EmojiNet: A Machine Readable Emoji Sense Inventory
Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, Derek Doran
Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis), Wright State University, Dayton OH, USA
{sanjaya, lakshika, amit, derek}@knoesis.org

Motivation
✓ Emoji symbols take on different meanings based on the context of a message as no rigid semantics are attached to emoji by design.
✓ Hence, for machines to understand emoji, a machine readable sense inventory that lists different emoji senses is needed.
✓ EmojiNet is a machine readable sense inventory which enables this.

Approach
For each emoji $e$, in EmojiNet, it records the octet $a_e = \{c_k, f_k, R, |S|\}$, where $c_k$ is the Unicode codepoint of $e$, $f_k$ is the description of $e_k$, $R$ is the set of related emoji of $e_k$, $|S|$ is the set of categories, $a_j$, the set of emoji senses of $e_j$.

Emoji Resource u c d K I R M S
Unicode Website ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔
Emojipedia ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔
$i$Emoji ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔
Emoji Dictionary ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔ ✔

Emoji at Work
Let’s disambiguate the sense of the 😚 emoji in the tweets T1 and T2, where T1: Pray for my family 😚 god gained an angel today and T2: Hard to win, but we did it man 😚 lets celebrate 😚. EmojiNet lists two senses for 😚 and they are pray(V) and highfive(N). We extract words from sense definitions available in EmojiNet for the above two senses. They are: highfive(N) – (palm, high, hand, slide, celebrate, raise, person, head, five) and pray(V) – (worship, thanksgiving, saint, pray, higher, god, confession). We then calculate the overlap of the words in tweets with the words extracted from the sense definitions in EmojiNet. This leads us to decide that the 😚 emoji in T1 refers to pray(V) and the 😚 emoji in T2 refers to highfive(N).

Evaluation
✓ The Emoji Dictionary was merged with other open resources by matching the images in The Emoji Dictionary with images found in other resources. A nearest neighborhood-based image processing algorithm was used for this task and its accuracy was 98.42%.
✓ Two Word Sense Disambiguation (WSD) algorithms based on the Most Frequent Sense (MFS) and the Most Popular Sense (MPS) were used to assign meaning to emoji senses extracted from The Emoji Dictionary. Their combined WSD accuracy was 85.18%.

Real World Examples

<table>
<thead>
<tr>
<th>Laugh(N)</th>
<th>Can’t stop laughing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy(N)</td>
<td>Got all A’s but 😚</td>
</tr>
<tr>
<td>Sad(N)</td>
<td>Can’t buy class la</td>
</tr>
<tr>
<td>Annoyed(N)</td>
<td>Oooohhh shots fired!</td>
</tr>
</tbody>
</table>

Future Work
✓ Expand EmojiNet sense definitions with words extracted from tweets, using word embeddings model trained on tweets with emoji.
✓ Evaluate the usability of EmojiNet using Emoji Sense Disambiguation and Emoji Similarity Finding tasks and expose EmojiNet as a web service. EmojiNet demo is available at http://emoji.net.knoesis.org/

Reference – Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, Derek Doran. EmojiNet: Building a Machine Readable Sense Inventory for Emoji. In 8th International Conference on Social Informatics (SoCinfo 2016), Bellevue, WA, USA.