Introduction to Metapinyin

https://github.com/pffy/pinyinbase

The Pffy Authors

Hanyu Pinyin is fine.

- Helps learners speak Mandarin Chinese
- Gives tourists useful signage
- Lowers barrier for business transactions
- **The problem is pinyin document search.**

Pinyin Search is broken.

- Learners spell pinyin words wrong
- Native speakers use wrong pinyin tones
- Knowledge base input is inflexible
- Pinyin documents are poorly indexed
- **The solution is metapinyin.**

What is Metapinyin?

Metapinyin describes any hashing construct designed to improve the storage and retrieval of pinyin documents. For example: **pmash, pbash, psmash,** and **phash.**

pmash: pinyin tones with no spaces

- **pmash - very specific pinyin search terms**
- pbash
- psmash
- phash

PINYIN: sheng1 ri4 kuai4 le4

PMASH: sheng1ri4kuai4le4

pbash: pinyin without tones or spaces

- pmash
- **pbash - used in modern search engines**
- psmash
- phash

PINYIN: sheng1 ri4 kuai4 le4

PBASH: shengrikuaile

psmash: initials of each pinyin word

- pmash
- pbash
- **psmash - used in Pinyin IMEs**
- phash

PINYIN: sheng1 ri4 kuai4 le4

PSMASH: srkl

phash: unique set of alpha characters

- pmash
- pbash
- psmash
- **phash - used to correct spelling in search**

PINYIN: sheng1 ri4 kuai4 le4

PHASH: aeghiklnrsu

Metapinyin already works.

Metapinyin requires the transformation of Hanyu Pinyin to hashes that can be searched with standard keyboards. **Using multiple hashing constructs in pinyin search is the key to a better user experience.**

Thank you.

https://github.com/pffy/pinyinbase

The Pffy Authors