



Edutainment for adults
by exploiting the Semantic Web

Poon Bing-chun

TEMPUS E SPATIUM



Download APK

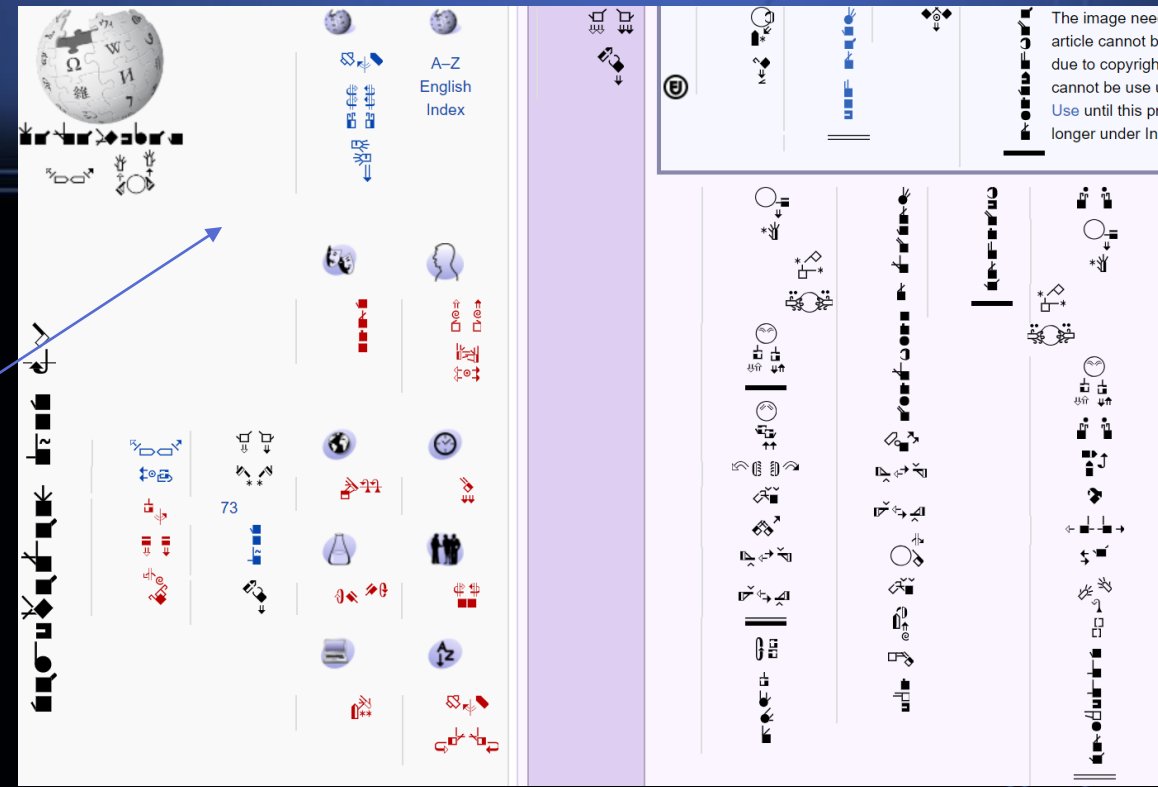
- Since I don't have a Google Play dev account, visit this page to download the APK:
- <https://github.com/SoftFeta/tempusespatium/releases>

Why?

- Most **edutainment** apps are for **kids** and specialise in some topics, too **easy** (I'd prefer super-duper hard), not so **educative**
- Never heard of a game that uses **natural language processing**
- Never heard of a game that **generates levels** real-time using external resources (hmm... Wikipedia has tens of thousands of articles...)
- Never heard of a game that has multi-lingual actual **content** (NOT interface)
- Most multi-player games are **pass-and-play**, don't make use of **multi-touch**

Wikipedias

- Not just English Wikipedia
This is a real Wikipedia
 - Tens of thousands of articles
 - Thousands of WikiProjects
 - You can learn whatever you want
-
- Why does an Wikipedia article has so many hyperlinks? Even 'animal' is a link? Readers are not that stupid aren't they?
 - ▶ Semantic web



Why was `<i>` replaced by `` (emphasis)?

► Semantic web (to discourage people using italics not for emphasis)

In Wikipedia, hyperlinked terms are important terms!

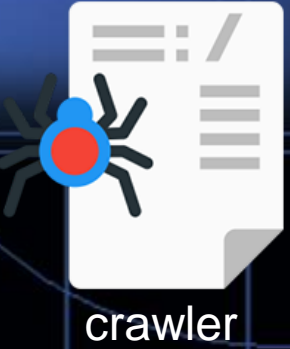
Semantic Web

Definition from Wikipedia

A web of data (or data web) that can be processed by machines—that is, one in which much of the meaning is machine-readable.

Feature: carefully designed DOM

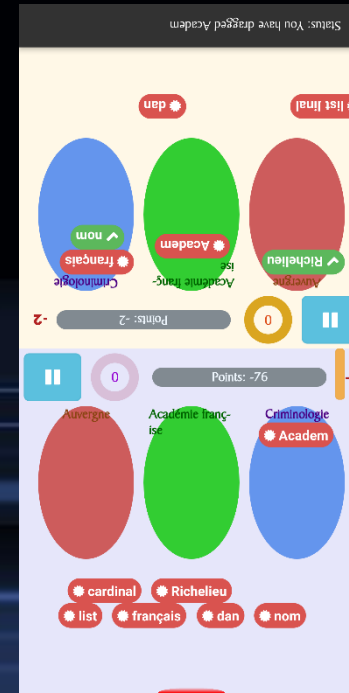
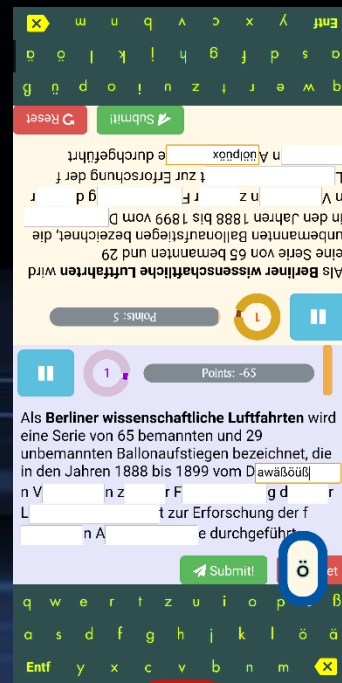
I use 2 approaches in my app



1. The **traditional** way – Web scraping

Used in 2 of my mini-games:

- Fill in the blanks (left)
- Match the category to the most relevant words (right)





crawler

1. The **traditional** way – Web scraping

Q: What is web scraping?

A: In short, we specify a 'base website' and get its content, then follow its hyperlinks until some useful data is found.

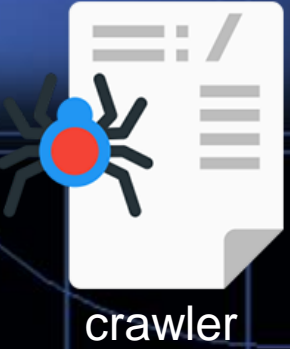
Q: How is web scraping done on Android?

A: Obscure, but there are 2 libraries *org.w3c.dom* and *javax.xml.xpath.xpath* on Android for writing your own crawlers. I send GET requests to get webpages using the OKHttp library. I use RegEx too to look for interesting data.

Q: What is XPath? Any example?

A: XPath is a query language, like a DOM selector but with **predicates**, **wildcards**, math functions and operators.

```
//*[@id="mw-content-text"]/div/table/tbody/tr/td/table/tbody/tr/td//p[b/text()='fr']/a[not(../img)][1]
```



1. The traditional way – Web scraping

Q: How do you group Wikipedia articles by topics?

A: Depending on the language of Wikipedia. In English and French Wikipedia, I use **WikiProjects**. In German and Ukrainian Wikipedia, I use **Portals**.

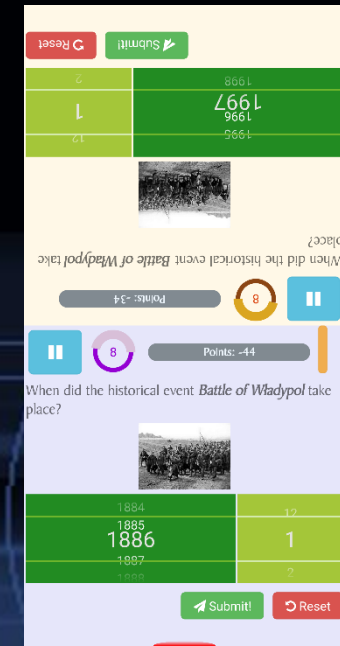
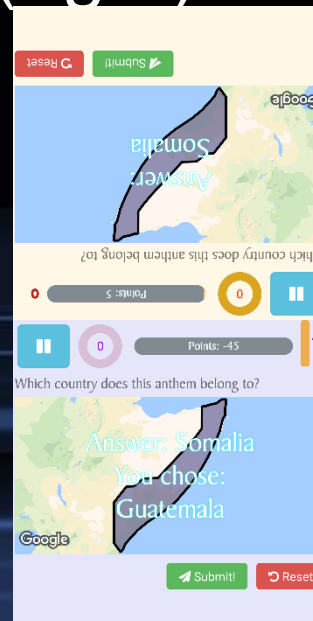
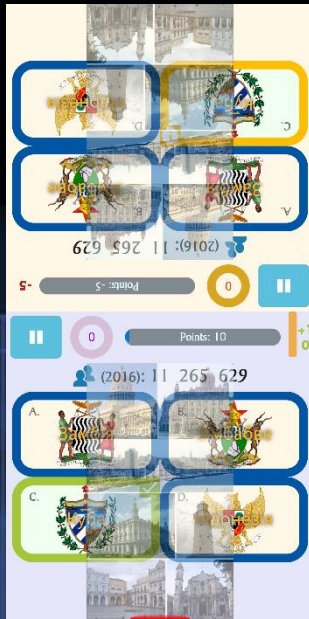
Q: How do you ensure the articles are significant enough?

A: Depending on the language of Wikipedia. In English Wikipedia, I sort articles by recent **popularity** since the info is available. In other Wikipedias, I sort articles by the **WikiProject quality scale** (Featured > A > Good > B...) or sieve by **importance scale** (Top > High > Mid > Start > Stub).

2. The **newer** way – Semantic queries

Used in 3 of my mini-games:

- Choose the correct arms/flag (left) (even the country names are in Ukrainian, as Wikidata has almost any language)
- Hear the song and pin the map (centre)
- Scroll the correct date (right)





2. The **newer** way – Semantic queries

Q: What is SPARQL?

A: In short, another query language developed by W3C. It looks like SQL and a bit like Prolog. It is specialised to interpret **1. named graphs, 2. linked-data, or 3. triples (which are, subject-predicate-object tuples. Now you know what the Semantic Web is!)**

Q: SPARQL does not support many SQL features, even the AND OR operators. Why do you use it?

A: Because Wikidata, a sister project of Wikipedia has a query service using SPARQL. You have to use the Prolog semicolon **(as well as many syntaxes)** for AND, and the keywords like UNION and BIND **(as well as many syntaxes)** for OR.

Q: I still don't get the appeal?

A: [Link 1](#) [Link 2](#)

What is sacrificed?

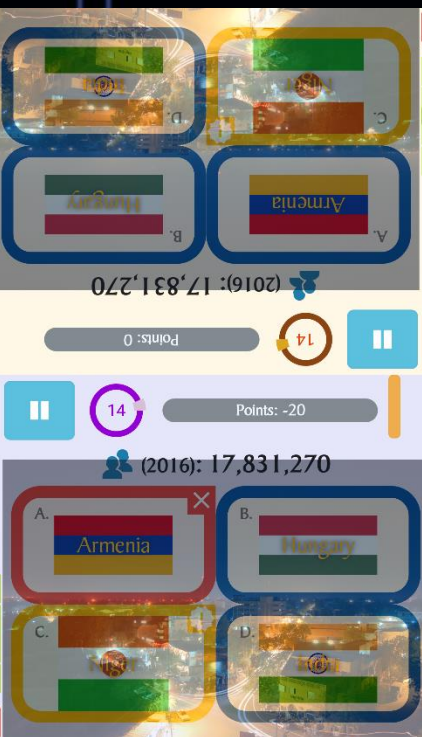
- Speed, Internet connection needed (using existing database is faster than obtaining data real-time)
- Accuracy (both Wikipedia and Wikidata can be edited by everyone)

However we get...

- Storage
- No need to update database (except highscores)
- Great quantity of data (customisable!)

However we get...

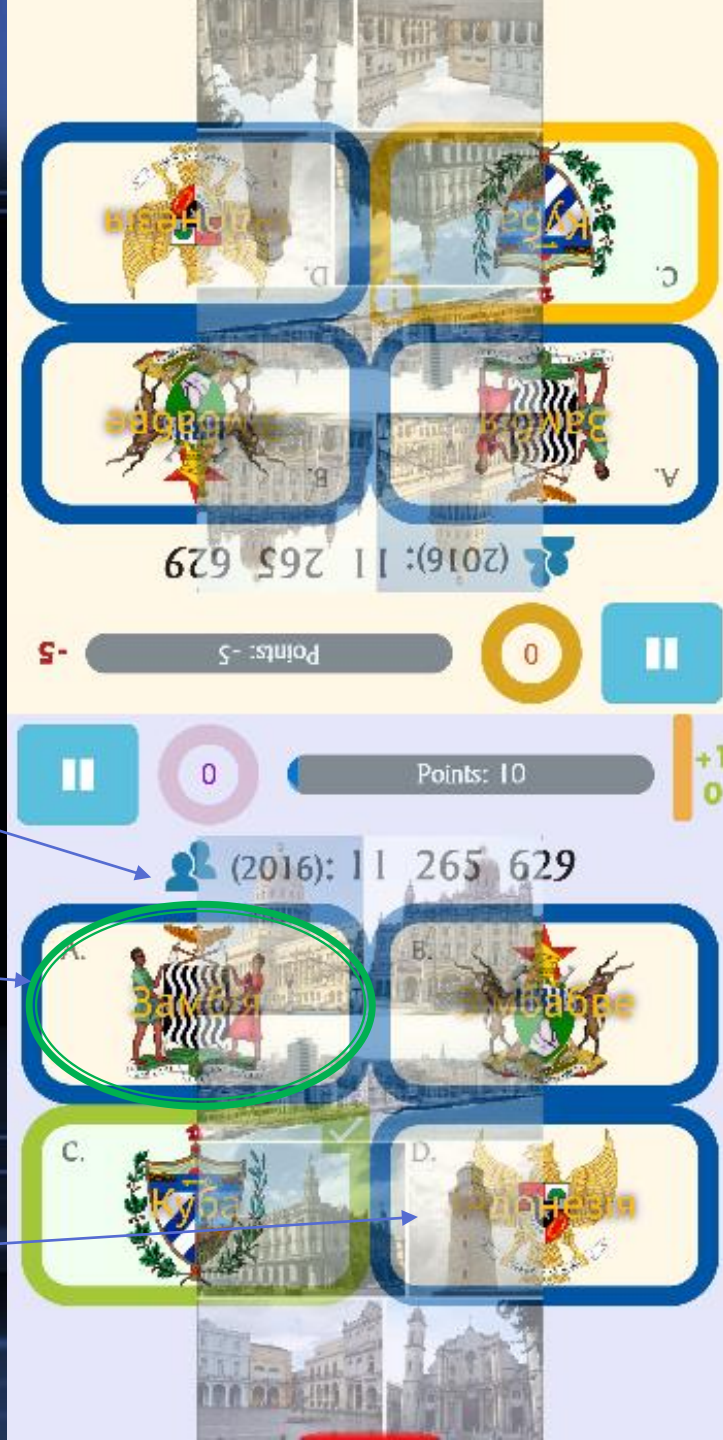
- Increased customisability (or more random content)



May be country name, capital name, population, etc, ...

May be flag or coat of arms

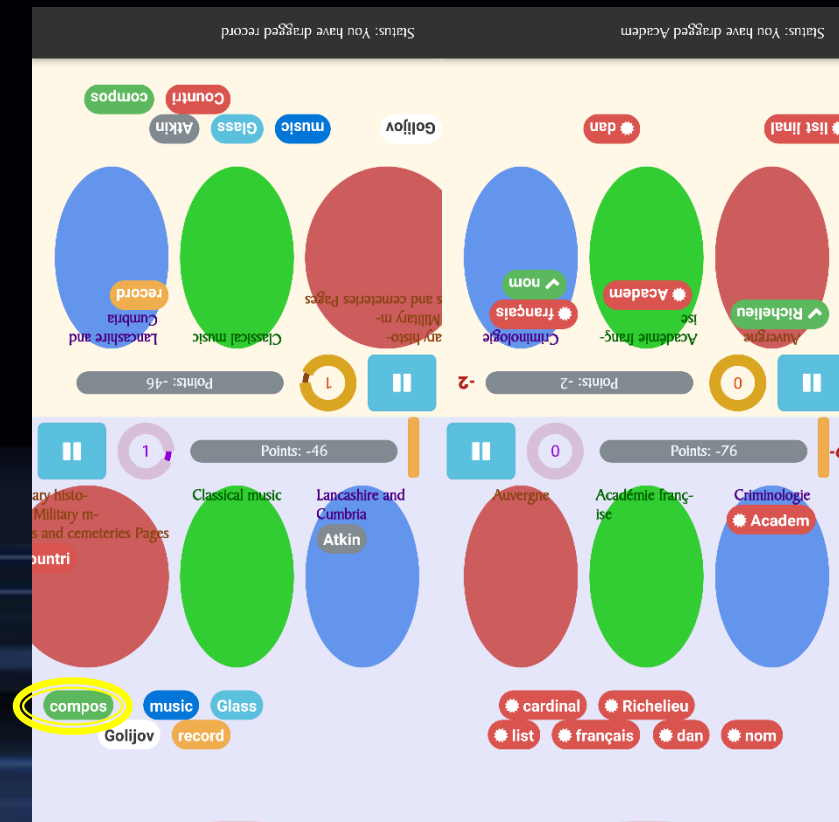
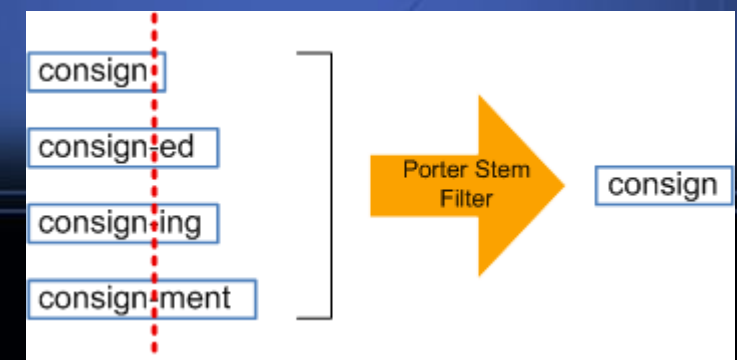
Localised content (Indonesia in Ukrainian is 'Індонезія')



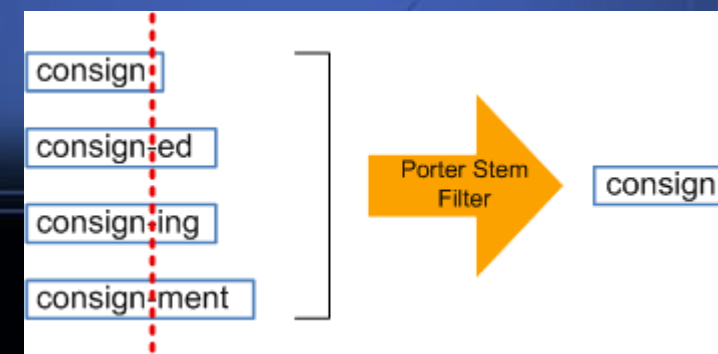
Stemming

Used in 1 of my mini-games:

- Match the category to the most relevant words



Stemming



Q: How do I count the occurrence of words (or find word relevance) on a page?

A: Difficult because of **grammar**. The most accurate way is to have a corpus (**huge data!**), and you convert every word to its noun form and count them.

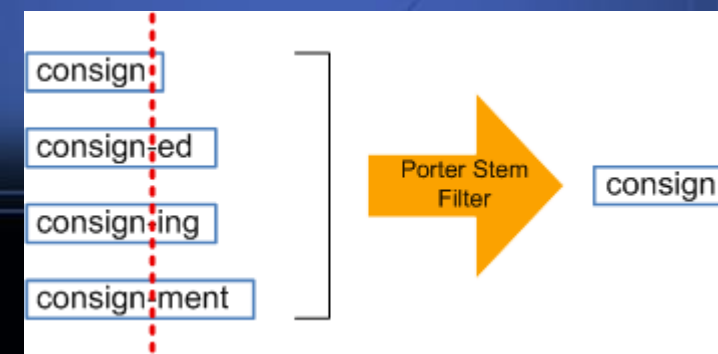
Or maybe you can try **lemmatisation**, which many lemma rules are hardcoded (**still big! But it's the only way you can know that the original form of 'bought' is 'buy'**)

The simplest way is **stemming**, which is based on naïve rules like affix removal, sometimes you will get weird results like 'amusing → amus' (**without an 'e'**). But since we want speed and storage over accuracy, it is the best!

Also, you have to skip '**stop words**' like 'the', 'and', 'because', 'among'... since they are meaningless!

Do it for French, Spanish and German as well! (**different rules of course**)

Stemming



Q: How do you stem on Android?

A: The only way seems to be Snowball stemmer generator (which is old software since 2002!). You can't use spacy, nltk (obviously), even Apache Lucene on Android... (learn some linguistics dude.)

Keyboards

Q: I don't want to install a Ukrainian keyboard, but I want to learn in Ukrainian?

A: The embedded soft keyboards were written by me, so you don't have to download any. Now you can input Cyrillic (А Б В Г Д Е Ж...). My French keyboard has dead keys to minimise the keyboard size.

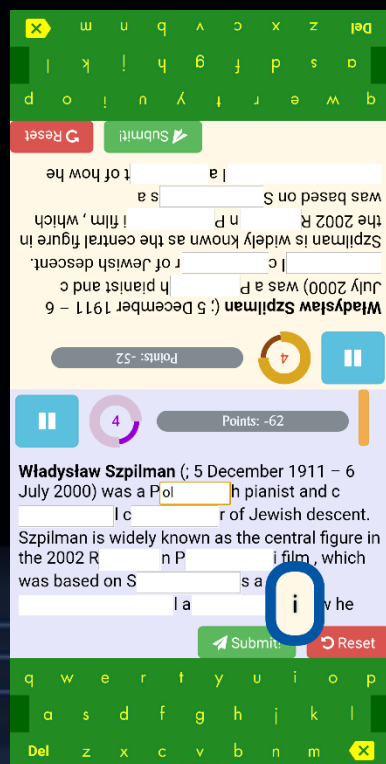
Q: I am a beginner in German. I want to learn German but if I change the language, I don't understand the UI instructions?

A: Don't worry! UI language and Gameplay language are separated settings. You will see in a moment.

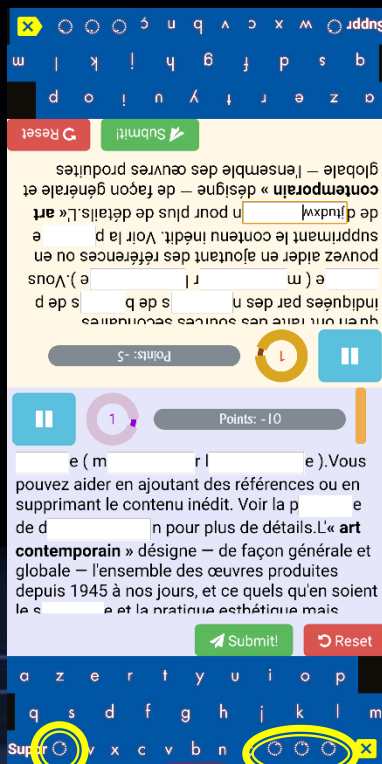
Language Challenges

Keyboards

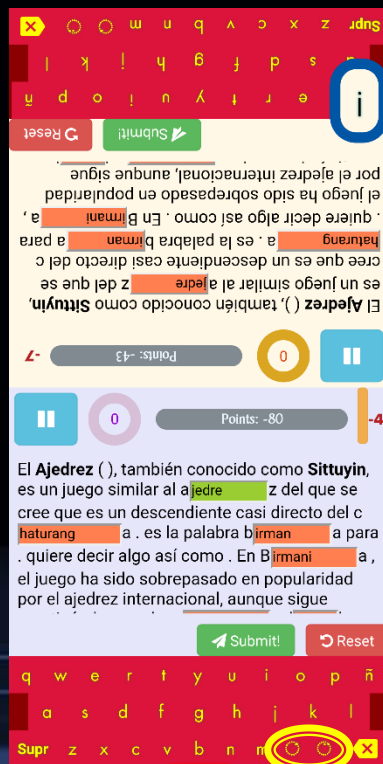
5 keyboards for 5 different languages (2 of them uses **dead keys** for typing accents):



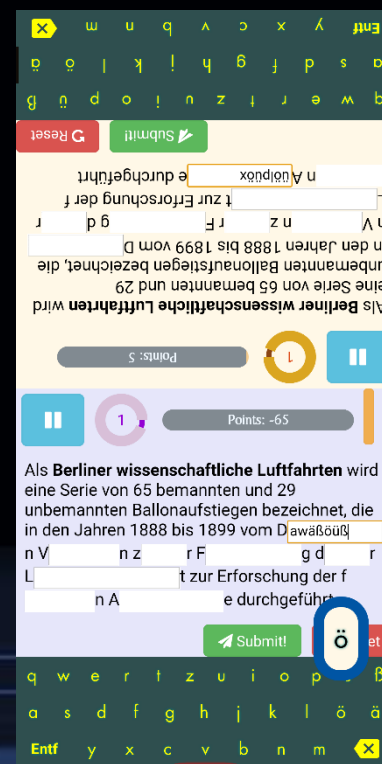
English QWERTY



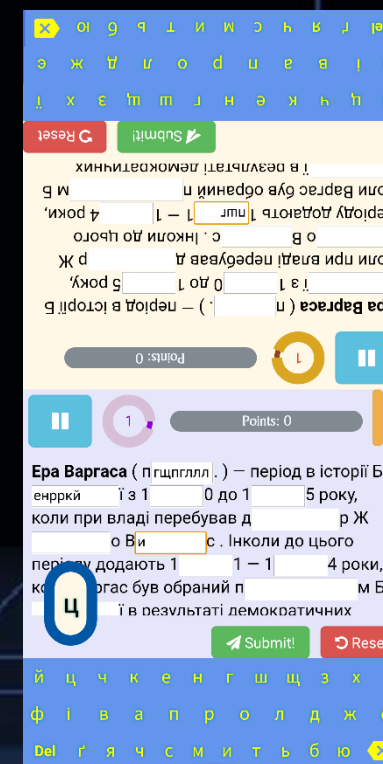
French AZERTY
(w/ dead keys)



Spanish QWERTY
(w/ dead keys)



German QWERTZ



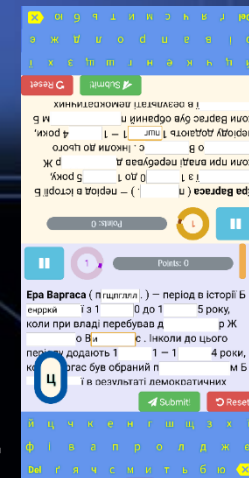
Ukrainian

Pretend that there are two cursors

Q: In the 'fill in the blanks' game, how do you mimic that there are two cursors?

A: The so-called 'cursor being inactive' is just visually inactive. The JavaScript engine will still consider the text field to be the active element.

There is also a reason why I hide the default keyboard and make my own keyboards. In this way keyboards on each side will call JavaScript to the WebView on the same side only. It does not actually input a character.



Other challenges

- Async everywhere
 - You cannot do any networking on the main thread
 - Pausing the game needs Semaphore (**CountdownLatch**)
 - Timer needs Semaphore
 - Exchanging data with the WebView needs `SystemClock.sleep` and `CountdownLatch` (**cannot exchange data directly even with @JavaScriptInterface**)
 - ...
- Loading SVG

