شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

HOSSAM MAGHRABY

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار

HOSSAM MAGHRABY

Cairo University

HILATSA: A HYBRID INCREMENTAL LEARNING APPROACH FOR ARABIC TWEETS SENTIMENT ANALYSIS

By

Kariman Mahmoud Hamouda Elshakankery

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Computer Engineering

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

HILATSA: A HYBRID INCREMENTAL LEARNING APPROACH FOR ARABIC TWEETS SENTIMENT ANALYSIS

By
Kariman Mahmoud Hamouda Elshakankery

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Computer Engineering

Under the Supervision of

Prof. Dr. Magda Bahaa ElDin Fayek

Professor
Computer Engineering
Faculty of Engineering, Cairo University

Dr. Mona Farouk Ahmed

Associate Professor
Computer Engineering
Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

HILATSA: A HYBRID INCREMENTAL LEARNING APPROACH FOR ARABIC TWEETS SENTIMENT ANALYSIS

By
Kariman Mahmoud Hamouda Elshakankery

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Computer Engineering

Approved by

Prof. Dr. Magda Bahaa ElDin Fayek, Thesis Main Advisor

____

Prof. Dr. Nevin Mahmoud Darwish, Internal Examiner

____

Prof. Dr. Ahmed Rafea, External Examiner, Professor at American University in Egypt

____

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

Engineer's Name: Kariman Mahoud Hamouda Elshakankery
Date of Birth: 19/10/1992
Nationality: Egyptian
E-mail: Kariman_Elshakankery@hotmail.com
Phone: 01111246043
Address: Nasr City

Registration Date: 1/10/2014
Awarding Date: …./…./…2021…..
Degree: Master of Science
Department: Computer Engineering

Supervisors:

Prof. Dr. Magda Bahaa ElDin Fayek, Thesis Main Advisor
Dr. Mona Farouk Ahmed, Advisor

Examiners:

Prof. Dr. Magda Bahaa ElDin Fayek, Thesis Main Advisor
Prof. Dr. Nevin Mahmoud Darwish, Internal Examiner
Prof. Dr. Ahmed Rafea, External Examiner, Professor at American University in Egypt

Title of Thesis:

HILATSA: A Hybrid Incremental Learning Approach for Arabic Tweets Sentiment Analysis

Key Words:
Sentiment Analysis, Opinion Mining, Hybrid Approach, Arabic Twitter Sentiment Analysis, Sentiment Classification

Summary:

Sentiment analysis (SA) is the process of analyzing writers' opinions, emotions and attitudes from documents and determining whether they are positive, negative or neutral. It is used for many reasons as developing product quality, adjusting market strategy and improving customer services. Since the evolution in technology and the tremendous growth of social networks, a huge amount of data is generated. In spite of the availability of data, there is a lack of tools and resources. Though Arabic is a popular language, there are too few dialectal Arabic analysis tools. This is because of the many challenges in Arabic due to its morphological complexity and its dynamic nature. Approaches used to classify the opinions are categorized into lexicon based, machine based and hybrid based approach. This thesis introduces a semi-automatic learning system for sentiment analysis that is capable of updating the lexicon in order to be up to date with language changes. It is a hybrid approach which combines both lexicon based and machine learning approaches in order to identify Arabic tweets sentiments polarities. It proved to be able to cope with the dynamic nature and improved the accuracy by 17.55%.

Disclaimer

I hereby declare that this thesis is my own original work and that no part of it has been submitted for a degree qualification at any other university or institute.

I further declare that I have appropriately acknowledged all sources used and have cited them in the references section.

Name: Kariman Mahmoud Hamouda Elshakankery Date: ../../…

Signature:

Acknowledgements

Thanks to Allah for everything. Thanks to my family for their patience, support and understanding. Thanks to my supervisors especially Dr Mona for her guidance and patience.

Table of Contents

List of Tables

List of Figures

List of Acronyms

Table I.1: List of Acronyms

Acronym	Definition
API	Application Program Interface
CBOW	Continues Bag of Words
CNB	Complement Naïve Bayes
CNN	Convolutional Neural Network
DDO	Diabetes Diagnosis Ontology
DT	Decision Tree
FN	False Negative
FP	False Positive
KNN	K Nearest Neighbor
LR	Logistic Regression
LSTM	Long Short Term Memory
MSA	Modern Standard Arabic
NB	Naïve Bayes
NN	Neural Network
POS	Part of Speech Tagging
RF	Random Forest
RNN	Recurrent Neural Networks
SA	Sentiment Analysis
SG	Skip Gram
SVM	Support Vector Machine
SWN	SentiWordNet Lexicon
TDOC	Term Degree of Correlation
tf	Term Frequency
tf-idf	Term Frequency Inverse Document Frequency
TN	True Negative
TP	True Positive
URL	Universal Resource Locator