Preliminary communication Received: 14. IV. 2022. Accepted: 6. VI. 2022.

UDK
811.163.42'374:004
811.163.42'374:793.7]:004

https://doi.org/10.33604/sl.16.30.6

Model for developing educational games based on structured dictionary data[1]

Josip Mihaljević
The Miroslav Krleža Institute of Lexicography, Zagreb
jmihaljevic@lzmk.hr
Lana Hudeček
The Institute of Croatian Language and Linguistics, Zagreb
lhudecek@ihjj.hr

ABSTRACT: The paper describes a model for transferring data from dictionary structure into games. The model is developed on the example of *Croatian Web Dictionary – Mrežnik*. *Mrežnik* has a complex structure with fields and subfields for entering data created in TLex. Dictionary data organised in a fixed structure can be selected from the dictionary, exported, and used for automatic or semi-automatic creation of educational games like spelling games, accentuation, splitting words into syllables, idioms, and word relations (e.g. synonyms, antonyms). Dictionary data can be processed by a game algorithm that creates game assignments and corrects answers based on data structure. Game types based on structured dictionary data are quizzes, fill-in-the-blank games, drag-and-drop games, typing games, and crossword puzzles. The development process will be analyzed based on these initial steps: understanding dictionary structure, conceptualizing games based on structured dictionary data, choosing game data, exporting data, processing data, and gamifying dictionary data. After explaining each of the steps, the authors propose modifications to the steps and analyze how they can be used for other e-lexicographic works.

Keywords: data export; data processing; educational games; e-lexicography; gamification; structured dictionary data; websites

1. Introduction

Gamification is a process in which we take certain game elements and put them into situations that are not considered playable by nature (Deterding et al. 2011, 10). Gamification is implemented in education, workouts, sales, and many other activities

https://orcid.org/0000-0002-7482-7663 [Josip Mihaljević]
https://orcid.org/0000-0003-4667-7781 [Lana Hudeček]

[1] This paper was written within the research project *Croatian Web Dictionary – Mrežnik* conducted at the Institute of Croatian Language and Linguistics.

to motivate players and encourage them to work more or harder. The term originated in 2002 when the programmer Nick Pelling used it to describe designing user interface for electronic devices similar to video games (Pelling, 2011). Gamification can be conducted in two ways:

1. using gamification elements, which are elements taken from many parts of the game, and implementing them into systems or programs (Ashaari et al. 2016, 870). Most common gamification elements include scoring, badges, leaderboards, progress bars, skill trees, quests, levels, storylines, avatars, social elements, time limits, reward systems, etc. (Figueroa 2015, 39–40). A well-known example of gamification by using gamification elements are shopping systems that give points to their registered buyers for buying more items. These points can be used later for awards and discounts (e.g., DM-drogerie markt shop gives their buyers a card with the app PAYBACK).[2] Gamification elements can also be used in encyclopedia systems like *Baidu Baike*[3] in which editors are scored and get badges based on their activity. These activities include editing articles and answering forum questions. The users are also ranked on leaderboards for specific areas they participate in (e.g., politics, astronomy, etc.). There are also weekly assignments that users can finish to get more points.

Figure 1. An example of using gamification element leaderboards in *Baidu Baike* encyclopedia

[2] https://www.dm.de/services/kundenprogramme-services/payback (accessed 29. III. 2022.)

[3] https://baike.baidu.com/ (accessed 29. III. 2022.)

2. by creating educational games based on learning material. A well-known example is *Merriam-Webster*[4] dictionary, which has the whole page dedicated to games made from dictionary content. There are weekly quizzes and crossword puzzles for learning words and definitions. Some dictionaries have mobile apps which can include games. One of the examples is the app *Besedoboj*[5] based on the *Slovenian School Dictionary* data from *Franček* portal.[6] The app is a word game in which a player can play against another player randomly selected online or against multiple AI programs named after famous Slovenian writers. In the game, one player types a word, and then the other player types another word that starts with the same letter with which the last word ends. There is also a time limit, obligatory inclusion of specific letters, minimum word length, etc. Players are scored, and the scores can be used to improve the avatar in the app or buy game items that will help the players while playing the game. The words from the app are linked to the dictionary from the *Franček* portal where the user can also get rewards if he logs in the dictionary with the same username as in the app.[7]

The research of gamification is usually connected with the use of technology. Currently, the Scopus database (29th March 2022) shows that 7,439 papers published from 2011 contain the keyword *gamification*.[8] Most of these papers were published in 2021 (1,328 papers), and more than half of the papers are from the field of Computer Science (5,199). There are only nine papers containing keywords *gamification* and *lexicography,* and most of them have been published recently (six from 2019 to 2022). Gamification is usually connected with crowdsourcing[9] (210 papers in Scopus containing both keywords) and crowdsourcing and gamification are not yet common practices in lexicographic projects (Kallas et al. 2019, 529). One of the recent papers (Mihaljević 2020b, 17) analyzed gamification in web encyclopedias. The author concludes that from 94 analyzed web encyclopedias only 14 include games and ga-

[4] https://www.merriam-webster.com/word-games (accessed 29. III. 2022.)

[5] https://play.google.com/store/apps/details?id=si.francek.besedoboj&hl=hr&gl=US (accessed 29. III. 2022.)

[6] https://www.xn--franek-l2a.si/ (accessed 29. III. 2022.)

[7] More about *Franček* in Perdih (2021).

[8] https://www.scopus.com/search/form.uri?display=basic#basic (accessed 29. III. 2022.)

[9] Crowdsourcing is the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community, rather than from traditional employees or suppliers (Merriam-Webster, 2021).

mification elements, that gamification in lexicography is still in the initial phase, and that it can be assumed that more encyclopedias will include gamification in the future. In that paper, the model for gamifying the *Croatian Web Dictionary – Mrežnik*.[10] However, the relation between structured dictionary data and processing for creating games has not been analyzed. That will be the focus of the research presented in this paper.

2. Methodology

Since most lexicographic works have a fixed structure in which they organise data, it is assumed that there is a way to extract the data for creating games automatically or semi-automatically by using specific programs or algorithms. The research questions in this paper are:

- RQ1. Can we define steps for creating educational games based on structured dictionary data?
- RQ2. Can those steps be used for other lexicographic works?

Based on the dictionary structure of *Mrežnik* presented in the next chapter (Figure 2), the possibilities for creating many games for learning different aspects of the Croatian language, such as phonology, spelling, definitions, idioms, etc., are explored. Each entry has the same structure and contains data in most fields. In the dictionary program TLex in which the structured data is stored, entries can be filtered based on the data they contain in specific fields. The data can then be exported in many forms, processed, and used in games.

The initially defined steps for creating games based on structured dictionary data are:

1. understanding dictionary structure;
2. conceptualizing games based on structured dictionary data;
3. choosing data for games;
4. exporting data for games;
5. processing data for games;
6. gamifying dictionary data.

These steps were tested on games created for *Mrežnik* and published on the site https://rjecnik.hr/igre/. The user can browse through the games based on their con-

[10] More about the whole gamification process in Mihaljević (2020a).

tent, target user, and game type. Each game is also added as an external link to certain lemmas.[11]

Although they are derived from a single dictionary, on the compilation and gamification of which the authors worked, it is important to consider these steps to see if they can be implemented on other lexicographic works. Therefore, each step is explained in detail, and the possibilities for using that step in gamifying other lexicographic works are analyzed.

3. *The Croatian Web Dictionary – Mrežnik*

The Croatian Web Dictionary – Mrežnik project aims to create a free, monolingual, easily searchable, hypertext, born-digital, corpus-based dictionary of the Croatian standard language. *Mrežnik* is a hypertext dictionary, as its entries and sub-entries are interconnected and linked with entries in databases created within the framework of the *Mrežnik* project and databases created by project collaborators or other Institute members within the framework of other projects. The databases created in parallel with the creation of *Mrežnik* are: a language advice database (http://jezicni-savjetnik.hr/), language advice for elementary school children (http://hrvatski.hr/savjeti/)[12], a conjunction database with a description of groups of conjunctions and their modifications, a database of explanations of the origins of idioms (http://hrvatski.hr/frazemi/)[13], a database of demonyms and ktetics[14] (http://hrvatski.hr/etnici-i-ktetici/)[15], and an etymological database. The databases created within separate projects and linked with *Mrežnik* are *Croatian Special Field Terminology* (http://struna.ihjj.hr/),[16] *Collocational Database of the Croatian Language* (http://ihjj.hr/kolokacije/),[17] *Database of Croatian Verbal Valencies* (http://valencije.ihjj.hr/),[18] *Croatian Metaphor Repository* (http://ihjj.hr/metafore/),[19] etc.

---

[11] More about the site and games in Mihaljević (2021).

[12] More about this database in Lewis and Matijević (2021a).

[13] More about this database in Kovačević (2021).

[14] The database contains the names of places and their inhabitants (masculine and feminine) and the adjectives derived from those nouns.

[15] More about this database in Vidović (2021) and Vidović (in print).

[16] More about this database in Nahod (2020).

[17] More about this database in Blagus Bartolec (2021).

[18] More about this database in Brač (2021).

[19] More about this database in Despot et al. (2019).

Mrežnik is based on the *Croatian Web Repository Online Corpus*[20] and the *Croatian Web Corpus*.[21] As it is a corpus-based and not a corpus-driven dictionary, *Mrežnik* also considers data from specialised Croatian corpora created in the Institute of Croatian Language and Linguistics (*Linguistic Corpus,*[22] *Gender Corpus, Gamification Corpus,* and *Corona Corpus* – specialised corpora created in Sketch Engine)[23] and other available printed and web sources. In compiling a dictionary entry, meanings, examples, collocations, etc. can be added to a particular entry regardless of word sketches and corpus attestations. As there is currently no representative corpus of the Croatian language, the lexicographers make informed choices from word sketches to provide data representative of the Croatian standard language, not the corpus. This means that, for example, while the collocations are primarily based on word sketches[24] and the corpora mentioned above, other collocations can be added to the dictionary even if they are not attested in the corpora, but the lexicographer intuitively knows that they are commonly used in Croatian and can be found in other sources (Hudeček and Mihaljević 2020b, 81).

Mrežnik consists of these three modules:

- the module for adult native speakers of Croatian;
- the module for elementary school children;
- the module for non-native speakers of Croatian.

Data in all three modules are coordinated and synchronised. However, each module functions as a separate dictionary compiled for a different target group.[25]

The microstructure of the *Mrežnik* module for native adult speakers is shown in Figure 2. The microstructure of other modules is derived from this basic module (Hudeček and Mihaljević 2020a, 661) by reducing or merging some fields.[26]

---

[20] http://riznica.ihjj.hr/index.hr.html (accessed 5. IV. 2022).

[21] http://nlp.ffzg.hr/resources/corpora/hrwac/ (accessed 5. IV. 2022).

[22] More about this corpus in Mihaljević and Marković (2020).

[23] https://www.sketchengine.eu/ (accessed 6. IV. 2022).

[24] https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/ (accessed 5. IV. 2022).

[25] More on the module for elementary school children in Lewis and Matijević (2021b) and more on the module for non-native speakers in Hudeček, Mihaljević and Pasini (2021).

[26] For more information on *Mrežnik* see the website http://ihjj.hr/mreznik/ and the monograph *Croatian Web Dictionary – Mrežnik* (http://ihjj.hr/mreznik/page/monografija/34/).

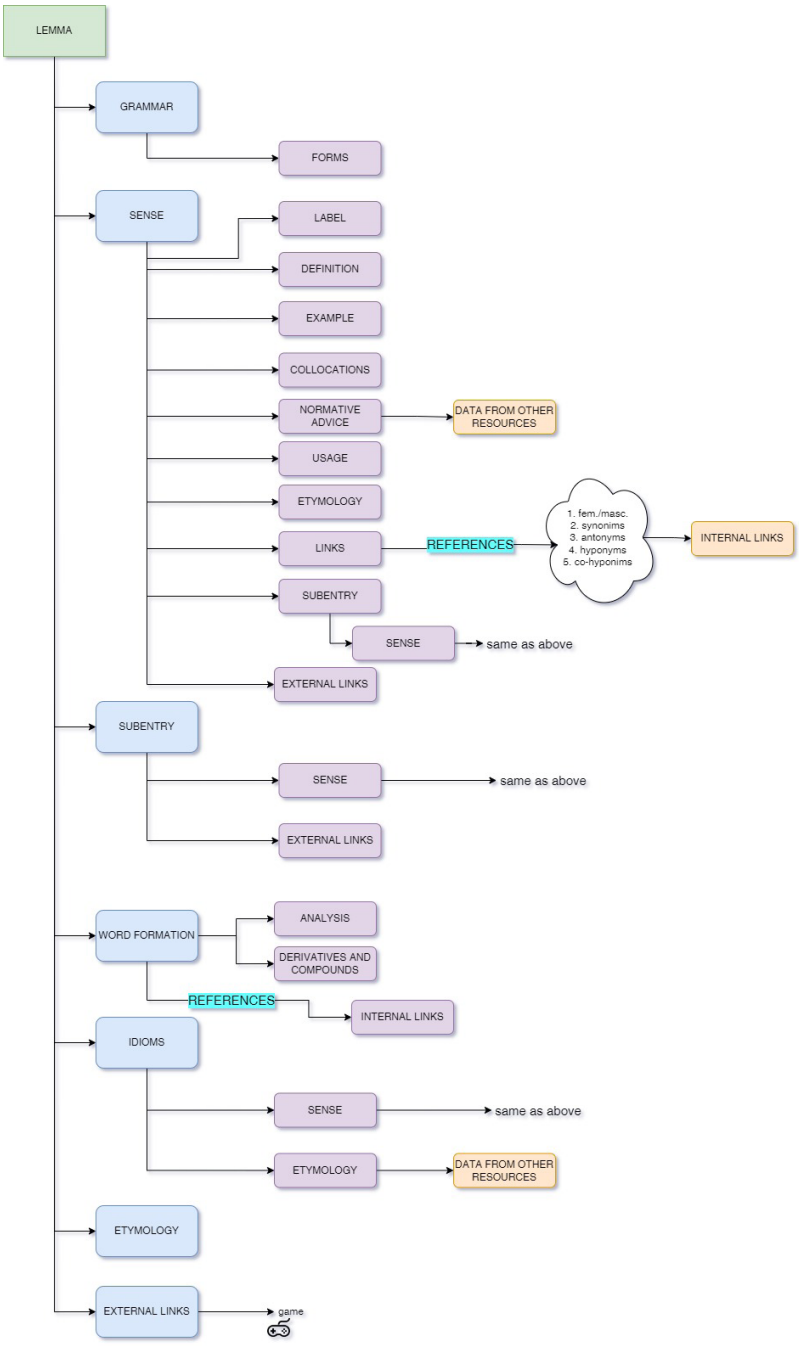

Figure 2. Data structure in *Mrežnik*

4. Steps in developing educational games based on lexicographical data structure

4.1. Understanding dictionary structure and structured dictionary data

The first step is to analyze dictionary structure and structured dictionary data in TLex. The *Mrežnik* structure in Figure 2 helps conceptualise and visualise the structure of each entry. It is important to know how the data in a dictionary or other lexicographic work is structured and stored so it can be accessed.

The *Mrežnik* TLex database is structured as a hierarchical tree of fields and subfields (e.g., the field *sense* under a *lemma* contains many subfields like *definition, collocations, examples,* etc.). Some dictionary entries have more than one field of the same type with different data (e.g., the entry *admiral* has three definitions, so three *sense* fields each have a different definition). Some fields contain different data types (e.g. text, numbers, date, hyperlinks), so this must be planned. Fields and subfields can also have values that are decided upfront and are only selectable by the lexicographer (e.g., subfield *internal link* has attributes such as *synonym, antonym, female or male pair,* etc.). It is also essential to know how data is to be interconnected. Particular field types have to be planned to enable the connection between different lemmas. Some fields can also reoccur in the structure (e.g., the field *sense* can also be used as a subfield in *subentry* and *idioms,* as shown in Figure 2).

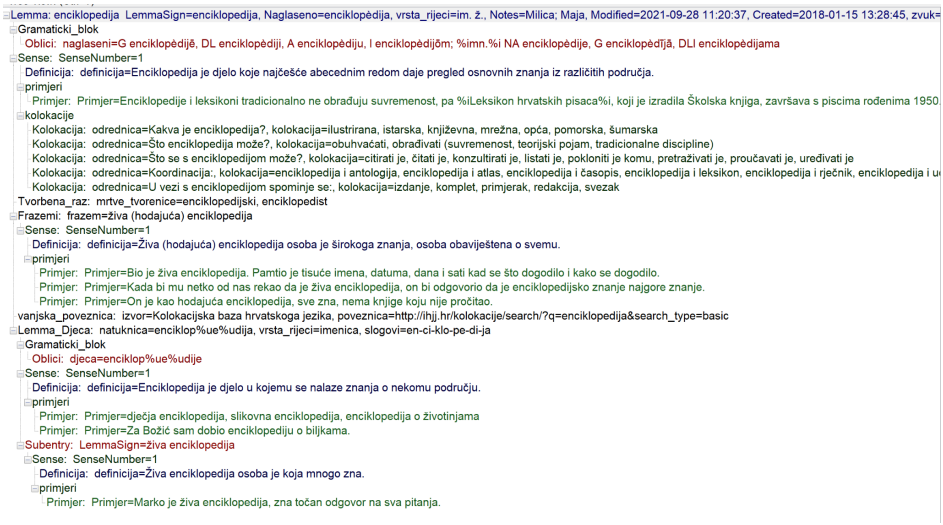

Figure 3. The entry *encyclopedia* in TLex

After the dictionary structure of *Mrežnik* was designed, creating all necessary options in TLex began. TLex uses a DTD structure[27] for compiling dictionaries. Each field and subfield was created with its attributes. Some fields were added later to expand the dictionary content, and thus, the graph showing dictionary structure was also updated (Figure 2). TLex displays the data in a hierarchical structure, so it is easier to follow the structure when working on a dictionary (Figure 3).

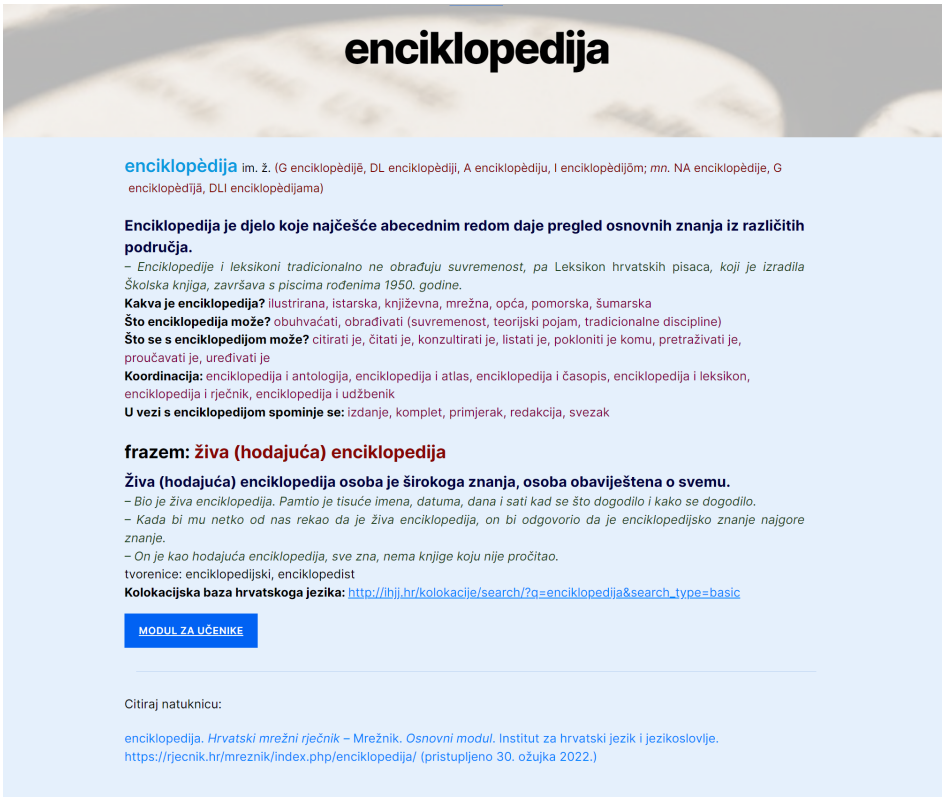

Figure 4. The entry *encyclopedia* on the *Mrežnik* website[28]

[27] DTD (Document Type Definition) is a language that defines an XML document's structure and legal elements and attributes. With DTD, we determine the XML structure of the data that has to be followed by editors when adding new data. The administrator can modify DTD data to add new XML elements and attributes to the existing structure. In dictionary software TLex, the DTD structure is created through a graphical interface, so it does not require coding skills (Joffe and Schryver, 2018, 55).

[28] enciklopedija. *Hrvatski mrežni rječnik – Mrežnik*. Osnovni modul. Institut za hrvatski jezik i jezikoslovlje. https://rjecnik.hr/mreznik/index.php/enciklopedija/ (accessed 30. IV. 2022).

This step applies to any online lexicographic work that stores the data in a database. However, certain online encyclopedic works may not use a strictly defined structure. Sometimes they are written in a free form with text and images that do not have predefined fields. Thus, certain data types are not marked. That makes defining the structure for other programs and algorithms much more complicated and future steps such as exporting data and automatic data processing almost impossible. However, games can still be made from unstructured encyclopedic data but not with this model.

4.2. Conceptualizing games based on structured dictionary data

Once we understand the dictionary structure and structured dictionary data in TLex and are familiar with the type of content, we can start conceptualizing games based on the data. This step can be broken down into three sub-steps:

1. defining the target users of the game (e.g., pre-school children, elementary school children, adult native speakers, non-native speakers);
2. selecting data from the structure for a game;
3. selecting game types that will be used for transferring data into games (e.g., quiz, fill-in-the-blanks, crossword).[29]

From most dictionary structures, we can extract data for different games. There are many possibilities for creating games from structured data. In the case of *Mrežnik*, the subfield *internal link* was used to extract data on synonyms, antonyms, male-female pairs, and aspectual pairs.[30] Games developed from these fields mostly targeted adult native speakers. It was decided at the initial stage of game development that fill-in-the-blank would be the best game type for learning semantic and grammatical relations between words; e.g. in the antonym game, the pair *snažno – slabo (strong – weak).* The game algorithm divides the pair, using an en dash (–) between elements, and chooses one of the words for the player to guess. The other word is replaced by an empty field where the player is asked to insert a synonym.

Words that contain *ije/je*, *č/ć,* and *dž/đ* (the spelling of which can be difficult even for native Croatian speakers) were chosen for a quiz game in which the player has the task to choose the correct spelling of a word from two variants; e.g. in games for practicing words with *č/ć* the player can get a question in which buttons *sreča* and *sreća (luck)* appear and if he clicks *the correct word sreća* he gets points. There is also an option to create fill-in-the-blank fields, and the player has to spell the word

[29] We also have to decide which gamification elements will be implemented in games, considering target players and possible technology restrictions.

[30] More on internal links in Hudeček and Mihaljević (2019) and Hudeček and Mihaljević (2021).

correctly. These games use scoring; no negative scores were added not to demotivate the players. There is also an option to choose the level by picking the number of questions. In a spelling game, the player can also select the question types, and more experienced players can make a game more challenging by turning on the time limit. Moreover, exported word lists can be used for creating a typing game in which the player has to type the words correctly in a short time.

The field *idioms* was also selected for creating two game types: drag-and-drop game and crossword puzzles. The drag-and-drop game was primarily created for elementary school children, so some difficult idioms were removed in choosing the game data. The crossword puzzle for adult native speakers included most of the idioms from *Mrežnik*.

The module for school children contains syllable analysis of words that can also be used to create a quiz-like game in which the player has to click a particular letter in a word to split the word into syllables. A similar game is the one for learning Croatian accentuation.[31] The player gets a word and has to click on a letter (denoting the place of the accent) and then select an accent for the word. The target users of this game are adult native speakers of Croatian (e.g., high school and university students of Croatian).

Table 1. Fields from *Mrežnik* chosen for different game types and gamification elements

Field	Content	Game type	Gamification elements
word link	synonyms, antonyms, male-female pairs, aspectual pairs	fill-in-the-blank	scoring, levels, time limit
lemma	spelling	quiz (user selects one of two options), fill-in-the-blank, typing game	scoring, levels, time limit
idioms	idioms, idiom definitions	crossword, drag-and- drop	scoring
syllables	splitting the word into syllables	quiz (click to split the word)	scoring, levels, time limit
accents	accent placement, accent types	quiz (click a letter to place the accent)	scoring, levels, time limit

[31] Standard Croatian has a four-accent language based on the Štokavian dialect. It differentiates between these four cases of accentuated vowels: a short vowel with a falling tone (\\), a short vowel with a rising tone (\), a long vowel with a falling tone (^), and a long vowel with a rising tone (/), e.g., *kȉša*, *nòga*, *grâd*, *rúka*. Some vowels can also be long after the accent. This accentuation system is complicated even for native Croatian speakers (especially non-Štokavian speakers). More of the accentuation of *Mrežnik* see in Vidović 2021.

This step can be applied to lexicographic works to which the first step has been successfully applied. However, let's consider this step separately from the previous and future steps. There is an option to create games based on the existing information found in lexicographic works even if dictionary data is not structured. In that case, gamification can be performed using the conceptual framework for gamification of lexicographic works from Mihaljević (2021, 230).

4.3. Choosing data for games

Once we understand what kind of structured data the dictionary contains and what games we want to create, it is also essential to determine if we will use all of the data to create games or select only some data because of game limitations or because certain content is not appropriate for targeted players (because of their age or level). This can also be done after we export the data or even after processing it and putting it into the game. However, it is essential to know the game concepts to decide what type and quantity of data are required. Specific technologies and game types have limitations on how much they can store or present.

In the case of *Mrežnik*, since there are three modules in the database, it was much easier to choose the data for creating game assignments aimed at target players, e.g., the spelling game[32] for writing words with *ije/je, č/ć,* and *dž/đ* uses three different sets of data from each of the modules. The player can choose the module at the beginning of the game. Thus non-native speakers or elementary school children playing the game will not get some of the words intended only for adult native speakers.

Games based on the field *internal link* mostly did not use filters since their target players were adult native speakers. Certain idioms were only chosen for drag-and-drop games, which targeted younger players. Most of these idioms were filtered from the module for elementary school children.[33] The table below shows the initial number of data items used in *Mrežnik* for different games and the final list of data used in the game.

[32] https://rjecnik.hr/igre/igre/osnovni/igra-pravopis/ (accessed 25. III. 2022).

[33] More on conceptualizing games for the dictionary in Mihaljević (2021).

Table 2. Data items used for games

Data type	*Mrežnik* data items	Filtered data items
synonyms	1642	526
antonyms	700	345
male-female pairs	568	296
aspectual pairs	380	188
idioms	1123	crosswords: 1072 drag-and-drop: 48
syllables	2611	2611
accents	7979	7907
word with *ije/je*	936	918
words with č/ć	1437	1431
words with *dž/đ*	131	126

At this phase, only prototype synonyms and antonyms were selected. The most common male-female and aspectual pairs and idioms were selected. Abbreviations, e.g., čv., were also excluded. Crossword puzzles included most of the idioms, and only very rare idioms were excluded. Syllables from the elementary school module did not have to be filtered and all words were used for the game. Homonyms with different accents were also excluded to avoid confusing the player.

This step can also be applied to lexicographic works that have the option to browse through the data based on the fields. In contrast, data selection would be much more complicated if the data was not structured.

4.4. Exporting data for games

Systems and programs in which dictionaries are created usually have options to export selected data in different formats; e.g. TLex formats for exporting include XML, HTML, RTF, and TXT. It is important that the developer can process the format in the next steps.

In most cases, data is exported into structured XML or HTML files, which allows it to be parsed and transferred to a game code using programming languages. In the case of *Mrežnik*, TLex has an option to export certain data in specific fields for filtered lemmas. Thus, data could be exported from the structure directly, and therefore, there was no need for parsing. However, the exported data was still processed. Games based on word relations were exported as an RTF file, which can be opened and edited using MS Word and contained each set of related words in sepa-

rate lines. The same procedure was followed when exporting idioms, syllables, and accents. Spelling games were the easiest to export and process because there is an option to alphabetically export lists of filtered lemmas in TXT files.

Even if the structure, the fields, and the technology for storing them are different in each lexicographic work, the options to access and export the data are in most cases the same. Thus, this step can be applied to lexicographic works created in a dictionary writing system or even a program. Static pages can still be exported as an HTML file or raw text, but that format is much harder to process in the next step.

4.5. Processing data for games

The exported data has to be processed to work as a game code. Data can be processed by converting it into the game code manually. In the case of big data, it is usually processed automatically by an algorithm written in a programming language such as Python or Ruby. If processing is done automatically, data can be read from many formats (e.g., XML, TXT). However, the structure of the data has to be clear to the algorithm which processes the data so the computer knows how to use the data appropriately; e.g. in quizzes the algorithm has to recognise what element of the data will be used as a question, what are the possible answers, what are the correct answers, and what is the explanation for the correct answer. Suppose changes are frequently made to dictionary data (linguistic data) that is used in the game, e.g., more questions are added. In that case, it is essential that the data that is being used in the game is easily editable. If more people want to make changes, it is also important to allow them access to data and explain how the data is being loaded into the game code to avoid errors. In the case of games made for *Mrežnik,* games for synonyms, antonyms, male-female pairs, and aspectual pairs were copied from the RTF file to the TXT files. Each pair was written in a separate line, and separated by an en dash (–) or in the case of synonyms by the equals sign (=). Those TXT files are then opened using JavaScript at the beginning of the game. Each pair is automatically separated by a line and a question for each pair is asked by splitting pairs based on the separation symbol (– or =) and choosing one of the pair elements to create an empty field for the player to fill. The questions in the game are randomly selected and ordered.

The algorithm for the game also gives the players feedback by showing the correct answer. All incorrect answers are stored in a list, and the player can see all mistakes at the end of the game. These mistakes are also anonymously submitted to Google Forms in the background and can be checked by the developers. This is done to enable further research on the most common mistakes.

Spelling games were processed similarly; only underscores had to be introduced within the word instead of the letters *č, ć, đ,* and combinations of letters *ije, je,* and *dž* to allow fields in which the player could write the correct letter. This was processed using Python, which converted lines from the TXT file into the JavaScript code, which generates questions. That code was later loaded into the game. However, certain words that contain multiple letters *č* or *ć* (e.g., čakovečki, čarobničin, ćevapčić) had to be processed separately. The player had to write only one letter in a word, and the game took into consideration only the first occurrence and ignored the rest. These words were filtered from TLex using regex, and the code for them was manually written since the automatic processing algorithm could not give satisfactory results.

Words for the typing game did not need a lot of processing. The list of words was copied into the array within the game code. The computer program compares what the players have typed with the word from the database.

Words containing syllable marks were copied for the syllable game within the array of the JavaScript file. The words were shown without syllable separation in the game, so the players had to place the hyphen (-) in the right place to mark the syllables. Their answers are then stored and compared with the dictionary words that contain syllable separation.

The game for learning accents consists of all words with accents from a TXT file. The algorithm checks each word and recognises accents on letters based on the array. The letters with accents are replaced with the same letter without the accent (the word with the accent is still stored). The letters (as marks of the corresponding sounds) containing accents are *a, e, i, o, r, u,* and each letter was coded in an array with all possible accents (see Figure 5). When the players are given the word they have to click on the letter that they assume should be accentuated. If they choose the correct letter, they then have to choose between four accent options *(ȁ, à, â* and *á)* and length *(ā)* (e.g., click letter *a,* then select between *ȁ, à, â, á,* and *ā*).

```
akcenti = ["a", "ȁ", "à", "â", "á", "ā", "e", "ȅ", "è", "ê", "é", "ē", "o", "ȍ", "ò", "ô", "ó", "ō", "u", "ȕ",
"ù", "û", "ú", "ū", "r", "ȑ", "ŕ", "ṛ", "ṛ́", "i", "ȉ", "í", "ī"],

slova_akcenti = {
    "a": "aȁàâáā",
    "e": "eȅèêéē",
    "i": "iȉíī",
    "o": "oȍòôóō",
    "r": "rȑŕṛṛ́",
    "u": "uȕùûúū",
}
```

Figure 5. Game code for storing accents for letters within the game

Games for idioms were made on the H5P platform. Since the platform uses only GUI for creating interactive content, processing had to be done manually. However, since all idioms with their definitions were exported within the RTF format that can be opened with the Word program, idioms with their meaning can be copied in the fields to create assignments for crossword puzzles and drag-and-drop games. In a drag-and-drop game, idioms were marked with the sign * (at the beginning and end of the idiom) with explanations for the correct and incorrect answers marked with + and -. This is done so the computer can determine which answer is correct. In the crossword, an underscore line was written instead of a missing word, so the player knows which word needs to be placed in the field, and the idiom is written in a separate field so that the system knows the correct answer (see Figure 6).

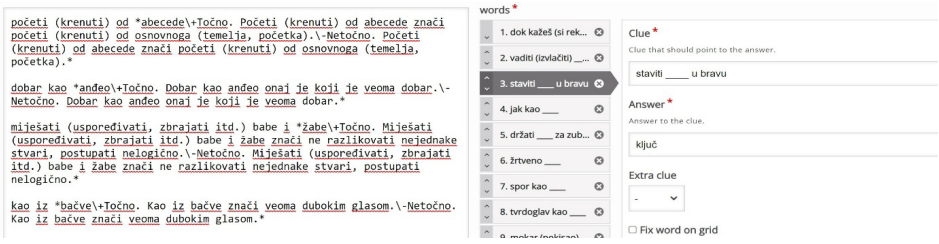

Figure 6. Processing data for the drag-and-drop game for idioms on the left and input data for the crossword puzzle on the right

Data processing is the most challenging step in this whole process. To apply this step successfully, the developers have to know gaming logic and have computer programming skills. However, this step is mainly applicable to other lexicographic works if the data is computer-readable (at least as a structured text). If we deal with raw unstructured text, processing must be done manually. This makes it almost impossible to gamify large lexicographic works.

4.6. Gamifying data into games

The last step of the initial model was applying gamification elements to the game and presenting the data from the game. Simple gamification elements such as scoring, levels, and time limits for assessment are usually easier to implement than social gamification elements such as avatars, leaderboards, and award systems because those elements require connecting games with the database (Mihaljević 2021, 110).

Social gamification elements usually require a database that stores user profiles for all players. The player's achievements and game status are stored in the user profile. Since this requires much more technical resources and skills, simpler games usually do not use these gamification elements.

Scoring is the gamification element that is the easiest to implement and gives the player instant feedback on the score. When players succeed, they get points. It has to be decided if the game will have negative scoring for incorrect answers and how many points specific questions or tasks bring. Scoring could also be connected to other gamification elements such as time limits and virtual rewards.

Time limit is used to make the game faster as it forces the player to solve problems or answer questions before the time runs out. However, it can also be stressful for some players. This is why the time limit is optional in most games in *Mrežnik*. Negative scoring was never implemented in any of the games in order not to demotivate the players. The time limit was connected with the scoring for the player who chose to play with it by giving the player more points based on response time.

Virtual rewards such as badges or some items are mainly used to acknowledge player skills. They are usually given to the player at the end of the game. Some games implemented medals (bronze, silver, and gold) for the three best players. The scores could be stored on leaderboards by the players choosing their username (no registration was required).

Some games have options to select levels based on the number of questions or content that will appear in the game; e.g., a spelling quiz has an option to choose between tasks that contain letters č/ć, *ije/je, dž/đ,* or all of them. Players can also select the number of questions (20, 50, or even 100).

When graphically designing games for websites or mobile apps, all content must be visible or accessible to the player. Modern games require a responsive design and controls to be played by players who use different systems, have different devi-

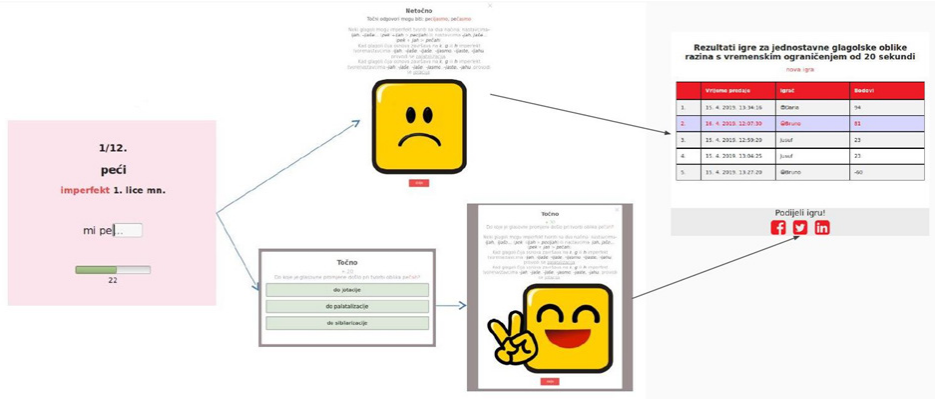

Figure 7. The display of data in combination with many gamification elements (time limit, scoring, instant feedback, leaderboards, social sharing)*

* Figure made from the game Quiz for verb forms: https://borna12.gitlab.io/igre-mreznik/kviz-mijesani/ (more about the game in Lazić and Mihaljević 2020)

ces, different resolutions, and control types (mouse and keyboard, touchscreens, and touchpads). From the beginning, games made from *Mrežnik* data were designed with this in mind. Most of the content in quizzes was centered, so it would always appear in the middle of the screen. The text size and images are defined for different resolutions (computer, smartphone, tablet). Quizzes have dynamic feedback for player actions and appear in a pop-up every time a player answers correctly or incorrectly or the time runs out. This pop-up can contain points that are colored green, symbolise something positive, and have different emoji images based on a correct or incorrect response. If the player answers incorrectly, the correct answer is given so the player can learn from playing the game.

The table below shows each of the analyzed games made using data from this model with their URL addresses.

Table 3. Games made from structured data in *Mrežnik*

Game	URL
Learn Male-Female Pairs	https://rjecnik.hr/igre/igre/osnovni/mocijski-parnjaci/
Crossword Puzzle with Phrases	https://rjecnik.hr/igre/krizaljka-s-frazemima/
Drag a Phrase	https://rjecnik.hr/igre/dovuci-frazem/
Learn Synonyms	https://rjecnik.hr/igre/igre/osnovni/sinonimi/
Learn Antonyms	https://rjecnik.hr/igre/igre/osnovni/antonimi/
Separate Word into Syllables	https://rjecnik.hr/igre/igre/ucenici/rastavi-rijec-slogove/
Recognise Accents	https://rjecnik.hr/igre/igre/osnovni/naglasci/
Spelling Game *(č/ć, dž/đ, ije/je)*	https://rjecnik.hr/igre/igre/osnovni/igra-pravopis/
Type the Word Game	http://hrvatski.hr/igra/4/

This step is only applicable to lexicographic works which have successfully implemented all previous steps. However, the control of the way the data is displayed in the game or gamification elements that were implemented may depend on the technology used, and the technical knowledge of design and programming, e.g., social gamification elements such as leaderboards are difficult to implement and maintain. Therefore, some of the initial concepts from the second step (conceptualizing the games based on the dictionary structure) can be left out due to technical limitations.

After successfully gamifying data based on this model, the games can be published and promoted, and then the user reception can be followed and the game potentially modified.

5. Results and discussion

We can see that the initially designed steps for our first research question are well defined based on analyzing each of the steps. However, some of them could not be conducted one after the other. Some steps are usually done in parallel. Choosing data for games and exporting them may be undertaken in a single step. After exporting data, we may still have to select the elements we want to keep in the game. Sometimes exported data is in a more readable format than in the original program or system it was exported from (e.g., within the exported RTF document, there was still a need to choose certain idioms for the game, and it was much easier to browse through the idiom lists in Word than in TLex). Processing and gamifying data could also be merged into one step because both involve programming and setting game logic. The gamification element of scoring is usually implemented directly when an algorithm marks the correct answer in the data processing step. There is also a possibility that after the end of the game, the developer returns to step three and then repeats the steps to choose and export, process, and gamify new data for the existing game. A new step that can be added to our initial model is the step of publishing the game. We mentioned publishing the game at the end of step five. It can be considered a process in which we have to decide who will have access to the game (for playing or testing purposes), where the game will be available to the player (on the lexicographic website, within the dictionary or encyclopedia entries, on an institutional website through an external link, or will it only be accessible in the form of a smartphone app), will it be able to collect and process data from the lexicographic database automatically or will this be done manually.[34]

Taking all this into consideration, here are the final five steps of the model for creating games based on dictionary structure and structured dictionary data:

1. understanding dictionary structure
2. conceptualizing games based on structured dictionary data
3. choosing and exporting data for games
4. processing and gamifying data for games
5. publishing games.

[34] More about game publishing in Mihaljević (2021, 266–271).

6. Conclusion

These steps still have to be tested on other lexicographic works and they could probably be further modified. This also answers the second research question of whether these steps apply to other lexicographic works. At this moment, they were tested on *Mrežnik*. The applicability mostly depends on the first step because if the data is not structured in a computer-readable way, the whole model cannot be applied as we cannot select and export the data. However, since most dictionaries and other lexicographic works are compiled in a dictionary writing system, modern lexicographic works have structured data. Another problem with applying this model are technical limitations connected with the processing and gamifying of data since this step requires a lot of knowledge in designing and programming.

REFERENCES

Ashaari, Noraidah Sahari, Firas Layth Khaleel, Tengku Siti Meriam Tengku Wook and Ismail Amirah. 2016. »Gamification Elements for Learning Applications«. *International Journal on Advanced Science, Engineering and Information Technology,* 6(6): 868–874.

Blagus Bartolec, Goranka. 2021. »Kolokacijska baza hrvatskoga jezika«. In: *Hrvatski mrežni rječnik* – Mrežnik*: Monografija, radna inačica*. Zagreb: Institute of Croatian Language and Linguistics.

Brač, Ivana. 2021. »Baza hrvatskih glagolskih valencija – e-Glava«. In: *Hrvatski mrežni rječnik* – Mrežnik*: Monografija, radna inačica*. Zagreb: Institute of Croatian Language and Linguistics.

Despot, Kristina, Mirjana Tonković, Mario Brdar, Mario Essert, Benedikt Perak, Ana Ostroški Anić, Bruno Nahod and Ivan Pandžić. 2019. »MetaNet.HR: Croatian Metaphor Repository«. In: *Metaphor and Metonymy in the Digital Age: Theory and Methods for Building Repositories of Figurative Language*. 123–146. Amsterdam: John Benjamins Publishing Company.

Deterding, Sebastian, Dan Dixon, Rilla Khaled and Lennart Nacke. 2011. »From game design elements to gamefulness: defining gamification«. In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments,* 9–15. Tampere: ACM.

Figueroa-Flores, Jorge F. 2015. »Using Gamification to Enhance Second Language Learning«. *Digital Education Review,* 27(21): 32–54.

Hudeček, Lana and Milica Mihaljević. 2019. »Croatian Web Dictionary – Mrežnik – Linking with Other Language Resources«. In: *Proceedings of the eLex 2019 conference: Electronic lexicography in the 21st century,* 72–98. Brno: Lexical Computing CZ, s.r.o.

Hudeček, Lana and Milica Mihaljević. 2020a. »The Croatian Web Dictionary – *Mrežnik*: Project – Goals and Achievements«. *Rasprave:* Časopis *Instituta za hrvatski jezik i jezikoslovlje,* 46(2): 645–667.

Hudeček, Lana and Milica Mihaljević. 2020b. »Collocations in the Croatian Web Dictionary – *Mrežnik*«. *Slovenščina 2.0*, 8(2): 78–111.

Hudeček, Lana and Milica Mihaljević. 2021. »Unutrašnje i vanjske poveznice u mrežniku«. In: *Hrvatski mrežni rječnik – Mrežnik: Monografija, radna inačica*. Zagreb: Institute of Croatian Language and Linguistics.

Hudeček, Lana, Milica Mihaljević and Dinka Pasini. 2021. »Modul za osobe koje uče hrvatski kao ini jezik«. In: *Hrvatski mrežni rječnik – Mrežnik: Monografija, radna inačica*. Zagreb: Institute of Croatian Language and Linguistics.

Joffe, David, de Schryver, Gilles-Maurice. 2018. *The TLex Suite (2012 – 2018).* Gordons Bay: Tshwane-DJe.

Kallas, Jelena, Svetla Koeva, Margit Langemets, Carole Tiberius and Iztok Kosem. 2019. »Lexicographic Practices in Europe: Results of the ELEXIS Survey on User Needs«. In: *Proceedings of eLex 2019.* Sintra: Lexical Computing CZ s.r.o.

Kovačević, Barbara. 2021. »Baza frazemskih etimologija«. In: *Hrvatski mrežni rječnik* – Mrežnik*: Monografija, radna inačica.* Zagreb: Institute of Croatian Language and Linguistics.

Lazić, Daria and Josip Mihaljević. 2020. »Morfološke igre na primjeru igara za učenje glagolskih oblika«. In: *4. Međimurski filološki i pedagoški dani: Jezik, književnost i obrazovanje – suvremeni koncepti,* 122–135. Čakovec: Učiteljski fakultet Sveučilišta u Zagrebu.

Lewis, Kristian and Maja Matijević. 2021a. »Baza jezičnih savjeta za učenike nižih razreda osnovne škole«. In: *Hrvatski mrežni rječnik* – Mrežnik*: Monografija, radna inačica.* Zagreb: Institute of Croatian Language and Linguistics.

Lewis, Kristian and Maja Matijević. 2021b. »Modul za učenike nižih razreda osnovne škole«. In: *Hrvatski mrežni rječnik* – Mrežnik*: Monografija, radna inačica.* Zagreb: Institute of Croatian Language and Linguistics.

Mihaljević, Josip. 2020a. »Igrifikacija Hrvatskoga mrežnog rječnika – *Mrežnika*«. *Rasprave: časopis Instituta za hrvatski jezik i jezikoslovlje,* 46(2): 407–434.

Mihaljević, Josip. 2020b. »Igrifikacijski elementi na mrežnim stranicama enciklopedija«. *Studia lexicographica: časopis za leksikografiju i enciklopedistiku,* 14(27): 15–32.

Mihaljević, Josip. 2021. *Konceptualni okvir igrifikacije hrvatskoga mrežnoga rječnika.* PhD thesis, Zagreb: Faculty of Humanities and Social Sciences, University of Zagreb.

Mihaljević, Josip and Mario Marković. 2020. »Korpus *Jene* – jednojezični terminološki korpus«. In: *Hrvatsko jezikoslovno nazivlje* – JENA, 110–123. Zagreb: Institute of Croatian Language and Linguistics.

Nahod, Bruno. 2020. »Terminološka baza Struna«. In: *Hrvatsko jezikoslovno nazivlje* – JENA. 1–7. Zagreb, Institute of Croatian Language and Linguistics.

Perdih, Andrej. 2021. »Učenje o slovarjih v šoli: Portal *Franček* kot most med splošno in pedagoško leksikografijo«. In: *Lexicography and Lexicology in the Light of Current Issues,* 1119–1132. Belgrade: Institut za srpski jezik SAZU.

Vidović, Domagoj. 2021. »Baza etnika i ktetika«. In: *Hrvatski mrežni rječnik – Mrežnik: Monografija, radna inačica.* Zagreb: Institute of Croatian Language and Linguistics.

Vidović, Domagoj. 2022. »The treatment of demonyms, ktetics and exonyms in the more recent printed and online sources of the Institute of Croatian Language and Linguistics«. *Studia lexicographica: časopis za leksikografiju i enciklopedistiku,* 16(30): 21–37.

WEB SOURCES

Baidu Baike: https://baike.baidu.com/ (accessed 29. III. 2022)

crowdsourcing – Merriam-Webster: https://www.merriam-webster.com/dictionary/crowdsourcing (accessed 1. VI. 2022)

Dovuci frazem – Mrežnik igre: https://rjecnik.hr/igre/dovuci-frazem/ (accessed 5. IV. 2022)

E-glava: http://valencije.ihjj.hr (accessed 29. III. 2022)

enciklopedija – Hrvatski mrežni rječnik – *Mrežnik*: https://rjecnik.hr/mreznik/index.php/enciklopedija/ (accessed 30. IV. 2022)

Etnici i ktetici – Hrvatski u školi: http://hrvatski.hr/etnici-i-ktetici (accessed 29. III. 2022)

Franček: https://www.xn--franek-l2a.si/ (accessed 29. III. 2022)

Frazemi – Hrvatski u školi: http://hrvatski.hr/frazemi/ (accessed 29. III. 2022)

Google play – Besedoboj: https://play.google.com/store/apps/details?id=si.francek.besedoboj&hl=hr&gl=US (accessed 29. III. 2022)

Hrvatska jezična riznica: http://riznica.ihjj.hr/index.hr.html (accessed 5. IV. 2022.)

hrWaC – Croatian web corpus: http://nlp.ffzg.hr/resources/corpora/hrwac/ (accessed 5. IV. 2022)

Jezični savjetnik: http://jezicni-savjetnik.hr/ (accessed 29. III. 2022)

Kolokacijska baza hrvatskoga jezika: http://ihjj.hr/kolokacije (accessed 29. III. 2022)

Križaljka s frazemima – Mrežnik igre: https://rjecnik.hr/igre/krizaljka-s-frazemima/ (accessed 5. IV. 2022)

Mrežnik igre: https://rjecnik.hr/igre/ (accessed 5. IV. 2022)

Mrežnik: http://ihjj.hr/mreznik/ (accessed 29. III. 2022)

Napiši antonim za zadanu riječ – Mrežnik igre: https://rjecnik.hr/igre/igre/osnovni/antonimi/ (accessed 5. IV. 2022)

Napiši mocijski parnjak za zadanu riječ – Mrežnik igre: https://rjecnik.hr/igre/igre/osnovni/mocijski-parnjaci/ (accessed 5. IV. 2022)

Napiši sinonim za zadanu riječ – Mrežnik igre: https://rjecnik.hr/igre/igre/osnovni/sinonimi/ (accessed 5. IV. 2022.)

PAYBACK bei dm: https://www.dm.de/services/kundenprogramme-services/payback (accessed 29. III. 2022)

Pelling, Nick – Funding Startups (& other impossibilities) (2011): https://nanodome.wordpress.com/2011/08/09/the-short-prehistory-of-gamification/ (accessed 5. IV. 2022)

Postavi naglaske na slova – Mrežnik igre: https://rjecnik.hr/igre/igre/osnovni/naglasci/ (accessed 5. IV. 2022)

Pravilno pisanje riječi – Mrežnik igre: https://rjecnik.hr/igre/igre/osnovni/igra-pravopis/ (accessed 5. IV. 2022)

Rastavi riječ na slogove – Mrežnik igre: https://rjecnik.hr/igre/igre/ucenici/rastavi-rijec-slogove/ (accessed 5. IV. 2022)

Repozitorij metafora hrvatskoga jezika: http://ihjj.hr/metafore/ (accessed 29. III. 2022)

Savjeti – Hrvatski u školi: http://hrvatski.hr/savjeti/ (accessed 29. III. 2022)

Scopus: https://www.scopus.com/search/form.uri?display=basic#basic (accessed 29. III. 2022)

Sketch Engine: https://www.sketchengine.eu/ (accessed 6. IV. 2022)

Struna – Hrvatsko strukovno nazivlje: http://struna.ihjj.hr/ (accessed 29. III. 2022)

Utipkaj riječi – Hrvatski u školi: http://hrvatski.hr/igra/4/ (accessed 5. IV. 2022)

Word Games – Merriam-Webster: https://www.merriam-webster.com/word-games (accessed 29. III. 2022)

Word Sketch – Sketch Engine. https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/ (accessed 5. IV. 2022)

MODEL IZRADE OBRAZOVNIH IGARA NA TEMELJU STRUKTURIRANIH RJEČNIČKIH PODATAKA

Josip Mihaljević
Leksikografski zavod Miroslav Krleža, Zagreb
jmihaljevic@lzmk.hr
Lana Hudeček
Institut za hrvatski jezik i jezikoslovlje, Zagreb
lhudecek@ihjj.hr

SAŽETAK: Ovaj rad prikazuje model prebacivanja strukturiranih podataka iz rječnika u igre. Model je osmišljen i oprimjeren na primjeru *Hrvatskoga mrežnog rječnika – Mrežnika*, u čijoj su izradi sudjelovali autori rada. *Mrežnik* ima razrađenu strukturu obrade natuknica unutar koje su određeni podatci koji se unose u određena polja. Ta polja sadržavaju podatke iz kojih se mogu napraviti igre poput definicija riječi, rastavljanja na slogove, naglašavanja riječi, sinonima, antonima, frazema, mocijskih parnjaka, vidskih parnjaka, tvorenica itd. Također riječi koje sadržavaju glasove ili skupove koji su povezani s pravopisnim problemima i za izvorne govornike hrvatskoga jezika *(ije/je, č/ć, đ/dž)* mogu se vježbati s pomoću pravopisnih igara. Podatci se iz tih polja u određenim natuknicama mogu izvesti iz rječničke baze te se ti podatci mogu iskoristiti za poluautomatsko stvaranje i ispravljanje zadataka. Primjerice, izvozom sinonima iz rječnika dobivaju se parovi riječi, pa se može računalno obraditi tekst kako bi se oblikovali zadatci i odredili točni odgovori. Na temelju izvezenih rječničkih podataka izrađeni su idući tipovi igara: igre popunjavanja praznina, kvizovi, igre dovlačenja, daktilografske igra i križaljka. Neke su igre trenutačno dostupne na portalu za igre https://rjecnik.hr/igre/. Prvotno izrađen model izrade obrazovnih igara na temelju strukturiranih rječničkih podataka objasnit će se iscrpno u koracima koji uključuju razumijevanje strukture rječnika, osmišljavanje igara, odabir podataka, izvoz podataka, računalnu obradu podataka te igrifikaciju podataka. Nakon što je svaki korak objašnjen, koraci modela malo su izmijenjeni na temelju novih zaključaka istraživanja te se analizira kako bi se novoosmišljeni model mogao upotrijebiti u drugim e-leksikografskim izdanjima.

Ključne riječi: e-leksikografija; igrifikacija; izvoz podataka; mrežne stranice; obrada podataka; obrazovne igre; strukturirani rječnički podatci