



Wikidata Powered Language Keyboards

Andrew Tavis McAllister

Senior Data Analyst: ResearchGate
Initiator and Lead Developer: Scribe

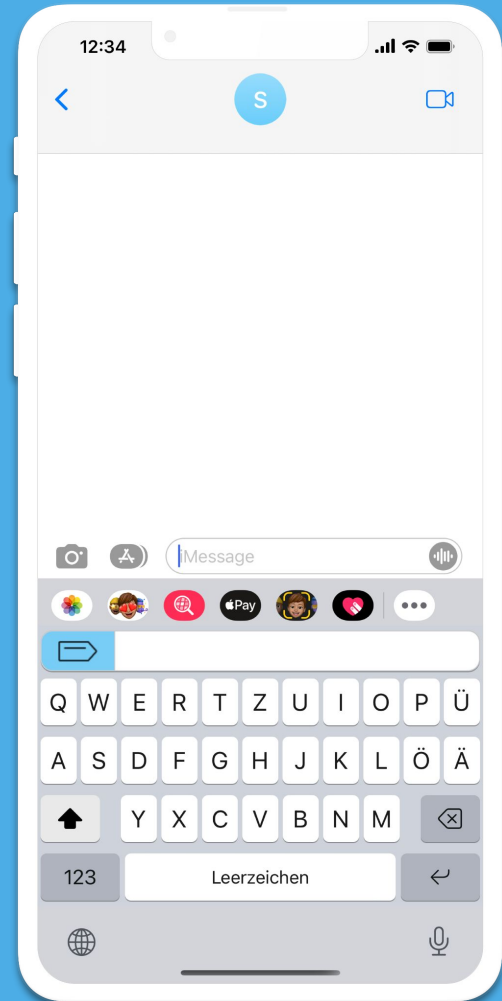


CocoaHeads Berlin

Schedule

- Presentation (~30 min)
 - Short Introduction
 - iOS Demo
 - App Structure
 - Wikidata + Data Process
 - Future Plans
- Discussion (rest of session)

* Questions always welcome!



Overview

- Keyboards for second language learners
- “Never leave your keyboard”
- Noun genders, verb conjugations, translations, preposition cases
- French, German, Italian, Portuguese, Russian, Spanish, Swedish
- Open-source + open-data + open-community
- [Wikimedia Featured Project](#) for new Developers

iOS Demo

Scribe's German Keyboard

App Structure

- Mostly using UIKit and Foundation
- Base: [KeyboardViewController.swift](#)
 - Each keyboard is a Custom Keyboard Extension
 - Each extension has a subclassed view controller
- Language data stored as JSONs

```
Scribe-iOS/  
├── Scribe/  
│   ├── ViewController.swift  
│   └── ...  
└──  
    ├── Keyboards/  
    │   ├── KeyboardsBase/  
    │   │   ├── Keyboard.xib  
    │   │   ├── KeyboardViewController.swift  
    │   │   └── ...  
    │  
    │   └── LanguageKeyboards/  
    │       ├── German/  
    │       │   ├── Data  
    │       │   │   ├── DECommandVariables.swift  
    │       │   │   ├── DEInterfaceVariables.swift  
    │       │   │   └── DEKeyboardViewController.swift  
    │       └── other languages...  
    └──
```

Wikidata

- A collaboratively edited knowledge graph (CC0)
 - From Wikimedia (DE)
 - Unique identifiers (QIDs)
 - Scribe uses [language data](#)
 - [Wikidata Query Service](#)
 - APIs: Python, JS, Java, Swift



Data Process

- Data in app (no internet)
- **WDQS** queries for each word type of each language
- JSON outputs saved
- Data formatting
 - Outputs to Scribe apps
- Takes ~10min

```
Scribe-Data/src/scribe_data/  
└─ load/  
    │ update_data.py  
    │  
    └─ extract_transform/  
        └─ German/  
            └─ nouns/  
                │ query_nouns.sparql  
                │ nouns_queried.json  
                │ format_nouns.py -> Scribe-iOS, etc.  
            └─ prepositions  
            └─ translations  
            └─ verbs  
  
        └─ other_languages...
```

```
scribe_data/load [main] » python update_data.py
```

```
Data updated: 0%| | 0/16
```

```
Querying German nouns
```

Current Data

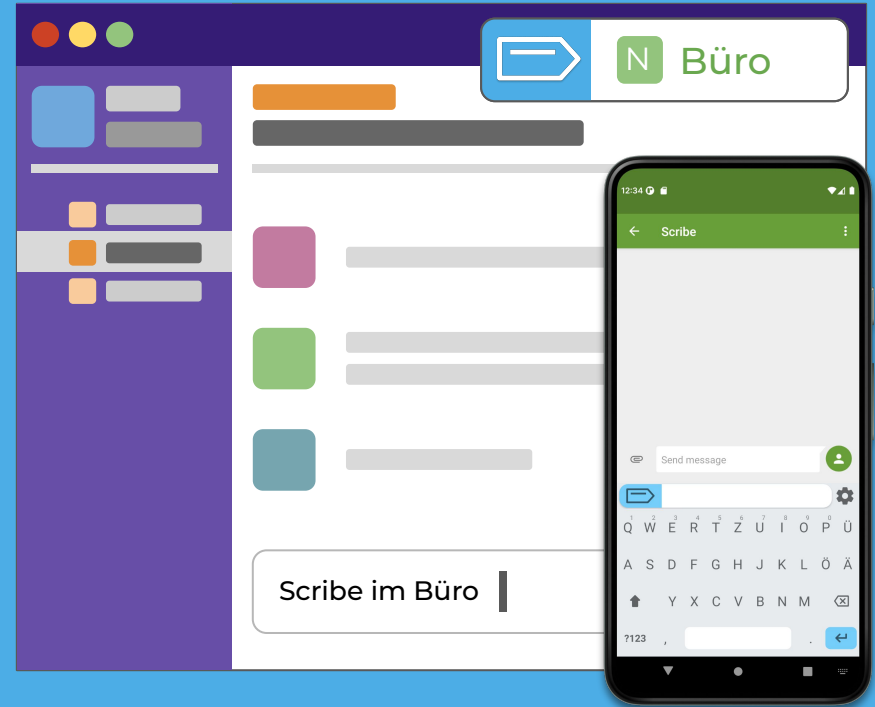
Language	L2 Speakers	Nouns (forms)	Verbs	Translations*	Preps
French	194.2 million	16,233 (3)	1,501**	67,652	-
German	59.1 million	28,745 (4)	3,381	67,652	187
Italian	13.4 million	847 (3)	73**	67,652	-
Portuguese	25.2 million	5,109 (3)	495	67,652	-
Russian	104.1 million	194,401 (4)	11**	67,652	12
Spanish	73.6 million	20,987 (3)	3,717	67,652	-
Swedish	3.2 million	42,002 (3)	4,160	67,652	-

* Machine translations

** Mostly just infinitives

Future Plans

- [Wikimania 2022](#) (8/11-8/14)
- [Scribe-Android](#) in Kotlin
- [Scribe-Desktop](#) GUI in Python
- Autosuggest, complete and correct (🤗 Hugging Face)
- More languages
 - Czech, Norwegian, Estonian, Hebrew and Danish all ready
 - See [Scribe-Data files](#)
- Language learning :)



Thank you for your interest in open-source!

To stay in touch or contribute:

<https://github.com/scribe-org>

Andrew Tavis McAllister
andrew.t.mcallister@gmail.com