Indonesian Wikidata User Research Report — Newcomers

Kenny Tjahjadi
26 September 2022

01   Background

Newcomers tell the story for us.

The Wikidata Software Collaboration team does not yet have an idea of how Wikidata Lexeme is perceived by newcomers (<1 year of experience editing Wikidata).

Therefore, this research was conducted as a way to determine what the newcomers in Indonesia thought of Wikidata Lexeme.

Research objectives

- Understanding the thought process and user journey of new Wikidata users when searching, adding, and editing lexicographical data on Wikidata

- Exploring issues that occur while interacting with lexicographical data projects on Wikidata and listens to suggestions from the user's perspective

- Finding out the condition of lexicographical data community and what they perceived about Wikidata software collaboration project

Participant Demographics

- Participants aged 18+

- Participants have been editing Wikidata for ≤ 1 year

- Participants are Indonesian speakers

- Participants join local Wikimedia community/have attended Wikidata events

Research method

Semi-structured interviews, each conducted for 60-90 minutes online on Zoom. All participants turn on their cameras and microphones.

Participants will be asked to do a screen share if the demonstration of using the feature needs to be carried out based on the initiative of the participant or as an instruction from the moderator.

The interviews focused on exploring the problems encountered when searching, understanding and editing lexicographical data on Wikidata.

To maintain privacy, research videos and participant data are anonymized.

Research hypothesis

- Newcomers don't know much about Wikidata, especially Wikidata Lexeme

- Newcomers cannot find information about Wikidata Lexeme

- Newcomers cannot search for specific lexemes from the Wikidata site

- Newcomers can add new lexemes to Wikidata

- Newcomers are not comfortable with the lexicographical data display on Wikidata

- Newcomers often edit Wikidata with their computer/laptop at night/holidays

- Newcomers feel the most need to improve is the look and friendliness of the site's user experience

02

Research Activities

Searching for a lexeme, adding a new lexeme

Looking for a specific lexeme

Present condiition

In order to search for Wikidata Lexeme, we have to add *the string* "L:" or "Lexeme:" before adding a lexeme, using advanced search filters, or using a tool like Ordia (ordia.toolforge.org).

However, this is not understood by newcomers so they cannot search for certain lexemes from the Wikidata site.

What was found

1. 4 out of 5 participants could not find the given lexeme
2. 5 out of 5 participants feel that Wikidata search results need to be grouped by namespaces into items and lexemes
3. 5 out of 5 participants said that lexeme results should be included in Wikidata search suggestions

Average SEQ: 1.8 / 7

The average SEQ results in various tests is 5.5. Source

Breakdown: search for lexemes

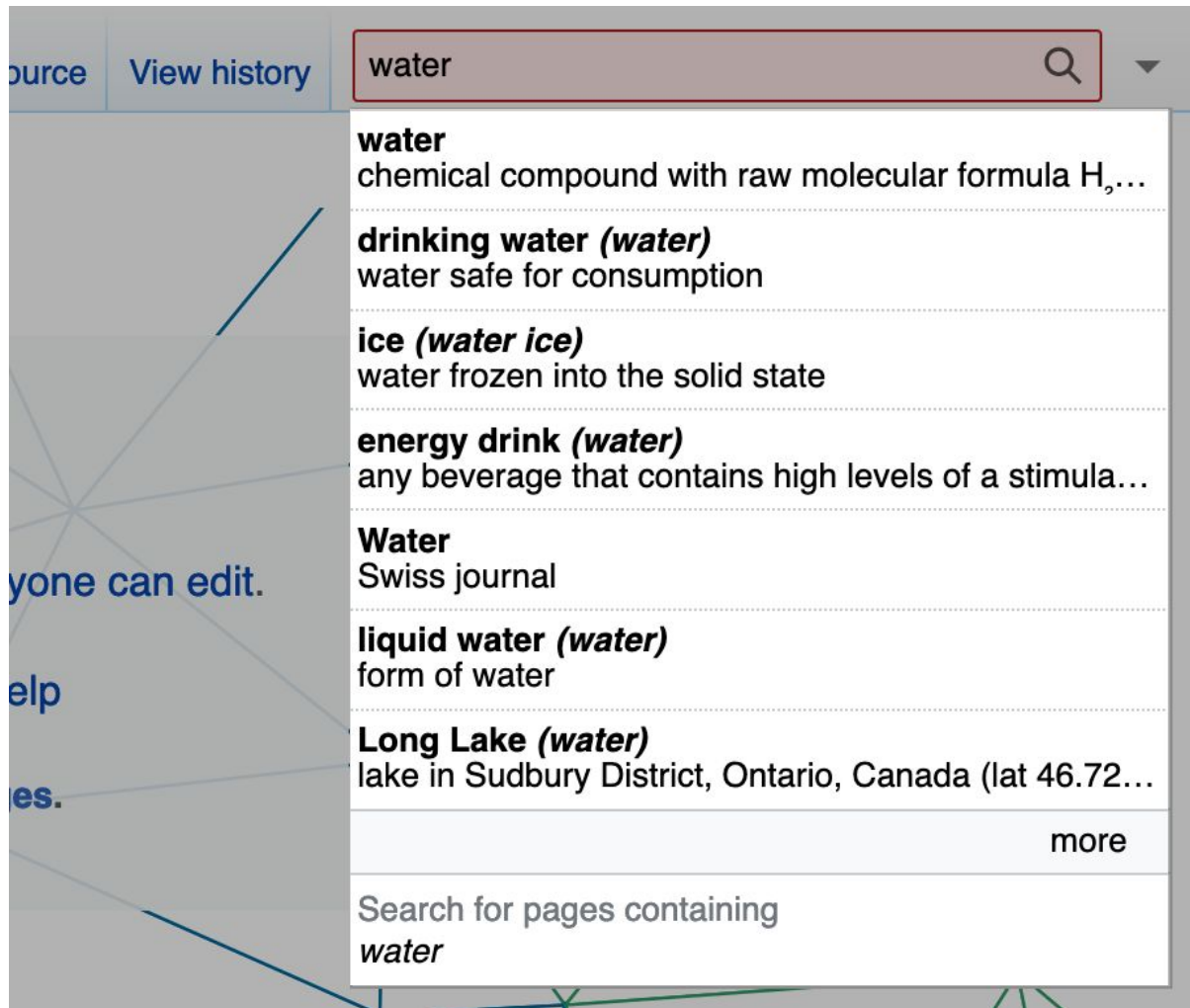| Activity	Search through the main page	Searching through random lexemes
Journey	1. User opens Wikidata site 2. Users search by entering lexemes directly 3. Users vote for Wikidata recommendation results 4. What is obtained is the result of the item	1. The user opens a random lexeme page 2. The user searches again in the search box, expecting to show only lexeme results 3. Users vote for Wikidata recommendation results 4. What you get is still the results *4 out of 5 users give up at this stage.*
Feeling	● **Frustrated** because I can't find lexemes, so I have to find another way	● **Frustration** because you still haven't found the lexeme you're looking for
User suggestions	● Adding lexeme results to search suggestions ● Added filter / search categorization by type	● Created a *scoped search* (search in lexemes/search in items/all)

The participant could not find the given lexeme in the search results.

All search results are entries in the Wikidata item namespace.

Breakdown: search for lexemes

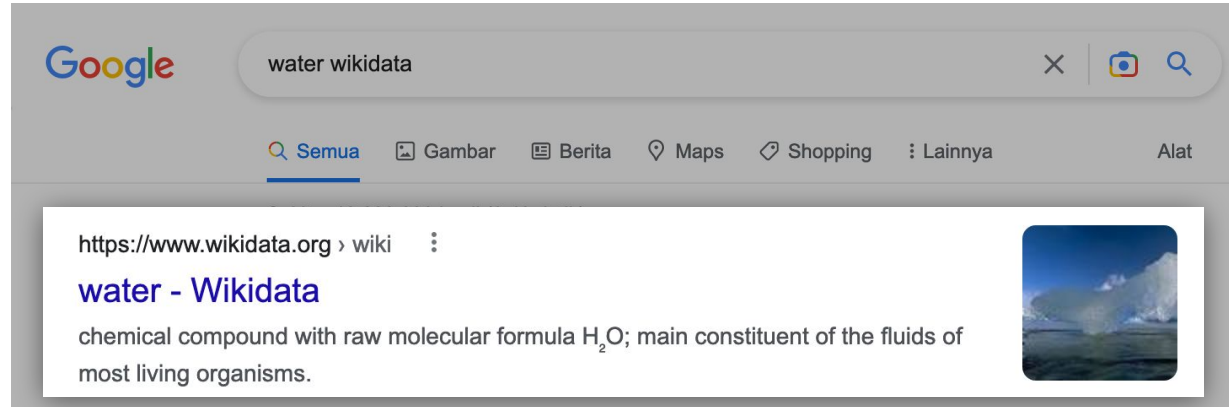| Activity	Searching via external search engine	Search again by adding "Lexeme:"
Journey	1. Users take the initiative to search for it via Google "water wikidata" 2. User selects the top search result 3. What was found was the result of *Only done by 1 of 5 participants.*	1. User returns to Wikidata main page 2. User views Lexeme's namespace model 3. The user tries to enter search results by increasing the namespace 4. Lexeme results found *Only done by 1 of 5 participants.*
Feeling	● **Frustrated** that lexemes don't appear, even after using Google	● **Annoyed** at not being able to see search suggestions like items ● **relief** to be able to find a lexeme
User suggestions	● Adding lexeme results to search suggestions ● Added filter / search categorization by type	● It shouldn't be this hard to find lexemes, no need to add namespace codes

Search results on external websites display Wikidata item results, not lexemes.

SEO for Wikidata Lexemes needs to be considered.

Google

water wikidata

Q Semua   Gambar   Berita   Maps   Shopping   Lainnya   Alat

https://www.wikidata.org › wiki

water - Wikidata

chemical compound with raw molecular formula H_2O; main constituent of the fluids of most living organisms.

Adding a new lexeme

Present condiition

The addition of new lexemes is available via the side menu in the desktop Wikidata view. Contents that need to be added initially are entries, spelling variations of the entry (optional, if non-Latin characters), language, and lexical categories.

The assumption is that newcomers can add new lexemes to Wikidata, although they may have additional suggestions.

What was found

1. 5 out of 5 participants can add new lexemes
2. 5 out of 5 participants feel that they need a guide to terms and examples on the display adding a new lexeme → not familiar with *placeholder text*
3. 1 out of 5 participants chose the wrong language
4. 1 out of 5 participants chose the wrong lexical category

Average SEQ: 5.6/7

The average SEQ results in various tests is 5.5. <u>Source</u>

Breakdown: add lexeme

Activity	Open the add new lexeme page (desktop)	Open the add new lexeme page (mobile)
Journey	1. User opens Wikidata home page 2. The user gets the add link in the sidebar and opens it	1. The user tries to search the hamburger menu (3 strips), can't find it 2. *desktop* 3. The user gets the add link in the sidebar and opens it
Feeling	● **Glad** that the link is easy to find	● **Annoyed** that can't find the feature in the hamburger menu ● **Frustrating to be** forced to use *desktop mode* to find this feature
User suggestions	*There isn't any*	● Added a way to access the add new lexeme page in the *mobile view*

Breakdown: add lexeme

Activity	Fill in new lexeme data (desktop)	Fill in new lexeme data (mobile)
Journey	1. User fills in data entry 2. User searches and selects language 3. Users search and select lexicographic categories 4. The user presses the "Submit" button	1. User fills in data entry 2. User searches and selects language 3. Users search and select lexicographic categories 4. The user presses the "Submit" button
Feeling	● **Confused** by the term used because it is only used by linguists	● **Annoyed** that the autocorrect and auto zoom system when filling in the fields, made the wrong choice of the suggested item ● **Surprised and happy** to be able to enter non-Latin characters
User suggestions	● Added a guide next to the lexeme view (seems to have been tried in the *opt-in test* before) ● Added an option to learn about lexemes in the side menu	● Users feel the display should be optimized for *mobile* ● Options for language and lexical categories should only display relevant data

Participants were not familiar with the lexicographic terms used, such as lemma and lexical category because they are rarely used, even in linguistics.

Create a new Lexeme

You are about to create a Lexeme (such as first (L2)) to store lexicographical data (e.g. language, etymology, inflections, etc.) about a word or phrase, **not general data about the entity or concept** to which it refers. If you want to store general data (e.g. date of birth, opening date, author, image, country, coordinates, website, etc.), you need to **create an Item** instead.

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

Create a new Lexeme

Lemma

Lemma is usually the shortest form of the word   *

Spelling variant of the Lemma

Lexeme's language

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

Lexical category

The Lexeme's category, e.g. 'numeral'   *

Create

03 Insights gained

Insights gained

Theme	Insight
Community nowadays	4 out of 5 active participants took part in community events held by local chapter or from WMID. They feel happy to meet new people, learn new knowledge, and compete in a healthy manner. The majority know about the event through social media, especially Instagram with its attractive visuals.
Learning curve	Participants found it difficult with the perception of the need to understand technology and language. Participants also unable to search for lexemes through the search box on Wikidata. They are also confused about adding new lexemes and editing lexemes.
Dialects and characters	Participants were still confused about the clarity of how to enter dialects and languages with non-Latin characters when adding new lexemes.

Insights gained

Theme	Insight
Example and exposure	Participants felt that Wikidata Lexeme should be given a separate page from Wikidata and present featured lexemes that can be used as examples. Currently, items and lexemes in Wikidata are unrelated so users don't know the relationship between items and lexemes.
Duplicate contributions	Participants want to contribute, but don't know how to check for duplicates in the lexeme. This causes them to be afraid of damaging the database and eventually not progressing from just casual editors.
Device usage	3 out of 5 participants have contributed with a mobile device. Wikidata, especially Lexeme, is not optimized for them.
Comparison with other projects	Some participants compared the appearance of Wikidata Lexeme with Wiktionary, Wikisource, and Commons.

Insights gained

Theme	Insight
WMID workshop is not optimal	So far, the WMID workshop is not optimal because it does not use a language that is easier to understand, and tend to use exact code (P-id or Q-id) to refer to attributes or values.
Localization	The appearance of Wikidata Lexeme has not been fully translated into Indonesian or regional languages.
Data source	Participants found it difficult to know what types of data could be entered into Wikidata Lexeme. In the typical workshop by WMID, the data to be entered has been provided and previously processed by the committee. The participant want to know how to find it and process it themselves.
Lack of lexeme socialization	WMID has not done much socialization of the Lexeme project. Participants want information about the lexeme, how to contribute to it, and how it relates to Wiktionary and similar wiki projects.

Theme 1: Community nowadays

Usually, newcomers know about Wikimedia and its projects from a friend's invitation. From there, they took the initiative to find information about the event through social media, especially Instagram with attractive visuals. They feel that WMID's events so far have been good and can be imitated.

So far, they have tentatively edited their own. However, when the local community invites them to participate, they will edit together.

"At that time, I tried to search on IG (Instagram) after being told by a friend, then I found out about wiki events."

Theme 1: Community nowadays

When they enter the community, they feel happy to meet new people, learn new knowledge, and compete in a healthy manner.

The most memorable events are competitive events, such as the Datathon and the Data entry Competition on National Museum Day because they have attractive prizes and a sense of achievement. This can be used as a consideration for gamification.

To date, there is no specific approach regarding Wikidata Lexeme to the community.

Their impression of the Lexeme project is that of a "dictionary", and they question how it differs from Wiktionary. However, they felt motivated that the Lexeme project could be a 'complete, structured and flexible dictionary'. Examples of its use include completing Wikipedia, creating a dictionary, becoming a database for language learning and language processing by computers.

Theme 2: Learning curve

Participants have the perception that to contribute to Wikidata requires an understanding of technology and language. You need a way to convince them to start editing.

For the layman, viewing Wikidata and Lexeme is not difficult, but it can be said that they need some time to get used to the look and terms.

"The Wikidata interface is easy to use."

"But, currently there are many steps that must be remembered, I have to go back and forth to match the codes (statements)."

Theme 2: Learning curve (continued)

Participants felt confused in making edits due to the unfamiliar interface and terms used.

The lexeme editing interface was considered more difficult than the item data editing interface because the participants were already familiar with the item display.

The terms used such as Forms, Senses, Statement are not commonly used, even in the literature though.
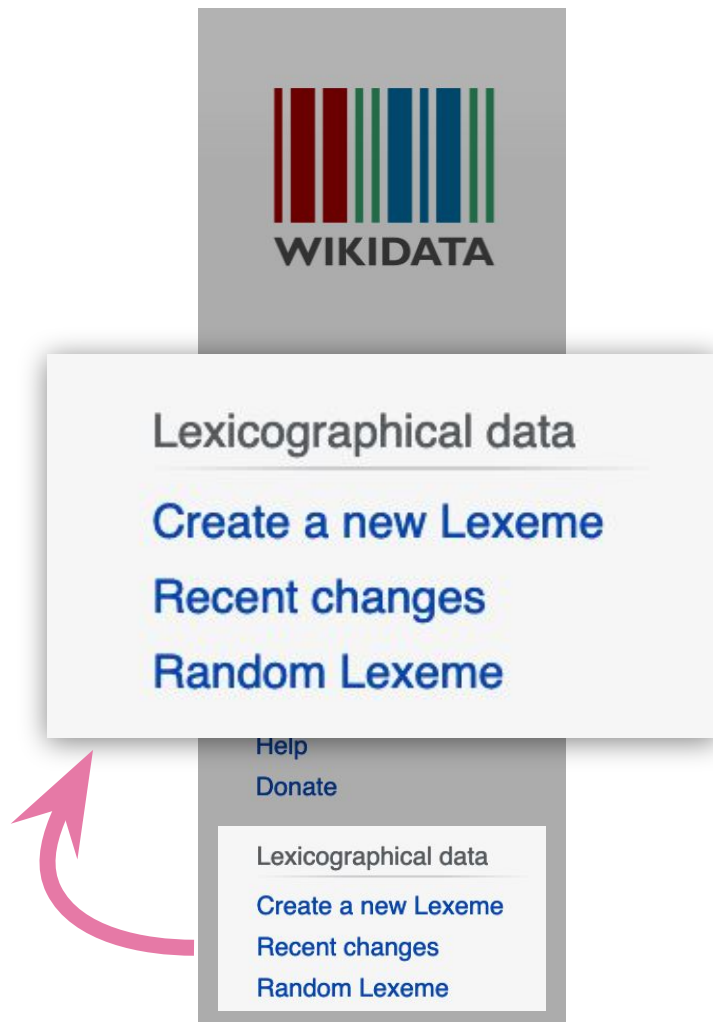
In addition, information about the Lexeme project and leksem in general is not on the Main Page or as a link in the sidebar.

Participants suggested moving the illustration of the lexeme up so that the lexeme display was easier to understand regardless of the language barrier.

Participants expect to find information about lexicographical data in the side menu.

Currently, the lexicographical data section only displays three functions: create a new lexeme, recent changes, and random lexeme.

WIKIDATA

Lexicographical data

Create a new Lexeme

Recent changes

Random Lexeme

Help
Donate

Lexicographical data

Create a new Lexeme
Recent changes
Random Lexeme

Theme 2: Learning curve (continued)

Participants were unable to search for lexemes through the search box on Wikidata.

Wikidata has provided this feature by adding the prefix "L:" or "Lexeme:", and only 1 participant can do it. The participant also admitted that it was not intuitive. In fact, there were participants who felt it necessary to *query* with SPARQL, which is clearly not friendly for new users.

Participants also could not get the results when using search engines such as Google, because the lexeme pages were not SEO optimized.

Participants suggest the following solutions:

1. Categorization of search results in both search suggestions and search results pages
2. Search with selectable namespace function
3. Special lexeme search page that is accessed via the *sidebar menu*

Theme 2: Learning curve (continued)

The participants also asked about the existing terms when adding lexemes, such as lemmas and lexical categories, because these terms were also not easy to understand.

Participants suggest the following solutions:

1. Giving tutorials (which are already on the Wikidata site but are hidden)
2. Guide next to the add lexeme form
3. External guide in PDF or presentation form

Participants also felt intimidated by the number of bots on Wikidata. Because bots are so active, they can't see what humans are actually doing inside Wikidata.

This can be used as the basis for creating filters to hide bot activity in the user's view.

Theme 3: Dialects and characters

Participants were still confused about the clarity of how to enter dialects and languages with non-Latin characters when adding new lexemes.

For example, there is a case where the dialect is read differently but the language remains the same as Javanese.

In addition, there are languages that have a level of politeness that Wikidata has not accommodated.

Participants were still unsure whether to add non-Latin characters when adding new lexemes.

Wikidata Lexeme can accommodate non-Latin characters, but they need to be convinced and told how to install tools to do this.

Theme 4: Example and exposure

Participants felt that Wikidata Lexeme should be given a separate page from Wikidata. At the very least, pay more attention to this project by adding useful links to explain this project and the lexeme in general.

They also want Wikidata Lexeme to present good lexemes and serve as examples that can serve as good editing references (see: featured articles on Wikipedia)

In addition, items and lexemes are not currently linked. Thus, users who edit items are not aware that many items about concepts or objects in Wikidata can actually be included as well as lexemes.

Theme 5: Duplicate contributions

Participants want to contribute, but don't know how to check for duplicates in the lexeme. This causes them to be afraid of damaging the database and eventually not develop from just casual editors.

In fact, when WMID holds events such as Datathon, a lot of duplicate data is entered by participants and must be tidied up by the committee.

Participants want to be notified of duplicate data when adding lexemes.

The form can be like a search suggestion or confirmation if there is an exact match of lemma, language, and lexical category.

Theme 6: Device usage

3 out of 5 participants have contributed with a mobile device. Wikidata, especially Lexeme, is not optimized for them.

Because in the mobile display there is no sidebar, users cannot add or access other lexem features.

The display of the form is only optimized for computers, because the layout is horizontal.

The "latest changes" feature, "my contribution" button, "new item" button is not available, some values cannot be expanded.

Theme 7: Comparison with other projects

Wiktionary

The data on the regional language dictionary on Wiktionary is incomplete, and it is better to directly ask the speakers of the regional language.

Wikistories

Wikistories is a mobile-centric project , but the pictures and illustrations on Commons are lacking. This can affect the visualization of lexeme meaning in this project.
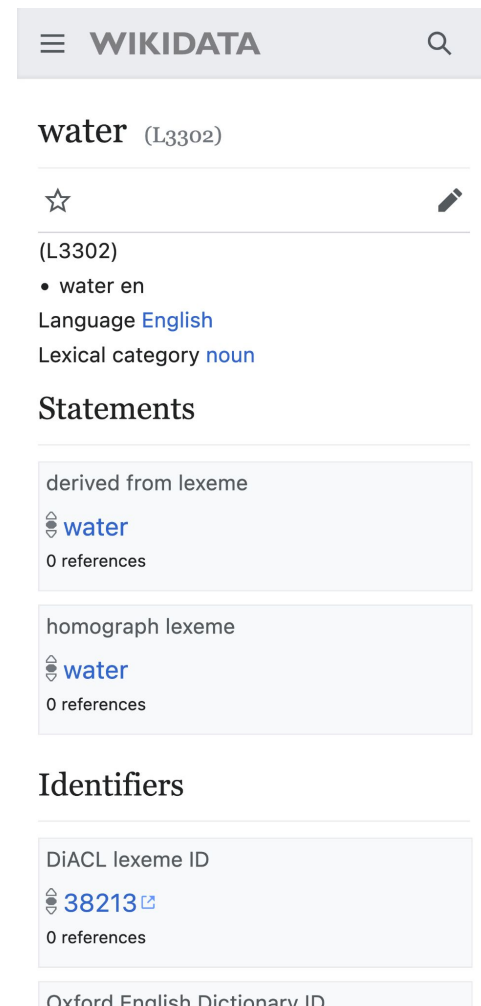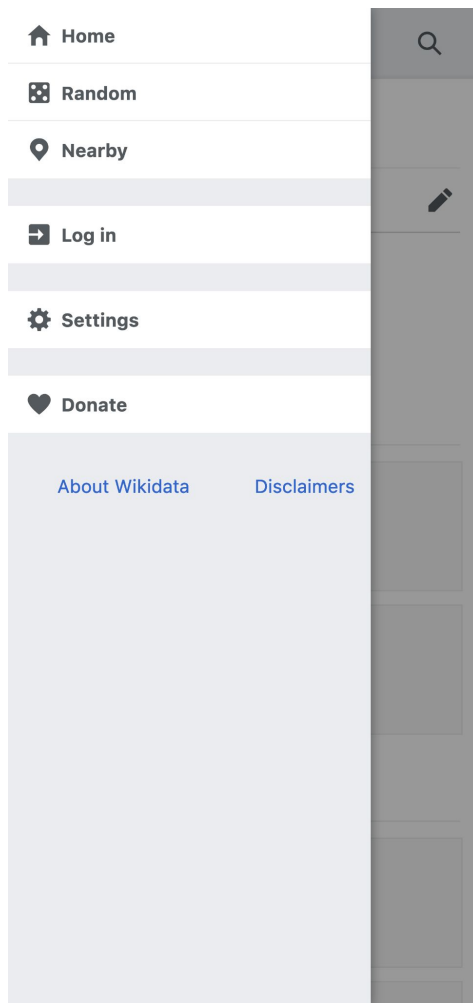
Wikisource

Viewing edits is easier because there is a special label whether a section has been worked on or not on the form.

The history display is easier because it is visually better and can display item changes more clearly.

Wikisource can display contributions by country of origin, not just the whole world which will be overwhelming.

The mobile view does not have a sidebar with the same features as the desktop and also cannot perform full editing.

Home

Random

Nearby

Log in

Settings

Donate

About Wikidata   Disclaimers

☰ WIKIDATA

water (L3302)

(L3302)

• water en
Language English
Lexical category noun

Statements

derived from lexeme
water
0 references

homograph lexeme
water
0 references

Identifiers

DiACL lexeme ID
38213
0 references

Oxford English Dictionary ID

Theme 7: Comparison with other projects (continued)

Wikimedia Commons

Participants are interested in adding data
such as sound files, pronunciation and images
that can be integrated with Lexeme.

Theme 8: Less than optimal WMID workshop

So far, the WMID workshop is not optimal because it does not use a language that is easier to understand, and tend to use exact code (P-id or Q-id) to refer to attributes or values.

In the future, participants should be given direction by explaining how the codes actually have labels that are easy to understand and eliminate the participant's perception that Wikidata = coding. This can make them intimidated so they don't want to contribute, especially for those who don't have an IT background.

Theme 9: Localization

Wikidata, especially Wikidata Lexemes properties and documentations, have not been fully translated into Indonesian and other regional languages.

This is important to do so that the terms in the lexeme can be uniformed and make it easier for the team to make guidelines for editing lexemes in their respective languages.

The translation to Indonesian has not been done completely.

There are still terms that need to be discussed with the community.

(L6546)

air

id

Language bahasa Indonesia
Lexical category nomina

Forms

L6546-F1   air
             id

Grammatical features tunggal, jamak

Senses

	Inggris	water
	Bahasa Indonesia	senyawa dengan rumus kimia H_2O yang berwujud cair pada suhu dan tekanan udara ruangan dan dapat ditemukan di sungai, danau, laut, dan saat hujan
L6546-S1	Jepang	水
	Jawa	banyu
	Madura	aèng
	Minangkabau	aia
	Sunda	cai
	Turki	su

Theme 10: Data sources

Participants found it difficult to know what types of data could be entered into Wikidata Lexeme.

So far, community workshops have not taught how to find data sources, but have been provided and processed. This gave rise to the perception that participants had to be 'fed' with data, even though some of them were actually very diligent and independent in digging and exploring data.

What can be suggested for this problem is to conduct training or create a guide on the types of data that can be used as a reference source for Wikidata Lexeme for non-IT and IT users.

Participants said that the data should be available online, because it is easier to add to the computer.

Theme 11: Lack of lexeme socialization

WMID has not done much socialization of the Lexeme project, so the participants who attended the interview did not really know about this project.

The participants wanted this event to be held online, given the limitations of their location which was quite far from the usual training venues.

This socialization is expected to be a means for WMID to hear and involve the community.

Participants want to be explained about:

1. General lexeme information
2. Lexeme Wikidata project information
3. How to contribute
4. How does this project relate to Wiktionary and similar wiki projects
5. Future benefits of this project
6. "lightweight" training and practice

Comparison with hypothesis

- Newcomers **know about Wikidata** , but **are not familiar with Wikidata Lexeme**

- Newcomers find it **difficult to find information** about Wikidata Lexeme

- Newcomers **cannot search for specific lexemes** from the Wikidata site

- Newcomers **can add new lexemes** to Wikidata

- Newcomers **are not comfortable** with the display of lexicographical data on Wikidata, especially on mobile devices

- Newcomers edit Wikidata with **laptops and mobile devices without a regular schedule**

- Newcomers feel **the most need to improve is project communication to the community** . The appearance and friendliness of the site's user experience **are not the most important.**

04 Next steps

Next steps

1

Report results to stakeholders

This user research can be a reference for the Wikidata Software Collaboration team to bring the perspectives of newcomer contributors in Indonesia to WMDE as one of the stakeholders of Wikidata Software Collaboration.

2

Add research demographics

This research only describes the demographics of the novice contributors. More experienced contributors also need to be included in the next research.

3

Test for usability heuristics

The themes that have been found and explained need to be categorized using Jakob Nielsen's 10 Usability Heuristics: https://www.nngroup.com/articles/ten-usability-heuristics/

4

Comparison between Wikimedia projects

There are participants who compare Wikidata Lexeme to projects such as Wiktionary and Wikisource. The team must find out the advantages and disadvantages that can be useful to note.

Thank you