

Kasneci, G., Suchanek, F.M., Ifrim, G., Ramanath, M. & Weikum, G. (2008). NAGA: Searching and Ranking Knowledge. In *IEEE 24th International Conference on Data Engineering, 2008*. ICDE 2008, pp. 953–962.

Menzel, Th., Hentschel, G. (2005). *Wörterbuch der deutschen Lehnwörter im Teschener Dialekt des Polnischen. 2nd, enlarged and revised ed. online*: Accessed at http://www.bkge.de/14451.html.

Meyer, P., Engelberg, S. (2011). Ein umgekehrtes Lehnwörterbuch als Internetportal und elektronische Ressource: Lexikographische und technische Grundlagen. In H. Hedeland, Th. Schmidt & K. Wörner (eds.) *Multilingual Resources and Multilingual Applications*. Hamburg: Universität Hamburg, Sonderforschungsbereich 538 Mehrspra- chigkeit, pp. 169-174.

Meyer, P. (2013). Ein Internetportal für deutsche Lehnwörter in slavischen Sprachen. Zugriffsstrukturen und Datenrepräsentation. In S. Kempgen, N. Franz, M. Jakiša & M. Wingender (eds.): *Deutsche Beiträge zum 15. Internationalen Slavistenkongress, Minsk 2013*. München: Otto Sagner, pp. 233-242.

Robinson, I., Webber, J. & Eifrem, E. (in press). *Graph Databases*. Beijing etc.: O'Reilly Media.

Sanchez, E. & Yamanoi, T. (2006). Fuzzy ontologies for the semantic web. In H.L. Larsen, G. Pasi, D.O. Arroyo, T. Andreasen & H. Christiansen (eds.) *Proceedings of the 7th International Conference on Flexible Query Answering Systems, Milan, Italy*. London etc.: Springer, pp. 691-699.

Striedter-Temps, H. (1963). *Deutsche Lehnwörter im Slovenischen*. Wiesbaden: Harrassowitz.

van der Sijs, N. (2010). *Nederlandse woorden wereldwijd*. Den Haag: SDU Uitgever.

Wood, P. T. (2012). Query Languages for Graph Databases. *SIGMOD RECORD*, 41(1) (March), pp. 50-60.