Practice of Smart LSP Lexicography: The Case of a New Botanical Dictionary with Latvian as a Basic Language

Silga Sviķe, Karina Šķirmante

Ventspils University of Applied Sciences, Inženieru Street 101, Ventspils, LV-3601, Latvia
E-mail: silga.svike@gmail.com, karina.krinkele@gmail.com

Abstract

The article provides an insight into the project "A New Botanical Dictionary: Terms in Latvian, Latin, English, Russian, and German" implemented in the second half of 2017 and in 2018 within the Ventspils University of Applied Sciences (VUAS) internal call for proposals "Development of Scientific Activity at the VUAS". The VUAS Faculty of Translation Studies in collaboration with the Faculty of Information Technologies in their scientific and research work along with other Latvian universities aim to occupy a niche in the branch of applied linguistics, therefore the research is related to this discipline and offers solutions in practical lexicography.
The study describes a new botanical dictionary (NBD) – a mobile application prototype – with Latvian as a basic language. An insight into the macrostructure of the dictionary and the structure of entries is given. The research deals with questions concerning IT solutions in general (simple) and semantic search in particular. It also introduces a general search – a morphological approach developed by the authors of the research specifically for the Latvian language; this approach is used to search for Latvian botanical terms in both singular and plural forms. The extracted and linked data methodology developed by the authors is described in detail, as well as the NBD technical solutions and architecture, technologies used, database model, and additional features.

Keywords: LSP lexicography; botanical dictionary; mobile application

1. The Need for a New Botanical Dictionary

One of the indicators of a well-structured and successful process of developing and coordinating field-specific terms is using qualitative, topical and useful terminology resources related to a particular field (TTC, 2007: 38). Although approximately 30% of Latvian lexicography consists of dictionaries of a terminological nature (Helviga & Peina, 2016: 127), the translators' need for them is still not satisfied (Balode, 2012: 40; Sviķe, 2018: 228-241); besides, the importance of specialized dictionaries for society in general should be noted. (Fuertes-Olivera & Tarp, 2014: 2) The need for compiling a new botanical dictionary is proved by the fact that more than half a century has passed since in 1950 the first issue of Galenieks's *Botanical Dictionary* (Latvian: *Botaniskā vārdnīca*) was published, thus it is necessary to compile a new dictionary of botanical terms with Latvian as a basic language. Within this study, the term NBD means a terminological work in the form of a multilingual translation dictionary (mobile

application) that can be used when translating from and into different languages. As the Latvian part of the dictionary has more specific implementations and offers a wider range of solutions (e.g. search options: see Section 4.2, definitions retrieved from *www.tezaurs.lv*), the Latvian language is defined as the basic language of the dictionary, while the other languages (English, Russian, German) as contrasted languages.

Plant names are an important part of botanical terms. However, some of the currently available electronic dictionaries and databases (e. g. the database of terms compiled and approved by the Terminology Commission of Latvian Academy of Sciences – *www.termini.lza.lv*) do not include the names of several important genera and species, like translations of the Latvian *ārstniecības izops* (*hyssop* in English) and *zilā vizbulīte* (*liverleaf* in English) into German and English (see *termini.lza.lv*). Translations of the names of many crops and economically important plants into English, Russian, and German are also not found in the electronic encyclopaedia *Latvian Nature* (see *Latvijasdaba*.lv), which mostly includes the names of Latvian species of flora. A conceptually new botanical term dictionary is needed not only for professional translators, but also for media professionals, science students, and natural science teachers or students.

Before compiling the dictionary, a survey and a statistical processing of survey data were conducted to identify potential users of a future product. The conclusions drawn from the analysis of survey data (see Sviķe, 2018) were taken into account when developing the prototype of a mobile application. One of the respondents' preferences was an electronic botanical dictionary with an offline option, so a mobile application with the dataset included in a local application database was considered to be the right solution. Initially, the intention was to develop an Android version of the dictionary, as, for instance, in the period from June 2018 – July 2019 in the Latvian market around 65–75% of smartphones were Android devices, and only 24–32% were iOS ones (see *Statista.com*). The situation could be similar elsewhere in the world. However, during the upcoming stages of improving the mobile app, the production of an iOS version will also be considered by using the Cross-Platform Mobile Development App "Ionic" or other possibilities. New approaches to the structure, as well as the functionality of the NBD, are described in the following sections of the study. The aim of this article is to show practical lexicographic solutions for the development of a new botanical dictionary (mobile application), specifying the problems encountered when using the Latvian language as the main one, and offering innovative solutions in developing search functions.

Compilation of the dictionary was conducted within two stages and financed by the VUAS. The first stage was implemented during the project "New Botanical Dictionary: Lexicographic Concept and Working Model" (project duration – five months), when the term search functionality, plant and flower structure visualization and linkage with terms, representation of pictures and literature lists were developed. The second stage was implemented during the project "New Botanical Dictionary: Supplementation of

Lexicographic Material and Modernization of the Mobile Application Prototype" (project duration – six months), by introducing possibilities to change the interface language from Latvian into English and vice versa; adding images of seed and root structure, of simple and compound leaves; creating interactivity between visual and search parts; making a list of publications; supplementing entries with photos; and introducing semantic search.

2. The Macrostructure of the NBD

This section provides only an insight into the macrostructure of a dictionary to explain the overall structure of the app.[1] A brief overview of the macro- and microstructure of the dictionary in this article is also required to describe programming solutions in the following paragraphs.

The macrostructure of the dictionary (see Figure 1) includes the main body – the lexicographic database level of the mobile application (at the presentation level the user sees the term search view when starting the mobile app) – and several sections (mobile application menus):

1. About the NBD – a view providing the description of the project, the authors' names and useful information about the mobile application in Latvian, English, Russian, and German.

2. Entry structure – describes all the components of the entries and their functions, as well as features used for increasing functionality: simple and semantic search.

3. The Designations Used – a view showing the table of designations and abbreviations used in the app, as well as explanations and translations into all contrasted languages.

4. Pictures – containing the following parts: Plant, Flower, Root, Seed and Leaf Structure – a view showing a picture of a plant or its part, where one can interactively translate the term of the chosen part of a plant into any of the contrasted languages.

5. Sources Used – a view showing all the sources used when developing the botanical dictionary prototype.

6. Publications – a view containing links to articles on botanical terminology that are potentially useful for users of the dictionary.

[1] Reported at the international scientific conference "Meaning in Translation Illusion of Precision" (*Semantiskais aspekts tulkošanā: precizitātes ilūzija*) organized by Riga Technical University in Riga, May 16–19, 2018.

7. Semantic search – searching only according to the scientific (Latin) name of the plant included in the dictionary (a detailed description is given in Section 4.2.2).

8. Selection of app language – changes the language of the user interface (English or Latvian).

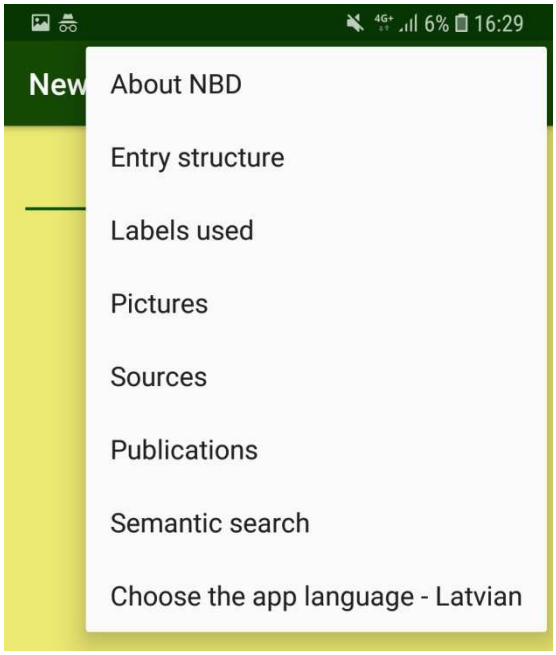

Figure 1: Menu of the mobile application in English

Initially, the macrostructure of the dictionary did not include the section of publications and the selection of app language; those were added during the second stage of compiling the dictionary, when it was supplemented and upgraded. However, this does not exclude the possibility of adding other useful sections to the macrostructure (such as external plant image databases or plant and plant structures schemes) during subsequent stages of upgrading the dictionary prototype and supplementing the language material.

3. The Microstructure of the NBD

Sylviane Granger (2012: 2) lists the six most significant innovations offered by the electronic medium: (1) corpus integration; (2) larger and better data; (3) efficiency of access; (4) customization; (5) hybridization; and (6) user input. The NBD compilers have attempted to include at least five of these, as follows: (1) linking a consolidated corpus of dictionaries with the corpus integration; (2) additional data from other free-access sources according to users' preferences: *www.tezaurs.lv* and *www.wikipedia.org*; (3) internal hyperlinks (from the main view of the mobile app to included pictures) and external hyperlinks (to external sources); (4) lexicographic surveys as a form of

customization; (5) some aspects of hybridization, such as linking the encyclopaedic and linguistic approaches (in the further processing of automatically extracted definitions). These aspects are more clearly evident in the structure of entries elaborated for the dictionary.

Terms included in the dictionary form a so-called *block* structure on the home view of the mobile app. On the home view, the dictionary shows the equivalents of a word searched in all the contrasted languages, thus creating a block. The entries consist of the following structural elements[2] (see Figure 2):

1. A word or words searched in the input field (box).

2. The functional search button is on the right of the input field.

3. Below the functional search button there is a photo icon which, when being touched, shows on the smartphone screen a photo of the plant that was saved in the resource directory created during the development of the mobile app (only for entries with an image saved in the app database).

4. Below the input field, there is a block of contrasted languages (arranged under each other) and term equivalents. In the NBD, after terms in Latin (put in italics), translations into other languages are arranged in alphabetical order. For German, Latvian, and Russian equivalents grammatical references are also given: gender (female – *f* (*femininum*), male – *m* (*maskulinum*), neuter – *n* (*neutrum*)), singular – *sg* (*singularis*) and plural – *pl* (*pluralis*).

   The explanatory part of an entry appears below the language block.

5. A hyperlink to the entry description in the free-access multilingual encyclopaedia *Wikipedia*. This function was mentioned by potential users of the dictionary in the lexicographic survey (see Sviķe, 2018). However, not all plant names included in the dictionary could be provided with a hyperlink to the plant photos, so it was decided to include photos in the dictionary itself; the user can view a photograph of the plant by using a pictogram.

6. A hyperlink to *www.tezaurs.lv* and definition of an entry retrieved from *www.tezaurs.lv* – the website consolidating different Latvian dictionaries – with the help of specially developed software. The abbreviations and markings used are shown below the explanatory section.

7. Glossary of Latin abbreviations used in the dictionary.

[2] The microstructure of NBD was discussed at the conference "The Word: Aspects of Research" (*Vārds un tā pētīšanas aspekti*) organized by Liepaja University in Liepaja, November 29–30, 2018.

8. Indications and markings of taxonomic levels: the word or words searched in the entry are coloured in the related colours.

9. Explanation of "T*" marking.

10. The INFO section (not shown in Figure 2) – a commentaries part made by the compilers of the dictionary for a relevant entry. In the future, it will be possible to keep in the INFO section not only the corrected or updated definitions automatically retrieved from *www.tezaurs.lv*, but also other comments about the entry. To implement this idea, during the subsequent phases of the dictionary compilation project it is necessary to analyse all the definitions automatically retrieved and to develop new definitions for those cases when an automatic retrieval is not accurate or is incorrect (a detailed description of this is given in Section 4.2.4).

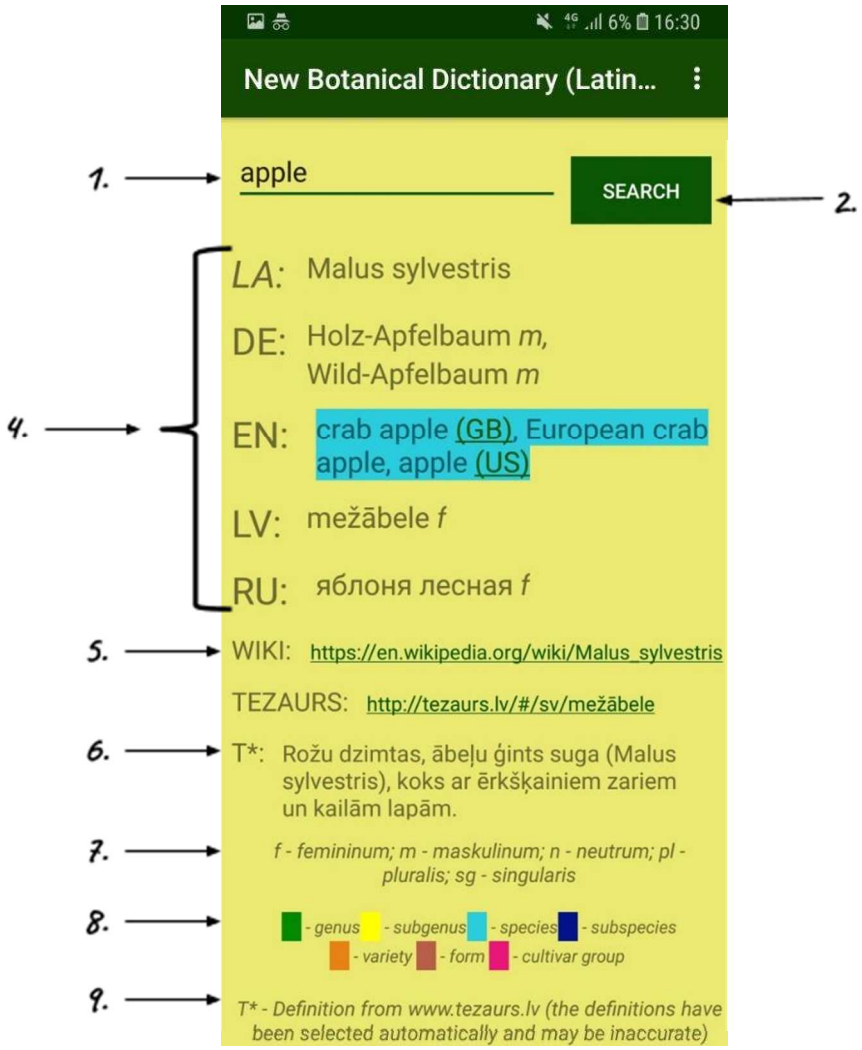

Figure 2. Term searching view

Figure 2 also shows accompanying explanatory notes (especially useful for translators) for English plant names – GB and US, which indicate the use of a plant name in Great Britain or the USA. When expanding the variety of entries, it has been found that in the same language one plant is named and called differently in various countries, e.g. *shadbush* in Germany is more often referred to as *Felsenbirne*, but in Austria as *Edelweißstrauch*. That is why the following markings were introduced: German equivalents used in Germany have a country code (DE), in Austria – (AT), in Switzerland – (CH). For English equivalents the codes US (United States) and GB (Great Britain) are used. The two-letter country codes were selected according to ISO 3166-1. Of course, such country codes are just one of the solutions offered in the dictionary (app prototype) that would make it easier for translators to choose the best equivalent. However, during the next stages of improving the dictionary it should be decided how to distinguish national varieties.

It is intended to supplement the NBD with the sixth innovation mentioned by S. Granger – user input – by adding extension and reduction signs "+/-", for instance, to the definition and INFO sections. Thus, the vocabulary user will be able to open or hide the information section of an entry. To create the "show and hide" functionality of large or small texts, the expandable TextView component of Android may be used. Possible further solutions – entering data by users, and thus personalizing the app, e.g. by adding comments in one of the sections if the user needs it. However, further improvements of the mobile app prototype require additional funding, deeper research, and extra programming work to develop a new system module. New solutions will be described in future research done by the authors.

4. The Functionality of the NBD Technical Solution

4.1 Technologies Used and Database Model

After researching the most popular OS (operating systems) of mobile phones, the Android mobile platform was selected with the start level Android API 19. The open source Android Studio was used for the development of the application. The sqlite (small local store) database was used to store data, because one of the requirements for a mobile application is the ability to operate with entries without using the internet (as desired and emphasized by potential dictionary users in the previously mentioned lexicographic survey), and without any need to keep data on the distributed server. For application testing the ASUS ZenFone 2 Laser mobile (with Android 6.0.1 version and API 23 level), Oneplus 5 A5000 mobile (with Android 8.1.0 version and API 27 level) and Samsung Galaxy Tab 9,6 E (with Android 4.4.4 version and API 19 level) devices were used.

The database model is based on the dictionary document structure with the following information: (1) a term, its designation and priority in Latvian; (2) a term in Latin; (3) a term in English; (4) a term, its designation and priority in German; (5) a term, its

designation and priority in Russian; (6) a wiki link (automatically generated); (7) notes to relate the term to its visualization in interactive structure images; (8) a definition from *www.tezaurs.lv* (automatically retrieved); (9) a link to the term in *www.tezaurs.lv* (automatically generated); and (10) an info field. The dictionary document content is automatically imported into the application's *insert.sql* resource file, where all INSERT SQL queries with each term's parameters are stored by the script developed based on the Java programming language and Apache POI (Java API for Microsoft Documents). During the process of compiling the application the *insert.sql* resource file is read and executed to create a local database, tables and records. When the dictionary document is updated with some new entries, the developed script automatically updates the database and creates a new android archive **.apk* file.

4.2 The Functionality of the NBD

Special attention was paid to improving the functionality of the NBD in the second phase of dictionary supplementation and mobile application modernization. The related improvements mainly concerned a maximally user-friendly and simplified search in the main section of the dictionary, which was done by developing a special morphological approach for Latvian terms and elaborating a semantic search function – when the user sees the visualization of a taxonomic link between the plant name searched for and the language material included in the dictionary database. A semantic search function offers searching for a taxonomic category represented by a (Latin) name of the plant, i.e. higher and lower taxonomic units (genus – subgenus – species, etc.) included in the dictionary (see Figure 3 below).

4.2.1 A Simple Search Option and its Elaboration

An improvement and special development of the search function is related to the specifics of Latvian as a basic language of the dictionary. Traditionally the Latvian names of plants species are used in the singular, but genus names in the plural. Paragraph 1 of the "Botanic Term-Building Principles", approved by the Botanical Terminology Subcommittee of the Terminology Commission of the Latvian Academy of Sciences, states that in Latvian the genus names of organisms should be put in the plural and species names in the singular (LZA TK TJ No. 10, 2004: 22), so the Latvian names of plants genus included in the dictionary are given only in the plural forms. However, in spoken language the singular form of a genus is often used (although incorrect), so the dictionary has a search function for both cases. As word endings in Latvian in the singular and plural forms are different, the programmer had to find a solution for cases when the user enters the word in the search box in either singular or plural forms. Thus the possibility of listing both forms in the database and getting the needed form through a simple lookup is not used in the app, although that might seem a simpler solution.

As Latvian is a flexive language, there are very different ways that word endings can change (see Table 1). The most difficult are the cases when one word has two or more different endings in the grammatical category of number – singular or plural. This section of the article offers an overview on how to make it possible to find words with both endings in the application database, and what was the programmer's approach and solution to this issue. In the lexicographic survey (Sviķe, 2018), one of the users' preferences was a simple search function as well as the ability to search the database, even if the word was entered in the search box slightly differently. In order to improve the functionality of the app, it was necessary to introduce an additional function – a search option regardless of the singular or plural form of the Latvian word is entered. When implementing the dictionary development project, a method was developed that performs an change in ending recognition algorithm and finds the combination of corresponding changes in word endings, as in Table 1.

In most cases the ending *-as* in plural changes to *-a* in singular (e.g. *aronijas* (pl.) and *aronija* (sg.)), but there are also some more difficult cases: for example, if in plural the ending is *-ņi*, then in singular that could be *-nis* (*alkšņi* (pl.) and *alksnis* (sg.)) or *-ņš* (*amoliņi* (pl.) and *amoliņš* (sg.)). To implement the solution for the singular and plural substitutions, all ending variants were stored in the application resource file. First of all, the word is searched for in the database with no substitution of an ending. If the query returns a positive result from the database, this means the word was found, and the process of translating into others languages and searching for a definition starts. If the word is not found, ending substitution starts: (1) the last symbols of the word are compared with the endings (shown in Table 1), and the algorithm starts searching for a suitable ending pattern; (2) if there is more than one corresponding pattern of an ending, a list with all of them is created, if only one pattern is suitable, then it is stored in the list as the only element; (3) regarding the list of suitable ending patterns, the ending of a word is substituted for an ending from the list, and the algorithm checks whether the changed word is available in the database. If so, the word substitution is successful and the process of translating and searching for a definition can begin. If not, then the next pattern in the list is checked. For small databases (such as the NBD, with 2,000 entry words in Latvian and their equivalents in contrasted languages) the algorithm is quick, but for larger databases the algorithm update may be needed.

Plural	**Singular**	**Example**
-s	-a	aronija**s** → aronij**a**
-s	-e	purene**s** → puren**e**, lapegle**s** → lapegl**e**
-či	-cis	laka**či** → laka**cis**
-dži	-dzis	da**dži** → da**dzis**
-i	-s	artišok**i** → artišok**s**, bērz**i** → bērz**s**
-i	-š	ceriņ**i** → ceriņ**š**, augstiņ**i** → augstiņ**š**
-ji	-is	ķirb**ji** → ķirb**is**
-ļi	-lis	āmuļ**i** → āmu**lis**, paegļ**i** → paeg**lis**, fizāļ**i** → fizā**lis**
-ļļi	-llis	amariļļ**i** → amary**llis**
-ši	-sis	buk**ši** → buk**sis**, o**ši** → o**sis**
-ši	-tis	sunīš**i** → sunī**tis**, jānīš**i** → jānī**tis,** žibulīš**i** → žibulī**tis**
-šļi	-slis	grīšļ**i** → grīs**lis**
-šņi	-nis	alkšņ**i** → alk**snis**
-ži	-dis	skābarž**i** → skābar**dis**
-ņi	-nis	doņ**i** → do**nis**, apiņ**i** → api**nis**
-i	-us	zeltlieti → zeltlietus

Table 1: Change of word endings

In order to improve the search function, different cases of endings changing from plural into singular were analysed (plural → singular). Considering these changes, as well as the fact that some words also have changes of consonant in the root, e.g. *alkšņi*, *ķirbji*, the search methodology was adapted to the tradition of using Latvian plant names – plant genus. All consonant substitutions are included in Table 1, and the methodology of changing endings is the same as described above.

The algorithm developed during the study also performs its function in reverse, from singular to plural. The word searched for is displayed in the app's input or search window, but after the recognition of a change in ending this word appears in the results section. The related algorithm is developed for the material compiled in Latvian, i.e. for the Latvian part of the application, but in further stages of improving the app it could also be developed for other languages used in the dictionary.

4.2.2 Semantic Search

One of the most characteristic features of hybrid and printed dictionaries is an innovative search function (Tono, 2009: 65). Such solutions are also found in the NBD, for which a semantic search system was elaborated. During the implementation of the project, work was performed on the representation of taxonomic categories, e.g. a link between genus and species, or a display of the semantic search function (referring to plant names). A new section, semantic search, was created, which performs a semantic search only according to the scientific (Latin) name of the plant included in the dictionary. The algorithm that was developed can successfully process simplified cases (see Figure 3).

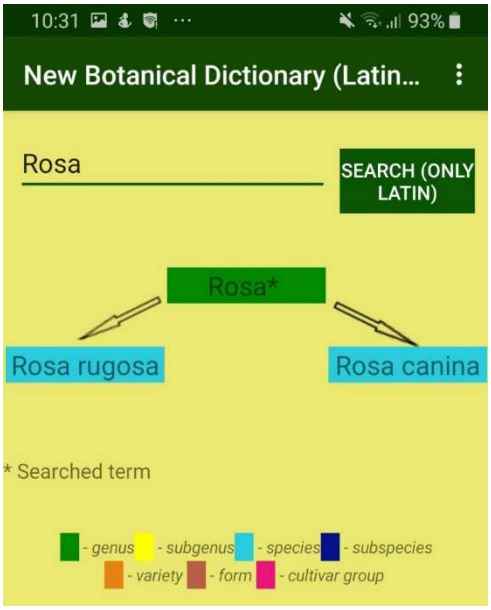

Figure 3: Semantic search (Rosa view after a semantic search)

Figure 3 shows that the dictionary includes two species of roses – *Rosa rugosa* and *Rosa canina*. The user sees a visualized link in the taxonomic categories between the genus and species.

When implementing the project, the main task related to the semantic search was to verify whether it is possible to use this function in the application. The results of the study confirm this possibility: the algorithm is able to perform semantic search, but only for the Latin plant names included in the database. The algorithm currently being developed performs data selection from a database taking into account a Latin equivalent of the term searched, in this case – the scientific name of a plant. For example, searching for *Rosa* (at the highest taxonomic level – genus), the app searches for terms in the database at lower taxonomic levels, where the first part of terms includes the keyword *Rosa*. For the scientific names of plants in Latin, the names of lower taxonomic levels will always include the first name of the highest level (for example, genus *Rosa* and species *Rosa rugosa*). When searching from a lower taxonomic level to a higher

taxonomic level, the database searches for a word at a higher taxonomic level; this identifies the semantic tree root of the word being searched for. Moreover, in order to build a full semantic tree of a the word being searched for, both the taxonomic level and the lower levels of the word are searched for. In forming the algorithm, a "tree" data structure is used to store the selected data at specific taxonomic levels and make it easier to display semantic search results in the semantic search view of the mobile app. Usually such challenging tasks are performed by groups of computer linguists and lexicographers within long-term projects (implemented over several years). The two phases of the NBD prototype project lasted for less than a year, so the development of the semantic search function could be implemented during future upgrades of the app. This task should be carried out within possible future projects along with broadening the research task and implementing it in a more detailed way (and also for other botany terms, not just for plant names) and offering specific solutions to the related problems.

4.2.3 An Extracted and Linked Data Methodology

This subsection provides an insight into the automated data selection methodology from free-access resources, and shows how linkage with other sources of information was performed.

As mentioned earlier, one of the application's features is retrieving the definitions from the Latvian Definition Information System *www.tezaurs.lv* (referred to as Tezaurs in this subsection). For the purposes of the project a script was developed to retrieve the definitions of all dictionary entries in Latvian from Tezaurs by using the *tezaurs.lv* API (Application Programming Interface). The developed script automates the definition retrieval from Tezaurs and stores the results in the dictionary document – in MS Word or Google Sheets format (Microsoft Word was used for storing entry units within the first stage of the project, but Google Drive Sheets was used during the second stage). The script was written using Java programming language and the external library was developed using Apache POI (Java API for Microsoft Documents), with this needed to retrieve words from a Sheets or Word document and to store entry definitions in the same document. The Tezaurs API returns HTML code with tags, and filtering of results is necessary. The script algorithm includes three data filtering and processing methods:

1. The result stream from the Tezaurs API is filtered using the external library JSoap (Java HTML parser) and by using an eliminator for the division of HTML tags "div", "sv_Sense", "span", "sv_NO". It is important to note that also multiple definitions might be retrieved, and it is necessary to automate choosing the right one. This is done by comparing the scientific (Latin) names, because most definitions in Tezaurs include the scientific names. For example, when searching for a definition of the Latvian *ābols* (*apple* in English), Tezaurs retrieves three definitions in Latvian from which only one is related to *ābols.* 1. *Sulīgs daudzsēklu auglis, raksturīgs ābelēm, bumbierēm, cidonijām, pīlādžiem u.*

c. (in English: juicy multi-seeded fruit, common for apple trees, pear trees, quince trees, rowan-trees, etc.)*;* 2. *Āboliņš* (it is a Latvian plant name (*clover* – in English), not an apple as required by the NBD); 3. *Parastais ķirbis* (it is a *field pumpkin* regionally called *ābols*, not an apple as required by the NBD).

2. If the word is not found in the Tezaurs database then additional filtering is carried out to search for synonyms in the NDB database and look for definitions of a specific synonym. For example, when the word searched is *ziemasteres* (Latvian plant name of a genus *Symphyotrichum*), but the Tezaurs API retrieves only the link to an entry *miķelītes* (*aster* in English), it is necessary to retrieve the definitions by using the link, because the definitions of both words given in Tezaurs are the same.

3. The Tezaurs API does not retrieve the definition of a word if this word was not entered in the correct form (plural or singular). In such cases the algorithm developed for word substitution from plural to singular or opposite is used. For example, when looking for the plant name *akanti* (*bear's-breech* in English), the Tezaurs API retrieves no results, but when substituting the term *akanti* to its singular form, *akants*, the Tezaurs API retrieves the definition. This implementation includes the algorithm described in Section 2, above.

After filtering and processing the data (when the Tezaurs API is used), the retrieved definitions are stored in a database table column "def_tez". The developed program is intended to be used only for obtaining definitions automatically from the Tezaurs database, and is not responsible for the correctness of the definitions and relevance to the term searched.

4.2.4 Analysis and Correction of Automatically Retrieved Definitions

The definitions included in the entries and automatically retrieved from *www.tezaurs.lv* are important additional information, as the NBD provides both a translation and explanation of the entry words it includes. It should also be noted that the study revealed that the definitions which are retrieved automatically are only a temporary solution in providing an explanatory function of the dictionary. The desire to link the newly compiled electronic dictionary with other existing lexicographic sources was mentioned by respondents in a lexicographic survey conducted before the dictionary was developed (Sviķe, 2018: 228-241). An insight into the problem of automatically retrieved definitions was given at the international scientific conference "The Word: Aspects of Research" (*Vārds un tā pētīšanas aspekti*) organised by Liepaja University on November 29-30, 2018 in Liepaja. The study concluded that the automatically retrieved definitions have many inaccuracies and even errors, so they need to be corrected and aligned with their information layout. As an example, the definition of *aronijas* (*chokeberry* in English) retrieved automatically from *tezaurs.lv* is translated into English and described below:

Rose family (genus "Aronia") deciduary shrubs with glossy, elliptical leaves, white flowers, black berries, 3 species (native to eastern part of North America, from Ontario to Florida), all introduced in Latvia.[3]

The derived definitions in the original – Latvian – language are given in footnotes (for comparison). First of all, it should be noted that there is a mistake at the beginning of the definition – *Rose family ("Aronia" genus)*, because the scientific (Latin) name in Latvian should not be put in quotation marks. Similarly, the wording of the definition needs to be corrected. It should also be noted that the fruits of chokeberries are pomes. The correct definition translated into English would be:

The rose family genus of a deciduary plant. Shrubs have glossy, whole and elliptical leaves. Flowers are white. Fruits are black pomes. The genus has 3 species.[4]

By analysing the definitions automatically retrieved from *tezaurs.lv*, a methodology has been elaborated for developing a basic variant of the definition, where the definition has been applied and adapted to the taxonomic level of a plant name in the NBD, as shown with the following example of *lotus*.

When searching for the word *lotosi* (*lotus* in English) in *tezaurs.lv*, the following definitions were found, which describe the order, the family of this order and the genus (*lotosi* in Latvian):

1. Divdīgļlapju klases gundegu apakšklases rinda ("Nelumbonales"), kurā ir tikai viena dzimta; 2. Šīs rindas dzimta ("Nelumbonaceae") ar 1 ģinti; 3. Šīs dzimtas ģints ("Nelumbo"), kurā ir 2 sugas, ūdensaugs ar lielām lapām un krāšņiem ziediem, kas sakņojas zemē, bet zieds atveras virs ūdens.

The translation of the definitions into English is: 1. An order of dicotyledon class, crowfoot sub-class ("Nelumbonales") with only one family; 2. Family of this order ("Nelumbonaceae") with 1 genus; 3. The genus of this family ("Nelumbo"), consisting of 2 species, aquatic plants with large leaves and bright flowers rooted in the ground and the flowers opening above the water.

Since the NBD requires a definition that characterizes genus, the third definition is appropriate, but there is still a need for corrections. The above definitions could be combined into one by correcting them as follows:

Lotosu dzimtas (Nelumbonaceae) ģints. Ģintī ir 2 sugas. Ūdensaugs ar lielām lapām un krāšņiem ziediem, kas sakņojas zemē, bet zieds atveras virs ūdens. (In English - *The*

[3] In Latvian: Rožu dzimtas ģints ("Aronia"), vasarzaļi krūmi ar spīdīgām, eliptiskām lapām, baltiem ziediem, melnām ogām, 3 sugas (Ziemeļamerikas austrumu daļā no Ontārio līdz Floridai), visas introducētas Latvijā.

[4] In Latvian: Rožu dzimtas ģints vasarzaļi augi. Krūmi, ar spīdīgām, veselām un eliptiskām lapām. Ziedi balti. Augļi melni āboli. Ģintī 3 sugas.

genus of a lotus-lily family (Nelumbonaceae). The genus has 2 species. An aquatic plant with large leaves and bright flowers rooted in the ground and the flowers opening above the water).

Pursuant to the taxonomy category that specifies the NBD entry – genus, the higher taxonomic name of a genus is added in Latvian, in genitive case – *lotosu* – (which replaces the pronoun *šīs*) and the word *family* with its scientific (Latin) name in brackets without quotation marks (but italicized) according to the Latvian punctuation traditions. The scientific (Latin) name of the genus is not needed in the definition, as the scientific (Latin) name is included in the translating section of the dictionary after the label LA, so in the definition it was deleted. The word *kurā* used as the link (..*kurā ir 2 sugas*..) is replaced by the taxonomic category *ģintī* (in English – genus). In order to maintain a structure similar to the corrected Aronia definition, an auxiliary clause was not used, but a new sentence has been started in which the word *ģints* is written with a capital letter. The description of the plant with the word *ūdensaugs* is also given in a new sentence.

As mentioned before, the Tezaurs API was used for to retrieve a definition from the *www.tezaurs.lv* database. The result is the HTML output stream of the Tezaurs API filtering using the JSoup parser and specific HTML tags. In this case, the verification by scientific (Latin) name in all retrieved definitions is carried out. If the definition consists of the words "Šis" or "Šī" ("This" in English in plural and singular forms), then it is a wrong definition, so processing is necessary. The concatenation of both definitions is done by cutting out the repeating parts of the concatenated definition. The algorithm works with multiple definitions as well.

The examples described may be one of the possible solutions for further reviews and corrections of new definitions done by the dictionary compilers. It is certainly important to verify the correctness of all definitions. The explanation should include the most important features only – this is a lexicographic axiom (Baldunčiks, 2012: 118). It should also be noted that for plant names, which make up the majority of the NBD entries, there is no strict difference between the encyclopaedic and philological definition formulation approach described by Melita Stengrevica (Stengrevica, 1998: 115-120). Without describing the appearance, lifestyle or function of the plant concerned, the meaning of the name of the plant cannot be formulated. The definitions added in offline mode provide the dictionary user with a concise, precise definition of the essential features of the denoted realia. However, in online mode it is possible to quickly access more information by using hyperlinks. Due to the limited length of this article, these aspects have not been addressed, but further research by the authors is certainly required in order to retrieve, combine, correct, and write definitions.

5. Short Summary, Conclusions and Future Plans

This research paper describes a prototype of the mobile app – a dictionary structure that includes a basic part and a visual part (images with terms). The paper specifically analyses problematic cases that required some special solutions, i.e. the development of search function in Latvian both in singular and plural, as well as the semantic search for displaying the taxonomic categories of plant names.

The authors of the study have researched 18 different types of changes in ending in the language material collected in the database (e.g. the plural ending *-ņi* in singular might be *-nis* (*alkšņi* (pl.) and- *alksnis* (sg.)), or *-ņš* (*amoliņi* (pl.) and *amoliņš* (sg.)), therefore, a new methodology for processing language material was developed.

The study concludes that automatically retrieved data (definitions) should still be reviewed by an experienced lexicographer in collaboration with an industry expert to develop an optimal language material (definition) solution. It is still necessary to test the already developed NBD functionality and evaluate users' feedback, as well as implement possible corrections and improvements.

During the next stages of improving the application it will also be necessary to include a feature that could hide an automatically retrieved definition, e.g. by adding an information extension and reduction function (+/-). The authors hope that in the near future users will receive the NBD app described in this article, which is intended to be supplemented with 2,500 to 3,000 entries, and the dictionary will be useful not only for translators, but also for students of science, educators, and all others interested in the world of flora.

6. Acknowledgements

The research is supported by funding received within the VUAS internal call for proposals "Development of Scientific Activity at the Ventspils University of Applied Sciences" for years 2017 and 2018.

7. References

Baldunčiks, J. (2012). Pārskats par nozīmīgākajām vienvalodas vārdnīcām: skaidrojošās vārdnīcas, svešvārdu vārdnīcas, etimoloģijas vārdnīca, slenga vārdnīca. In: A. Lauzis (ed.). *Vārdnīcu izstrāde Latvijā 1991–2010.* Pētījums J. Baldunčika vadībā. Rīga: Latviešu valodas aģentūra, pp. 108–190.

Balode, I. (2012). Vācu-latviešu un latviešu-vācu leksikogrāfija (1991–2010). In: A. Lauzis (ed.). *Vārdnīcu izstrāde Latvijā 1991–2010.* Pētījums J. Baldunčika vadībā. Rīga: Latviešu valodas aģentūra, pp. 16–61.

Fuertes-Olivera, P.A. & Tarp, S. (2014). *Theory and Practice of Specialised Online Dicionaries. Lexicography versus Terminography.* Berlin/Boston: Walter de

Gruyter.

Granger, S. (2012). Introduction: Electronic lexicography – from challenge to opportunity. In: S. Granger, M. Paquot (eds.) *Electronic Lexicography*. New York: Oxford University Press, pp. 1–11.

Helviga, A. & Peina, E. (2016). Mūsdienu latviešu terminogrāfijas raksturojuma daži teorētiskie un praktiskie aspekti. In J. Baldunčiks, I. Jansone, A. Veisbergs (eds.) *Latviešu valodas vārdnīca*. Valsts valodas komisijas raksti. (8). Rīga: Zinātne, pp. 127–158.

Latvijasdaba.lv. *Latvijas daba: sugu enciklopēdija*. Accessed at: www.latvijasdaba.lv (31 May 2019)

Statista.com. *Mobile Operating System Market Share Latvia June 2018 – July 2019*. Accessed at: www.statista.com (22 July 2019)

LZA TK TJ No. 10 (2004). *LZA TK Terminoloģijas Jaunumi*. Nr. 10. V. Skujiņa (resp. ed.). Rīga: LZA Terminoloģijas komisija.

Stengrevica, M. (1998). Daži vārdi par Latviešu literārās valodas vārdnīcu (sakarā ar pēdējā sējuma iznākšanu – pēcvārda vietā). In: *Linguistica Lettica*. 2. Rīga: Latviešu valodas institūts, pp. 115–120.

Sviķe, S. (2018). A New Dictionary of Botanical Terms: Data Analysis of a Lexicographic Survey. *Economics World*. May-June 2018, 6 (3), pp. 228–241.

termini.lza.lv. *Latvijas Zinātņu Akadēmijas Akadēmiskā terminu datubāze*. Accessed at: www.termini.lza.lv (31 May 2019)

tezaurs.lv. *Electronic dictionary developed by Artificial Intelligence Laboratory (AILAB)*. Accessed at: https://tezaurs.lv/ (25 July 2019)

Tono, Y. (2009). Pocket Electronic Dictionaries in Japan: User Perspectives. In: H. Bergenholtz, S. Nielsen & S. Tarp (eds.) *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang, pp. 33–67.

TTC (2007). Valsts aģentūra "Tulkošanas un terminoloģijas centrs" (State Agency "Translation and terminology Centre"). *Latviešu valodas terminoloģijas resursu kvalitātes un pieejamības apzināšana dažādās zinātnes un praktiskas darbības nozarēs*. Available at: http://www.vvc.gov.lv/export/sites/default/LV/publikacijas/vardnicu_petijums_TTC_16102007.pdf