Features for the Dictionary Portal

## Introduction

One of the task of WG1 is the design and implementation of a dictionary portal. Moreover, the WG should investigate possible features that can be added in future developments. In the following we will present a set of possible features for such a portal. However, only a very limited set can be realized in the course of the present project.

The purpose of this document is to simply list possible features the portal may have. Further research is required to the usefulness and practicality of the features. Such research will have to define user groups and needs to produce a number of enticing use cases. Furthermore, it will need to outline a technical solution and an estimation of the effort to produce and maintain the solution.

The features below are grouped according to the main purpose the serve. Furthermore, the groups are ordered from easy to realize to very complicated.

## Information

The features in this section pertain to provide users with general information about dictionaries and the process of dictionary making. The content will mainly consist of links and static text. These features require active maintenance. One approach would be to create a wiki-like environment where with access limited to members of the community.

* Provide information about on-going dictionary project. Since most projects have their own project website, the portal could provide links to those project websites with a short description and perhaps highlight actual additions to the websites.
* Provide links to events in the field of e-lexicography.
* Provide links to publications in that field.
* Provide blogs and discussion lists on relevant subjects.

## Discovery

The most important function of a portal is to provide pointers to other information. In our case to online dictionaries. It is a matter of choice which kinds of dictionaries to include. Below are some features which enable a user to find dictionaries.

* Simple link farm, with links to available online dictionaries. Links may be grouped according to one or a few properties.
* Search facility to find dictionaries according to properties. Search is only useful if there are sufficient (e.g. > 100) dictionaries to choose from.
* A contact form, allowing others to add entries to the link - or reference database. Also, this option is only useful if we intent to include a large number of dictionaries.
* And, optionally, a list of references to available printed dictionaries.

## Search; Human Interface

In the previous section we mentioned a facility to search for dictionaries. In the present section we list options for a user to search for dictionary content.

* Search in single dictionary. The user indicates a dictionary and specifies a number of search options. The result of the search is presented directly to the user. So, the user stays within the realm of the portal and is not redirected to the online dictionary to be presented with the result. The presentation within the portal may be different from the presentation in the online dictionary and some usability might be lost. Think of content appearing when the mouse hovers above a certain section.
* Federated search in multiple dictionaries. Instead of one dictionary, the user selects multiple dictionaries or a certain language which automatically selects all corresponding dictionaries. The search will return a list of results from all dictionaries involved. Also here, the usability and content may be different from that available from the original online dictionaries. Moreover, search parameters may need to be restricted to those that all dictionaries have in common.

Below a few more possible features that allow a user to find a corresponding lemma’s (translations) in other languages. These features correspond to the activities of WG4

* Search a dictionary in the source language. The lemma will contain a direction to a lemma of a dictionary in the target language, also at the lemma level. This feature assumes that the monolingual dictionaries are linked at the level of the lemma.
* Same as above at level of separate meanings. This features requires the monolingual dictionaries to be linked at the level of meanings.

## Search; Machine Interface

Besides providing an interface for human interaction, it is possible to facilitate automatic access. An example might be useful to explain this. Consider a library which displays historic documents at its website. These documents contain unfamiliar word and the library wants to provide the user with an explanation of these words. The library might construct the website such that when the user hovers the mouse over a word, the website sends a message to the dictionary portal (over another channel and invisible to the user) requesting information about the word. The request might be extended with parameters which specify language, period, etc. The dictionary portal responds with data from a dictionary which fits best with the parameters provided. The website of the library uses the received data to display a popup with the dictionary content next to the word in the document where the user held his mouse.

The features of this type overlap with the activities in WG3. Below some examples. The list is far from exhaustive.

* Providing access to lemma's, meanings. Like in the example above, dictionary content is returned according to search parameters provided.
* Providing frequency information for word forms. The website sends the word form and other information like language, document type, period to the portal service and retrieves relative frequency information about the word or lemma. This might be useful to e.g. select the words in a document that require elucidation.
* Providing disambiguation tools as a web service. The website sends the word form and some relevant context of the word. The portal service returns dictionary content corresponding to the most likely meaning of the word.

## Cross-dictionary Content Creation

These features pertain to the process of linking information in dictionaries. These features, of course, transgress the domain of a mere portal and require advanced web applications. The processes can be partially automated, but will require a large effort from humans. These type of features overlap with the subject of WG’s 3 and 4.

* Linking lemma's monolingual. Lemma’s from several dictionaries of the same language are cross-linked.
* Linking lemma's multilingual. Lemma’s from dictionaries of different languages are cross-linked.
* Linking meanings monolingual. Dictionaries of the same language are linked at the level of meanings.
* Linking meanings multilingual. Dictionaries from different languages are linked at the level of meanings
* Cross-linking between descriptive dictionaries and etymological dictionaries.
* Linking etymological dictionaries cross-language. Many words in different European languages share some part of their etymology. Merging these etymological descriptions will result in a pan-European interactive genealogy of words and meanings (cf. etymological wordnet). Such an interconnected data set will provide many opportunities to e.g. visualize the historic development of languages and their geographical shift according to political and economical influences.

## Dictionary Content Creation

Features pertaining to creating and enhancing dictionary content are beyond the normal features of a portal, but can be seen as an extension of the service of a portal and depend largely on the results from WG3.

* Corpus analysis tools. The portal may provide a facility to store large corpora of all European languages and make them accessible by means of advanced tools for querying and analysis.
* Neologisms. A central service to analyze and compare corpora in order to detect possible neologisms.