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 guarantied that an emoji in the original text looked as it does on your screen.

Querying emojis might not always be easy. We therefor decided to encode them in texts. This emoji ⬜ would e.g. become *emojiQsmilingCatFaceWithOpenMouth*. This encoding system allows for easily finding individual or groups of emojis using Regular Expressions, e.g.:

- *emojiQ.**: finds all all emojis
- *emojiQ.*face.**: finds all faces
- *emojiQcat.**: finds all cats

You can thus query for individual emojis or for their encodings.