1.2.3. Emojis

Emojis are characters in Unicode. The application WhatsApp uses special fonts such as to have the same appearance of emojis on all operation systems. In our corpus browsers, emojis can be displayed, but they are represented in the font that is used by the user, thus, it cannot be guaranteed that an emoji in the original text looked as it does on your screen.

Querying emojis is not an easy task. We decided to encode them in the messages, e.g. as *emojiQsmilingCatFaceWithOpenMouth*. This encoding system allows for easily finding individual or groups of emojis using Regular Expressions, e.g.:

- `emojiQ.*` finds all emojis
- `emojiQcat.*` finds all cats
- `emojiQ.*[Ff]ace.*` finds all faces, both human and cats (and maybe others).

From:
https://whatsup.linguistik.uzh.ch/ -

Permanent link:
https://whatsup.linguistik.uzh.ch/01_corpus/02_preprocessing/03_emojis?rev=1587553019

Last update: **2022/06/27 09:21**