

# **Sentiment Analysis Algorithm for Arabic Reviews on the Movies Domain**

**Prepared by**

**Khalid Waleed Nassar Al-Mansoori**

**Supervised by**

**Prof. Riyad Al-Shalabi**

## **Abstract**

Sentiment analysis has become popular in recent years and has been used in many life aspects where the people reviews are very important to know their reaction on any product. Regarding movies, these impressions are important for people to choose the best film to watch.

Movie reviews are a very important tool for measuring the performance of the movie, most of the sentiment analysis research in those reviews are in English language, but our goal is how to analyze the feelings and reviews using Arabic language.

In this thesis, an Arabic lexicon for movies was created and the lexicon is used in applying movies sentiment analysis.

Movie Rating for Arabic Reviews System ( MRARS ) was built for sentiment analysis for Arabic movies reviews using Oracle Language.

The lexicon contains all the feelings words about movies which are positive, negative and neutral sentiment.

We collected about 1250 words, the number of positive words is 655, the number of negative words is 441 and the number of neutral words is 154.

The proposed system showed excellent performance varied between 94% - 100% and the accuracy can be raised after adding new words to the lexicon.

# خوارزمية لتحليل المشاعر للتعليقات العربية على نطاق الافلام

إعداد

خالد وليد نصار المنصوري

إشراف

الأستاذ الدكتور رياض الشلبي

## الملخص

أصبح تحليل المشاعر منتشراً في السنوات الأخيرة وتستخدم في كثير من مجالات الحياة حيث تعليقات الناس مهمة جداً لمعرفة مشاعرهم حول أي منتج. وفي ما يتعلق في الافلام فهذه الانطباعات مهمة للناس لاختيار افضل فلم لمشاهدته.

تعليقات الافلام هي أداة مهمة جداً لقياس تأثير الافلام ومعظم بحوث تحليل المشاعر في تلك التعليقات هي في اللغة الانجليزية ولكن هدفنا هو تحليل المشاعر باستخدام اللغة العربية هذه الرسالة تناولت هذا الموضوع و وضعت معجماً عربياً خاصاً للافلام ، الذي سيستخدم لتحليل المشاعر للافلام، وهذا المعجم يحتوي على جميع كلمات المشاعر للافلام والتي هي جيدة، سيئة ، ومحايدة . تم جمع حوالي 1250 كلمة، عدد الكلمات الإيجابية هو 655، عدد الكلمات السلبية هو 441 وعدد الكلمات المحايدة هو 154.

تم بناء نظام تصنيف الافلام للتعليقات العربية (م.ر.ا.ر.س) لتحليل المشاعر لتعليقات الافلام العربية باستخدام لغة اوراكل لكتابة البرنامج. واطهر النظام المقترح أداءً ممتازاً يتراوح بين 94% إلى 100% ويمكن أن تتأثر الدقة بعد إضافة كلمات جديدة إلى المعجم.