

# Things to think about when building a dictionary website

Michal Boleslav Měchura

Dublin City University and Foras na Gaeilge, Dublin, Ireland

E-mail: valselob@gmail.com

## 1. Introduction

Websites come in different genres: blogs, social networking sites, corporate brochure-like sites and so on. Dictionary websites constitute their own genre too. What distinguishes dictionary websites from others is the relative **structural complexity** of dictionary entries in comparison to other text types such as news articles, and the prominence of **search functionality** on dictionary websites.

This paper will review the factors that go into planning and building a successful dictionary website. Some of the points presented here are specific to dictionary websites while others are applicable to all websites in general. Together, they should give the reader a useful checklist of things to think about when planning a dictionary website, or a suggested list of things to ask for when hiring someone to build one.

## 2. Controlling access to the website

### 2.1. Should users be required to log in?

The consensus now seems to be that dictionary websites should be freely accessible, with no need to register or log in. Every click or interaction that delays the user from accessing the content they want will inevitably turn some percentage of users away. If your motivation for requiring users to log in is merely to track user behaviour, then it is probably better to rely only on log file analysis or on third-party services such as Google Analytics.

### 2.2. How will the website generate revenue?

Websites can generate revenue by carrying advertising or by selling subscription. A mixed ‘freemium’ model is also possible. Finally, a website can generate revenue indirectly as ‘advertisement’ for another product, such as a printed dictionary or a mobile app. Either way, it is a fact of life that dictionary websites generate less revenue than printed dictionaries used to. Luckily, many dictionary projects are funded from public money where revenue generation is irrelevant (and replaced by a motivation to attract traffic in order to justify investment into the dictionary as a ‘public service’).

## 3. Searching and finding

### 3.1. Which elements in an entry can match a search query?

A ‘search’ is basically a process of taking the text the user has typed (the ‘query’) and comparing it to every item on a list of strings of text to see which ones ‘match’ the query. The list of text strings that

can potentially match a query should obviously contain the dictionary's headwords, but perhaps also **multi-word expressions** inside entries, **example sentences** and others. In a bilingual dictionary you can decide to look for matches on the target side as well, bringing **reverse search** into your website.

### **3.2. How much of an entry do we display as a result of a match?**

When a search query matches a headword, the user would normally expect to be shown the entire entry. But when the query matches something inside an entry, you may want to show only an **extract**, for example only the relevant example sentence, or only the sense containing the translation the user has searched for. Such extracts should be accompanied by links to the entire entry.

### **3.3. How do we group, rank and order the search results?**

Not all matches are equal. Matches on a headword outrank matches on an example sentence, matches on an entire headword outrank those that only match part of a (multi-word) headword, and so on. In a complex search algorithm, the results must be ranked and grouped by **match quality** (or closeness) and by type and/or language of the matched element.

### **3.4. Do we ask the user which language he or she is searching in?**

On websites that offer a language selector next to the search box, users fails to select the correct language surprisingly often. It is probably smarter to do away with the language selector and to always search all languages simultaneously. This means that the results may need to be grouped by language when the search yields matches in more than one.

### **3.5. What is a 'match', anyway?**

A 'match' is a correspondence between two strings, the user's query on the one hand and a string of text in our dictionary on the other. Not all matches are exact. For example, when a user searches for a word, it is a good idea to return all elements that contain any inflected form of this word. Essentially, all dictionary searches are **fulltext searches** (with some tweaks: for example, function words should normally not be ignored).

### **3.6. Should we suggest similar words?**

The user's experience is enhanced greatly if the search algorithm can detect spelling errors and suggest corrections, and if it can suggest words based on semantic similarity. Understandably, the suggestions should go to headwords that actually exist in the dictionary!

### **3.7. Should we have an 'advanced search' feature?**

The trend in online search interfaces (not only in dictionaries) is to provide just a single text box, and let the search algorithm try to figure out what the user wanted. The opposite option is an 'advanced search' feature where a 'power user' can specify all options explicitly. Such interfaces are complicated

to develop and tend to be rarely used. Whether the investment is worth it depends mainly on who the target audience is.

### **3.8. How do we serve users who aren't necessarily looking for anything in particular?**

Sometimes users visit a dictionary website with no particular information need; they are here merely to explore. Therefore, in addition to search, it is advisable to develop functionality that enable 'browsing', such as an alphabetical headword index or a 'word of the day' feature.

## **4. Design and graphics**

### **4.1. How do we 'style' the entries?**

Dictionary entries are structurally much more complicated than any other text type web designers are used to dealing with. Consequently, it can be challenging to devise a formatting 'style sheet' that does justice to all the structural eventualities that can occur: an entry having just one sense or many, a sense having no example sentences or just one or many, usage labels appearing at different levels in the entry's tree structure, and so on.

### **4.2. How do we cater for mobile devices?**

Modern websites need a **responsive design**: they automatically adapt to the dimensions of the screen they are on, from the multi-column layout of large desktop screens to the narrow screens of mobile phones.

### **4.3. How do we cater for users with disabilities?**

Modern websites are coded in such a way that they can be read aloud by **screenreading software** for people with visual impairment. This is particularly challenging for dictionaries due to the complex structure of dictionary entries: they still need to 'make sense' when read aloud.

## **5. Sharing and linking**

### **5.1. What structure will our URLs have?**

It is desirable for every page on a dictionary website, including every search result and/or every entry, to have a unique URL, and for the URLs to be short and to 'make sense' to humans. Such URLs are easy to copy and paste into e-mails and social networking messages, encouraging sharing.

### **5.2. How do we play on social networks?**

It has become customary for websites to carry social networking 'widgets' which allow users to share a page quickly on a social network such as Facebook. It may also be advisable to set up a profile for your website on social networks, invite users to 'like' or 'follow' them, and to feed regular updates to

them. A modern website with a modern institution behind it needs a social-networking strategy.

### **5.3. What machine-readable metadata should we include?**

A well-designed website contains machine-readable metadata which is invisible to humans but used by third-party search engines such as Google and Bing to index your website and to bring users to you. Machine-readable metadata is also used by social networking applications such as Facebook and Twitter to generate ‘previews’ when sharing a link.

### **5.4. What plugins should we offer?**

A good way to promote your website is to provide ‘plugins’ which can be installed elsewhere and which will bring users back to you, such as an OpenSearch plugin that allows users to search your dictionary from the search box in their own browser, or a reusable search box to be embedded in other websites.

## **6. Summary**

Websites are now the output format of choice for dictionary projects, outranking both printed books and other computer media such as CD-ROMs and mobile apps. For this reason it is advisable to devote good care to the planning and building of a dictionary website.

On the one hand, dictionary websites constitute their own genre with their own specialized requirements, such as the emphasis they place on search functionality. On the other hand, a dictionary website is a website like any other and needs to comply with modern-day expectations such as responsive design. This paper is a checklist for navigating through these.