

Project Proposal: Developing a Gamified Interface and Lexicographic Database for Abstract Wikipedia's Multilingual Content Generation

Tadiwa Magwenzi

The University of Cape Town Computer Science
Department

Zahraa Hoosen

The University of Cape Town Computer Science
Department

1 INTRODUCTION

The advancement of technology has created an increasingly interconnected world, where access to information and knowledge-sharing are paramount, as evidenced by platforms such as Wikipedia. However, knowledge-sharing is limited by the languages in which articles are written, creating a need for an inclusive platform that transcends language barriers. Wikipedia seeks to address this disparity in knowledge across different languages with the Abstract Wikipedia project, which will utilise a language-independent structure that allows for the automatic generation of articles. This approach is more scalable than the current method of manually written articles and enables the same article to be available in multiple languages. However, extensive lexicographical data is required for this project, particularly for low-resourced African languages such as the Niger-Congo B languages.

2 ABSTRACT WIKIPEDIA PROJECT

The Abstract Wikipedia project aims to create a new type of multilingual knowledge repository that utilises natural language generation (NLG) and renderer functions stored in Wikifunctions to generate articles from abstract data and language-specific lexicographic data stored in Wikidata. The project's user interface enables users to create templates that can be filled in to generate articles.

However, the project's multilingual article generation is impeded by the lack of lexicographical data, particularly for low-resourced languages (LRLs). This hinders the project's ability to identify concepts and similarities across languages, making it challenging to produce articles in different languages.

To address this challenge, we will focus on improving the current interface for adding lexicographic data. We will also work on a lexicographical database within the Wikidata component. Our overall goal is aiding the Abstract Wikipedia project by increasing its lexicographical data store. This will help Wikipedia become more accessible to readers of LRLs and facilitate knowledge sharing, which is Wikipedia's goal. Figure 1 illustrates how the article generator uses Wikidata, Wikifunctions, and the interface to produce articles.

3 PROBLEM STATEMENT

Abstract Wikipedia aims to create a new kind of multilingual knowledge repository that is easily translatable and machine-readable. However, its multilingual article generation faces several challenges,

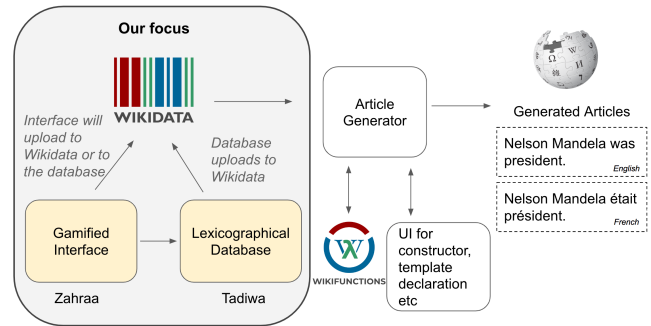


Figure 1: Abstract Wikipedia diagram showing the focus of our project

particularly the lack of lexicographical data for LRLs. This deficiency affects the project's ability to identify concepts and similarities across languages and produce articles in different languages.

Our project aims to address the lack of lexicographical data for LRLs specifically Niger-Congo B languages.

3.1 Problem List

The current problems faced by the lexicographical component Abstract Wikipedia:

- Insufficient lexicographical data for LRLs, which hinders multilingual article generation.
- A complex interface for adding lexicographical input on Wikidata.
- A lack of contributors for LRLs.
- Difficulty in identifying concepts and similarities across languages due to insufficient lexicographical data.
- Lack of language support for LRLs often requiring approval to add lexicographical data from these languages.
- A lack of a secondary, Wikidata-integrated database that can perform batch uploads to Wikidata.
- Limited accessibility and shareability of knowledge across different languages and cultures.
- There are limits to, or a complete lack of separate databases than can store and upload lexicographical data to Wikidata
- Inefficient tools and resources for representing and organising knowledge in LRLs.

In our project we will focus on two particular problems, in the context of LRLs:

- Problem A: A general absence of dedicated databases that can store and upload lexicographical data to Wikidata for LRLs. There are also severe limitations to the ones in existence.
- Problem B: The current Wikidata interface is aimed at expert users (as seen in Figure 2) and for LRLs, this pool of users is

Create a new Lexeme

You can check whether a Lexeme already exists by using the search. You can also learn more about Lexemes in the help box below.

Warning: You are not logged in. Your IP address will be publicly visible if you make any edits. If you log in or create an account, your edits will be attributed to a username, among other benefits.

Lemma *

Base form of a word, e.g. 'cat'

Lexeme's language *

The Lexeme's language, e.g. 'English'

Lexical category *

The Lexeme's lexical category, e.g. 'noun'

By clicking "Create Lexeme", you agree to the terms of use, and you irrevocably agree to release your contribution under the Creative Commons CC0 License.

Create Lexeme

About Lexemes

Lexemes contain lexicographical data which is data about words or phrases, such as language, etymology, inflections, etc. Here is an example:

cat Language: English
en Lexical category: noun

Lexemes don't contain general data (date of birth, opening date, author, country, coordinates, website, etc.) about the entity or concept to which they refer. If you want to submit general data, you need to create an Item instead.

Figure 2: Wikidata Interface in Problem B

small. Other interfaces are aimed at collecting data for Indo-European languages (which do not take the agglutinative structure of Niger-Congo B languages into account).

3.2 Goals

By accomplishing these goals, the project aims to address the under-representation of Niger-Congo B languages in Wikidata. This will benefit the speakers of these languages while enriching the linguistic diversity within Wikidata. The two specific goals for this project are as follows:

The first goal is to develop a separate lexicographical database, modeled after Niger-Congo B languages, where data can be collected and batch uploaded to Wikidata.

The second goal is to research and implement two gamified interfaces for adding lexicographical data to Wikidata to make it more accessible for more contributors.

We will describe in the next two subsections our specific goals for the project.

3.2.1 Development of a separate lexicographical database, that collects and batch uploads Niger-Congo B lexicographical data to Wikidata. The main aim of this section would be developing a separate lexicographical database that is modeled after Niger-Congo B languages. The idea is that this database would be fully usable for collection and batch upload of lexicographical data to Wikidata. Most importantly, it would be usable for any Niger-Congo B language.

The goal specific aims of this section are:

- (1) To develop a robust and stable database that can hold lexicographic data from a Niger-Congo B languages.
- (2) To model this database to cater for language family-specific grammatical features.
- (3) To ensure the database can simplify this data into a format that is acceptable and uploadable to Wikidata.

3.2.2 Research and implementation of a more accessible Wikidata interface. The main aim is to design 2 gamified interfaces that the average user can use and this will expand the pool of contributors, meaning more lexicographical data can be collected. The interface will also collect data specific to Niger-Congo B languages such as noun classes.

The research question that will be investigated is what specific gamification elements (from 2 different gamified interfaces) can be used to motivate users to contribute lexicographical data for

Niger-Congo B languages, and how effective are they compared to the current interface for collecting this data for Wikidata?

The goal specifics aims of this section are:

- (1) To research what should be included on the 2 gamified interfaces for user engagement.
- (2) To develop the above interfaces into web applications.
- (3) To test it with participants to see its effectiveness and how it compares to the current Wikidata interface.

4 RELATED WORKS

In this section, we explore related works in lexicographical databases and resources, as well as gamified interfaces for collecting lexicographical data. Firstly, we will look at several studies that discuss the creation and maintenance of lexicographical databases, including multilingual databases and wordnets. Secondly, we will also examine the use of GWAPs for collecting lexicographical data, highlighting the importance of participant motivation and validation techniques.

4.1 Lexicographical Databases and Similar Resources

Lexicographic data is used to describe and define lexemes [3], which are the basic units of meaning in a language [4]. This data is essential for the Abstract Wikipedia project, and lexicographic databases are necessary to store it.

Several works discuss lexicographical databases. Bergenholtz et al. [5] propose four essential components: a headword or lemma, sense or meaning, grammatical information, and illustrative examples. Fuertes-Olivera's [6] study highlights the advantages of electronic databases over traditional print dictionaries and discusses the challenges of creating and maintaining them. Corpus-based lexicography, which analyses large corpora to identify language usage patterns, is also explored [6].

Multilingual databases are crucial for translation and cross-linguistic communication in projects like Abstract Wikipedia. Fuentes et al. [6] emphasise the importance of considering translation quality, linguistic nuances, and cultural nuances when creating a multilingual lexicographical database.

Lexicographical resources, such as wordnets, have become popular in natural language processing due to their graph structure, which is easier for computers to understand. The Princeton WordNet served as a template for similar projects, maintaining the hierarchical structure and semantic relations while translating the content into the target language.

Wikidata serves as a structured data counterpart to Wikipedia, allowing users to collaboratively edit a knowledge graph. It features special pages for lexicographic data with a new namespace for lexemes. With over 668 languages represented, Wikidata is an extensive resource, but its coverage of LRLs is limited. Morshed et al. [7] proposed a method to represent syntactic dependencies within Wikidata lexicographical data, using a compact format applicable to different dependency grammars. Despite requiring modifications for specific syntactic structures, this representation allows for improved modeling of multi-part elements and their use in syntactic parsing of other texts.

4.2 Interfaces for Collecting Lexicographical Data

There are numerous projects that gather lexicographical data using crowdsourcing through a GWAP (game with a purpose) format. Crowdsourcing involves dividing lexicographical data collection into smaller tasks (called microtasks), allowing it to be done by non-experts. This expands the pool of who can contribute as there are very few experts around, especially for LRLs [8]. GWAPs use entertainment (by making data collection into a game) to keep users engaged while collecting lexical data in the background. GWAPs use validation techniques such as gold standard (checking annotator input against sample expert input) and inter-annotator agreement (multiple annotators agree on the same input) to check the accuracy of inputted data. Most GWAPs primarily rely on social (e.g. leaderboards, rankings and other competitive elements) and psychological (e.g. altruism) motivation [10]. Maintaining long-term contributor motivation is crucial, especially for LRLs with limited contributor pools, so participant motivation is important for a project's long term success [8]. A study on a GWAP for collecting isiXhosa lexicographical data [9], found that economic motivation strongly drove motivation and without it, there was little participation from users. This was the only study done in South African context and it should be taken into account since this project will be done in the same context. Another complementary issue to this is attracting users to first play the game.

Some notable GWAPs for collecting lexicographical data for African languages are the Kamusi Project [13] and the abovementioned GWAP for isiXhosa lexicographical data collection [9]. There are also projects focussed on collecting data for Wikidata such as MachtSinn [11] and the Distributed Game [12]. None of these projects took into consideration agglutinative languages, which will need to be addressed since Niger-Congo B languages are agglutinative.

Developing a new gamified interface requires carrying the concepts discussed above over successfully while also taking agglutinative language structures and participant motivation (along with attraction) into account to ensure a successful project.

5 METHODOLOGY

5.1 Lexicographical Database

The database will include all relevant grammatical features for the Niger-Congo B languages, including headwords or lemmas, meanings, grammatical information, and illustrative examples, as proposed by Bergenholtz et al. [5]. This database will facilitate efficient collection and batch uploading of lexicographical data to Wikidata. The development will largely follow a prototype model with functionality added sequentially, in iterations.

5.1.1 Development of a Lexicographic Database. To achieve this, the following steps will be undertaken:

- (1) Conduct a review of existing lexicographical databases, focusing on their structure, features, and limitations to identify best practices and potential improvements.
- (2) Analyse the data format requirements and specifications for Wikidata, focusing on lexicographical data.

- (3) Create a conceptual model, incorporating the specific linguistic features and requirements of the Niger-Congo B languages to ensure accurate representation.
- (4) Select an appropriate database management system (DBMS) that for storing the lexicographical data.
- (5) Implement and test the database structure, focusing on data integrity, security, and scalability to accommodate for potential growth.

5.1.2 Modeling the Database to Reflect Niger-Congo B Languages. Collaboration with native speakers is essential to accurately represent the Niger-Congo B languages. This involves gathering detailed information on their structure, syntax, and semantics, as well as identifying specific language features such as irregular verb forms, true and relative adjectives, tonal markings, and idiomatic expressions. A flexible data model must be developed to accommodate the variations and complexities inherent in these languages. Finally, testing the database model using sample data from Associate Professor Keet's data store and the SADiLar corpus is necessary to ensure linguistic accuracy.

5.1.3 Modifying Wikidata to Allow for Upload from the database. To enable the upload of lexicographical data on Wikidata, modifications to its data model and API can be made. First, the data model would need to be extended, potentially by creating a new data type or extending existing ones, to support and represent the Niger-Congo B data within the Wikidata ecosystem. Next, a new API endpoint specifically designed for data uploads would be created, accepting various formats such as CSV or JSON, and parsing it into the extended Wikidata data model. The data import process would be adjusted to accommodate the lexicographic data, including the implementation of preprocessing, validation, and normalization steps to ensure data consistency and quality.

5.2 Gamified Interfaces

Developing two gamified interfaces, for collecting lexicographical data, will require the following steps:

5.2.1 Interface Design. Firstly, a review of existing gamified interfaces and using them as a reference points for the new interfaces will be done. Then we will define the game layouts incorporating a format for agglutinative language structures as well as what game elements to include to motivate participants. We will create prototypes for both interfaces which will be subject to user testing by 3-5 test users each to identify improvements before the final versions are developed.

5.2.2 Development of Game. We will then develop 2 websites using feedback identified from last stage. JavaScript, SQL and PHP will mainly be used and validation techniques will be coded for verifying user input. The websites will be linked to server to upload and store lexicographical data. From this, the valid lexemes will be uploaded to Wikidata with SPARQL.

5.2.3 Run Study to Test Effectiveness of Game.

- (1) We will design how the study will be conducted and what data will be collected. The primary data collected will be the amount of lexicographical data inputted by users and user engagement metrics (such as session duration, frequency of

use, input time per a lexeme and number of active users). For the study design, details such as how monetary prizes will be used to motivate participants, exact dates for running study, how many experiments will there be (and how will they differ) etc will be specified.

- (2) We will advertise the games to get participants at UCT and elsewhere to play the game. Users will play the games over a week.
- (3) We will measure how effective our interfaces are in terms of collecting lexicographical data during and after study and draw conclusions from there.

6 REQUIRED RESOURCES

6.1 For Lexicographical Database

Our project will need native speakers of Niger-Congo B languages to deliver linguistic input and verify the accuracy of the data stored. We will employ database management system software (DBMS) to effectively handle the lexicographical data. In addition, we will utilize Wikidata's API and documentation to ensure compatibility and facilitate the data upload process. To further enhance our project, we require access to pertinent linguistic resources and research. This includes lexicographical databases provided by the university in-house, as well as corpora sourced from the SADiLaR repository, a digital language resource center.

6.2 Research and Implementation of a new Wikidata Interface

To ensure the success of our project, we will engage 3-5 users to provide feedback in order to refine the initial interfaces. Moreover, we will require a minimum of 30 participants who speak the Niger-Congo B language for the study, sourced from UCT and the researcher. In terms of technical resources, we will secure domain names for the websites and obtain a virtual server from the department to store data while the study is being conducted. As for monetary resources, we will allocate the first R2000 from the research budget for cash prizes, while additional funding and prizes will be provided by the researcher.

7 EVALUATION

7.1 Database Criteria

To assess the effectiveness of the database, the following evaluation criteria will be applied:

7.1.1 Database Robustness and Stability. To maintain data integrity, we will select a representative sample of data from the database and manually introduce errors, such as wrong data types, incomplete data, or incorrect data. We will then run a set of queries and reports that depend on the modified data, comparing the results with the expected outcomes. If discrepancies arise, we will identify the cause of the error and rectify it. For security, we will assess the database's resistance to unauthorized access, data breaches, and other security threats. To ensure scalability, we will generate a representative workload for the database, encompassing both read and write operations. Utilizing load testing tools, we will simulate

increasing levels of user and data activity. By measuring the database's performance metrics, such as response time and throughput under each load level, we will identify bottlenecks and limitations, including CPU, memory, or disk I/O constraints. Finally, we will implement and test changes to address these bottlenecks and limitations, potentially involving the addition of more resources or optimization of queries.

7.1.2 Linguistic Representation. To ensure completeness, we will identify a comprehensive list of relevant linguistic features for the Niger-Congo B languages and verify that the database includes all these identified features. We will review a representative sample of the data to confirm that all the identified linguistic features are accurately and completely represented. By comparing the database's coverage of linguistic features to existing linguistic resources for the Niger-Congo B languages, we will analyze the database's ability to handle new linguistic features that may emerge in the future. In terms of flexibility, we will examine the database's capacity to adapt to variations and complexities in the Nguni-B languages. To assess accuracy, we will evaluate the correctness of the data and its representation in the database through comparisons with existing resources.

7.1.3 Data Compatibility with Wikidata. To evaluate data format compatibility, we will assess the accuracy of the transformed data in meeting Wikidata's format requirements for lexicographical data. In terms of export efficiency, we will examine the ease and speed of exporting selected datasets in a Wikidata-compatible format. Finally, we will assess the upload success by examining the success rate of batch uploading the transformed data to Wikidata, ensuring no errors or data loss occur during the upload process.

By evaluating the project using these criteria, it will be possible to measure the success of the database and its ability to collect and batch upload data to Wikidata.

7.2 Interface Criteria

The primary criteria of the gamified interfaces will answer the research question while the secondary criterion that will be assessed are the quality of contributions and adaptability to agglutinative language structures.

7.2.1 Primary Criteria. The amount of lexicographical data collected for both interfaces will be compared to the amount of lexicographical data on Wikidata for the Niger-Congo B language the interfaces focus on. User engagement (which will show how motivated users are to contribute) will be measured with metrics such as the number of active users, session duration, and frequency of contributions for both interfaces. This will help identify which gamification elements are effective.

7.2.2 Secondary Criteria. Quality of contributions will be assessed by comparing the quality of inputted data to accurate data. Adaptability to agglutinative language structures will be evaluated by looking at the interface's ability to handle agglutinative language structures.

8 ANTICIPATED OUTCOMES

Based on the above, the outcomes will be discussed. Both the database and interface components will help with collecting lexicographical data for Niger-Congo B languages meaning more articles can be generated in these languages. This will reduce the knowledge gap for LRLs and help achieve Wikipedia's goal of knowledge sharing for all.

The database component of the project will create a comprehensive lexicographical database for Niger-Congo B languages, resulting in improved access to this data for researchers, language learners, and the general public. The lexicographical data will be batch uploaded to Wikidata, making it available to a wider audience. Since lexicographical standards will be followed and expert reviews will be used, this will result in a high-quality language resource.

For the gamified interface component of the project, its main outcome is that it will increase the pool of contributors meaning more lexicographical data can be collected resulting in more articles generated in the Niger-Congo B language the study is conducted in. This will also help pave the way for similar approaches in the future for collecting lexicographical data for LRL's.

8.1 Key Success Factors

- Develop a stable database for Niger-Congo B languages that can store and manage lexicographical data, including language family-specific grammatical features. The success of the database will be determined by its ability to seamlessly collect and upload lexicographical data to Wikidata.
- The ability to convert and batch upload the collected lexicographical data to Wikidata in a compatible format without loss of information or integrity.
- Collecting lexicographical data for a Niger-Congo B language through the gamified interfaces. This will address the under-representation of Niger-Congo B languages in Wikidata.
- Identifying which gamification elements from the gamified interfaces were effective and increases use of the interfaces.

9 PROJECT PLAN

9.1 Goals

The goals are to develop a lexicographical database modeled after Niger-Congo B languages for data collection and batch uploading to Wikidata and to research and implement more engaging interfaces for adding lexicographical data to Wikidata.

9.2 Risks

An insufficient understanding of Niger-Congo B languages' linguistic features (e.g. agglutinative structure), could lead to inaccuracies. There could also be technical challenges in developing a database capable of handling large volumes of data. Another risk comes with the difficulty in simplifying data to be compatible with Wikidata's format requirements, which might lead to the database failing. An important risk for GWAPs is that they rely on user participation and often struggle to attract and retain contributors over time which means insufficient engagement from users could lead to an unsuccessful study.

9.3 Deliverables

Attached to appendix.

9.4 Timeline and Milestones

Attached to appendix.

9.5 Work Allocation

Tadiwa will handle the database development. Zahraa will be the interface component.

10 CONCLUSION

In conclusion, this project will achieve two primary goals: (1) to research and implement gamified interfaces for adding lexicographical data to Wikidata, and (2) to develop a separate lexicographical database modeled after Niger-Congo languages, where data can be collected and batch uploaded to Wikidata. This project addresses the lack of lexicographical data for Niger-Congo B languages.

Future work may involve expanding the database to accommodate other LRLs and exploring ways to further improve the user experience when adding lexicographical data to Wikidata or how this approach can be used for other LRLs.

11 ETHICAL, PROFESSIONAL AND LEGAL ISSUES

11.1 For lexicographical database

11.1.1 Ethical issues. Collaborating with native speakers requires obtaining their informed consent, ensuring that they are fully aware of the project's objectives. All participants will be informed about the purpose of their involvement and their rights, including the right to withdraw. The project will acknowledge the contributions of native speakers and others involved ensuring that they receive proper credit.

11.1.2 Professional and Legal Issues. The project will respect the intellectual property rights of others and any third-party resources or data sets used will be properly licensed and attributed. To promote the accessibility, an open-source license will be applied, allowing others to use the project's work.

11.2 Gamified interfaces

11.2.1 Ethical issues. Ethical clearance must be obtained from University of Cape Town Research Faculty of Science ethics committee before the study is conducted. Informed consent will be required from all participants by giving them full information about the study and giving them the right to withdraw.

11.2.2 Professional and legal issues. User data will be collected through the gamified interfaces and this data must be protected and kept confidential while following the Protection of Personal Information Act. When developing the web applications, no resources that violate copyright agreements will be used.

REFERENCES

- [1] Denny Vrandečić. 2021. Building a multilingual Wikipedia. *Communications of the ACM* 64, 4 (2021), 38–41. DOI: <https://doi.org/10.1145/3442337>.
- [2] Bergenholtz, H. and Gouws, R.H. 2012. What is Lexicography?. *Lexikos*. 22, 1 (Nov. 2012). DOI: <https://doi.org/10.5788/22-1-996>.

- [3] Nielsen, F. (2020). Lexemes in Wikidata: 2020 status. In *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)* (pp. 82-86). European Language Resources Association.
- [4] Wikidata. (n.d.). Wikidata: Lexicographical data documentation. Retrieved March 14, 2023, from https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/Documentation.
- [5] Bergenholtz, H. and Nielsen, J.S. 2013. What is a lexicographical database? *Lexikos* 23 (2013), 77–87.
- [6] Paola Anna Fuertes-Olivera and Henning Bergenholtz (Eds.). *e-Lexicography: the internet, digital initiatives and lexicography*. A&C Black, October 20, 2011.
- [7] Morshed, M. (2021). Modeling Syntactic Dependency Relationships in Wikidata Lexicographical Data. In *Proceedings of the Wikidata@ISWC Workshop*.
- [8] Martin Benjamin and Paula Radetzky. "Multilingual Lexicography with a Focus on Less-Resourced Languages: Data Mining." In *Expert Input, Crowdsourcing, and Gamification Acquiring Lexical Data for LRLs. 9th edition of the Language Resources and Evaluation Conference*, 2014. <https://infoscience.epfl.ch/record/200375>.
- [9] Sean Packham and Hussein Suleman. "Crowdsourcing a Text Corpus is not a Game." In *Digital Libraries: Providing Quality Information: 17th International Conference on Asia-Pacific Digital Libraries, ICADL 2015, December 9-12, Proceedings 16*, pages 225–234, Seoul, Korea, 2015. Springer. https://doi.org/10.1007/978-3-319-27974-9_23. doi:10.1007/978-3-319-27974-923.
- [10] Jaka Čibej, Darja Fišer, and Iztok Kosem. "The role of crowdsourcing in lexicography." *Electronic lexicography in the 21st Century: linking lexical data in the digital age. Proceedings of the eLex*, volume 2015, pages 70–83, 2015. https://elex.link/elex2015/proceedings/eLex_2015_05_Cibej+Fiser+Kosem.pdf.
- [11] "MachtSinn – Das macht doch alles keinen Sinn!" <https://machtsinn.toolforge.org/>, 2023.
- [12] "Wikidata – The Distributed Game." <https://wikidata-game.toolforge.org/distributed/#mode=stats>, 2023.
- [13] Martin Benjamin. "Crowdsourcing microdata for cost-effective and reliable lexicography." In *Proceedings of the 9th ASIALEX Conference*, Hong Kong, 2016.

12 APPENDIX

12.1

Deliverables

For the Database

20 March	Complete comprehensive review of existing lexicographical databases
20 May	Develop data transformation module that converts lexicographical data into a Wikidata-compatible format
10 July	Implement export functionality for datasets in a Wikidata-compatible format
15 September	Have lexicographical database modelled after Niger-Congo B languages Report also completed

For the Interfaces

20 March	Comprehensive review of GWAPs (Games With A Purpose)
12 July	Develop the code for the gamified interfaces
15 September	Report completed

12.2

Milestones

For the Database

25 April	Create a conceptual database model for Niger-Congo B
1st May	Evaluate and compare the different DBMS's and how they can be used to model Niger-Congo B languages
5 May	Submit final project proposal
3 May	Select an appropriate database management system
15 May	Implement and test the database structure
25 May	Complete initial database prototype
31 May.	Analyse data format requirements for Wikidata

8 June	Modelling the database to reflect Niger-Congo B languages-Gather detailed information on language structure, syntax, and semantics
13 June	Develop a flexible data model
15 June	Test the database model using sample data from Associate Professor Keet's data store
30 July	Simplifying data for upload to Wikidata-Develop a data transformation module
4 August	Implement an export functionality
15 August	Test data transformation and export functionalities
28 August	Complete draft of final report paper

For the Interfaces

12 May	Research of a gamified interfaces and defining game layout
12 May	Ethics applications
2 June	Create prototypes
7 June	Conduct user testing and gather feedback
12 July	Development of 2 gamified interface web applications (using feedback from user testing)
20 July	Link the websites to the server for uploading and storing lexicographical data
20 July	Implement functionality to upload valid lexemes to Wikidata with SPARQL
20 July	Design the experiment
27 July	Advertise the games and recruit participants
1 August	Run the experiment over 1 week (during this time, Zahraa will work on draft of research project)
7 August	Analyse the results of the experiment and see if goal met
28 August	Submit draft paper
11 September	Submit final paper
15 September	Submit code

For both

21 July	Progress demonstration
26 September	Final project demonstration
9 October	Poster made
16 October	Website
24 October	School of IT Showcase

12.3

Gantt Chart

Timeline

