Designing and Developing the Slovenian Dictionary Portal Fran

Andrej Perdih[1], Janoš Ježovnik[1]

[1] Fran Ramovš Institute of the Slovenian Language, Scientific Research Centre of the Slovenian Academy of Sciences and Arts, Ljubljana, Slovenia
E-mail: andrej.perdih@zrc-sazu.si, janos.jezovnik@zrc.sazu.si

Keywords: dictionary portal; dictionary; lexicography; user contributions

1. Introduction

At the Fran Ramovš Institute of the Slovenian Language, more than 30 dictionaries have been published in the last 15 years, and thus the need has arisen to offer a larger amount of language resources integrated on an internet portal and at the same time to make the concurrent use of several resources as informative and as straight-forward as possible. The main goal was to establish an internet portal with a simple navigation system and with a clear and intuitive graphical representation of the data. The minimalistic graphic approach makes it possible, by reducing the amount of data directly presented, to publish newer dictionaries (primarily designed for use in electronic form) as well as older ones (primarily published in print). The Fran portal is used by various users with various demands. Therefore, different help functions are offered to general users, while advanced search options, additional functionality and content are offered to advanced users.

The Fran portal (full title: Fran: The Dictionaries of the Fran Ramovš Institute of Slovenian Language of the Research Centre of the Slovenian Academy of Sciences and Arts, URL: www.fran.si)[1] was published on October 14th 2014. At the time of writing it contains 28 integrated dictionaries, 1 linguistic atlas, language counselling and several other language resources. All but one dictionary[2] are freely available.

---

[1] Original title: *Fran: Slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU*. The portal is named after the Slovenian linguist Fran Ramovš.

[2] On the same date as Fran came into existence, the second, updated and partly renewed edition of the general monolingual dictionary *Slovar slovenskega knjižnega jezika* was published (Perdih & Snoj, 2015). This is the only dictionary on the portal that requires a registration code. Nevertheless, members of Slovene National and University Library have free access to this dictionary.

Figure 1: Home page with search field

2. Content

On the Fran portal, information on the Slovenian language can be obtained from the view of contemporary standard language as well as from the historical view from the 16th century until the present or across all Slovenian dialects. The portal is divided into 7 sections:

- **general** dictionaries (7 dictionaries),
- **etymological** dictionaries (1 dictionary),
- **historical** dictionaries (5 dictionaries),
- **terminological** dictionaries (11 dictionaries),
- **dialectal** dictionaries (4 dictionaries + 1 linguistic atlas),
- language **counselling**,
- and **other collections** (including a collection of exhaustive descriptions and (unless copyrighted) full digitalized texts of all the grammars and orthographies of the Slovene language from 1584 up to the present day).

3. Functionality and design

Despite the simplicity of its user interface Fran offers a wide array of options in terms of searching the content while also providing links to other online language resources such as various corpora, terminological databases etc., and allowing for user contributions. An overview is presented in Fig. 2. and described in further detail below.

Figure 2: Search results and navigation panel

3.1 Search scope

By default, the search is conducted in all 28 dictionaries, the linguistic atlas and the 2 counselling websites (see Figure 2.). The results are shown on the left-hand side, each with a grey label indicating the short name of the dictionary wherein the result was found. The navigation panel (to the right) provides information on the number of results in all the dictionaries, the atlas and counselling websites as well as links to navigate into any one of them. Additionally, direct links to search results in several Slovene corpora are available as well as a link to post questions on counselling websites.

3.2 Navigation panel

There are 564,707 dictionary entries altogether on the Fran portal. In the navigation panel, clickable numbers in blue labels represent the number of entries with search results in headwords and subheadwords[3] in the selected dictionary. Clickable green labels represent the number of entries that contain search results outside of headwords and subheadwords.

3.3 Search types

The portal offers two search types: simple and advanced. Using simple search, typing of the search condition is facilitated by the incremental search functionality, offering the user up to five (sub)headwords (cf. Fig. 3.). If no result is found, up to five similar results are offered (first researches have shown that the technology used brings fairly good results, cf. Kosem, 2014: 179–181). Using advanced search, it is possible to select a dictionary or a group of dictionaries in which the search should be performed. Additionally, it is also possible to define structure elements where the search string is to be found. Depending on the dictionary structure it is possible to specify searching for headword, definition, label, terminology, phraseology etc. (cf. Figs. 4. & 5.). Registered users can choose to save their last advanced search parameters for possible future use.
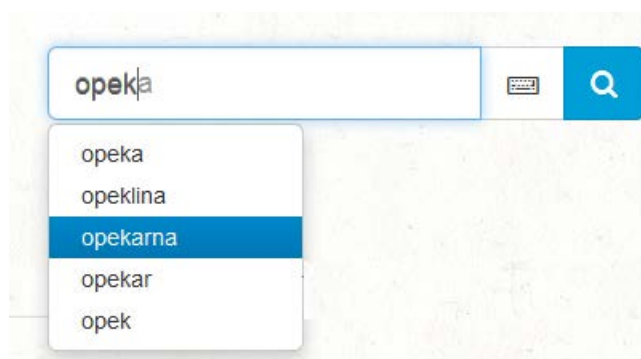

Figure 3: Incremental search

---

[3] By default, searches are conducted both in headwords and subheadwords, solving search issues in print dictionaries.

Figure 4: Advanced search in several elements (right) of the selected dictionaries (left).

Figure 5: Advanced search: highlighted results

3.4 Wildcards and regular expressions

In both simple and advanced search, wildcards can be used (* matches any string, ? matches any character), as well as double quotes (" " for phrase searching). In advanced search, regular expressions can also be used (inside slashes / /).

The default search setting uses normalization of accentuated diacritical characters (e.g. searching for *e* finds *éêě*...), but when searching for a specific diacritical character normalization is not used (e.g. *ê* does not search for *eéě*...). Since the portal is aimed at very different types of users, i.e. general and professional ones, including linguists, two virtual keyboards for inserting special characters have been created.

3.5 Keyboards

When using simple search, a keyboard containing 42 most common special characters is shown, including Slovene alphabet characters (ČŠŽ), while the keyboard for advanced search contains 244 special characters.

3.6 Search result ordering

By default, search results from the most frequently consulted dictionaries are shown first, i.e. the results from the first (or second, if available) edition of the monolingual dictionary (*Slovar slovenskega knjižnega jezika*) are shown on top, followed by results from the dictionary of neologisms (*Slovar novejšega besedja slovenskega jezika*) and the orthographic dictionary (*Slovenski pravopis* 2001).

The search results can be sorted in three ways. As explained above, results found in (sub)headwords of the most frequently consulted dictionaries are ranked highest by default. Afterwards, partial results are shown, e.g. when the result is found as a part of a multi-word unit. Search results can also be sorted according to the dictionary, i.e. in such a way that all results from one dictionary are presented together. If the user selects searching in one dictionary only, alphabetic entry sorting is also available, showing a clickable headword list (in alphabetic order) in a panel on the left-hand side while the entries are presented next to it (cf. Fig. 6.).

Registered users can change the default sorting to display search results by headword and set the number of search results shown to up to 100 (default is 20).

Figure 6: Alphabetic ordering in SSKJ[2] dictionary

3.7 Tooltips

In order to help users with understanding dictionary structures, abbreviations and

symbols that require consulting the front matter in printed editions, hints are displayed in the form of tooltips when hovering the text or symbols with the cursor or tapping the screen. For example, *Terminološko gnezdo* "Terminology section" is shown for the ♦ symbol (cf. Fig. 7.) and *elektrotehnika* "electrotechnics" is shown for *elektr.* abbreviation (cf. Fig. 8.).[4]
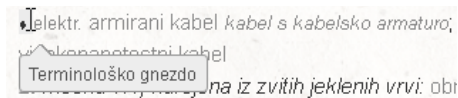
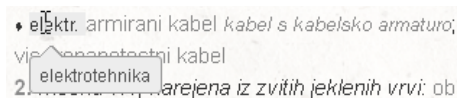**Figure 7:** Tooltip on symbols

Figure 8: Tooltip on abbreviations

3.8   User contributions

Users can contribute to the portal by sending suggestions and emendations regarding dictionary entries, suggesting new Slovene adequate words for recently loaned words, upvoting or downvoting other users' suggestions and sharing them on Twitter, and suggesting words not yet included in the portal, all via web form (cf. Fig. 9.).

3.9   Other content

For advanced users, especially for linguists, dictionary front and back matters are available (prefaces, introductions, lists of abbreviations etc.).

---

[4] This mechanism depends on structural elements and their content. In the process of retro-digitization, some dictionaries have not been adequately structured yet, therefore some mistakes are possible. These issues will be solved in the future.

Figure 9: User contributions

4. Future improvements

At least once a year the portal is upgraded in terms of new content, added functionality, and design. As of 2016, an English interface is being prepared for international users, opening the possibility of more languages to be added in the future. Several dictionaries are being prepared for inclusion in the portal, among others the first volume of the *New Dictionary of Standard Slovene* (cf. Gliha Komac et al., 2015), second volume of the *Slovene Linguistic Atlas*, and several dialectal dictionaries.

5. References

Abel, A. & Meyer, C. M. (2013). The dynamics outside the paper: user contributions to online dictionaries. In I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets & M. Tuulik (eds.) *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, pp. 179–194. Available at: http://eki.ee/elex2013/proceedings/eLex2013_13_Abel+Meyer.pdf.

Ahačič, K., Ledinek, N. & Perdih, A. (2015). Fran: The Next Generation Slovenian Dictionary Portal. In Gajdošová, K. & Žáková, A. (eds.) *Natural Language Processing, Corpus Linguistics, Lexicography. Eight International Conference Bratislava, Slovakia, 21–22 November 2015. Proceedings.* Bratislava: RAM-Verlag, pp. 9–16.

Ahačič, K., Ledinek, N. & Perdih, A. (2015). Portal Fran – nastanek in trenutno stanje.

In M. Smolej (ed.) *Slovnica in slovar – aktualni jezikovni opis (1. del)/Grammar and dictionary – Current language description (Part 1). Simpozij/Symposium Obdobja 34.* Ljubljana: Znanstvena založba Filozofske fakultete, pp. 57–66. Available at: http://www.centerslo.net/files/file/simpozij/simp34/zbornik%201/Ahacic-Led-Per.pdf.

Fran, slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU, 2014–, version 3.0. Accessed at: www.fran.si. (30 December 2015)

Gliha Komac, N., Jakop, N., Ježovnik, J., Klemenčič, S., Krvina, D., Ledinek, N., Mirtič, T., Perdih, A., Petric, Š., Snoj, M. & Žele, A. (2015). *Koncept novega razlagalnega slovarja slovenskega jezika.* Ljubljana: Inštitut Frana Ramovša za slovenski jezik ZRC SAZU. Available at: http://www.fran.si/179/novi-slovar-slovenskega-knjiznega-jezika/datoteke/Potrjeni_koncept_NoviSSKJ.pdf.

Kosem, I. (2014). Fran: pameten in intuitiven? *Slovenščina 2.0*, 2 (2), pp. 161–193. Available at: http://www.trojina.org/slovenscina2.0/arhiv/2014/2/Slo2.0_2014_2_09.pdf.

Lew, R. & Mitton, R. (2011). Not the Word I Wanted? How Online English Learners' Dictionaries Deal with Misspelled Words. In I. Kosem & K. Kosem (eds.) *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011, Bled, 10-12 November 2011.* Ljubljana: Trojina, Institute for Applied Slovene Studies. Available at: http://elex2011.trojina.si/Vsebine/proceedings/eLex2011-21.pdf.

Lew, R. (2012). How can we make electronic dictionaries more effective? In S. Granger & M. Paquot (eds.) *Electronic Lexicography.* Oxford: Oxford University Press, pp. 343–363.

Müller-Spitzer, C., Koplenig, A. & Töpel, A. (2012). Online dictionary use: Key findings from an empirical research project. In S. Granger & M. Paquot (eds.) *Electronic Lexicography.* Oxford: Oxford University Press, pp. 425–458.

Perdih, A. & Snoj, M. (2015). SSKJ[2]. *Slavia Centralis*, 8 (1), pp. 5–15.