

استخدام منهجية المجاميع لبناء معجم مشاعر باللغة

العربية

افنان عطية السلمي

إشراف

د. معظم أحمد صديقي

المستخلص

إن وجود الكمية الهائلة من محتوى مشاركات الرأي باللغة العربية على شبكة الإنترنت أثار حاجة كبيرة لتحليل المشاعر. تحليل المشاعر مهمة صعبة، ودقتها تعتمد على توفر العديد من المصادر والأدوات مثل معجم المشاعر. إن مصادر تحليل المشاعر باللغة العربية لا تزال محدودة للغاية. معجم المشاعر واحد من هذه المصادر التي ما تزال نادرة في اللغة العربية. يهدف هذا العمل إلى بناء معجم مشاعر باستخدام منهجية المجاميع. تم بناء المعجم باستخدام مختلف مقاييس التشابه وطريقة الانتشار. يعتمد استخدام قياسات التشابه على حقيقة أن الكلمات التي تظهر في نفس السياق ستكون لها قطبية مشابهة. كما تساعد طريقة الانتشار على توزيع قيمة قطبية بين إدخلات المعجم. طريقة تشابه جيب التمام تفوقت على باقي المقاييس في بناء المعجم العربي.

A Corpus Based Approach to Build Arabic

Sentiment Lexicon

Afnan Atiah Alsolamy

Supervised By

Dr. Muazzam Siddiqui

ABSTRACT

The dramatic amount of posts expressing Arabic opinion on the Internet has raised a huge need for sentiment analysis. Sentiment analysis is a challenging task whose accuracy depends on the availability of a number of resources and tools, including sentiment lexicon. This work builds an Arabic sentiment lexicon using a corpus-based approach. The lexicon has been built using different similarity measurements and propagation method. The use of similarity measurements depends on the fact that the words are appearing in the same context will have similar polarity. Also, propagation method helps to distribute a polarity value between lexicon entries. Cosine similarity has outperformed other similarities in building the Arabic lexicon.