



UNIVERSITY OF THE PELOPONNESE
SCHOOL OF SOCIAL AND POLITICAL SCIENCES
DEPARTMENT OF POLITICAL SCIENCE & INTERNATIONAL
RELATIONS

Master of Arts in
“Global Risks and Analytics”

Comparative analysis of the hashtags #RefugeesWelcome and
#StopRefugees

Tweets mining and examination with the Knime Analytics Platform

Master's Dissertation

Vasileios Lagos

Supervisors:

Professor A. Katsis

Associate Professor N.S Koutsoukis

Lecturer S. Vandoros

Final Version

Korinth, 2017



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ
ΣΧΟΛΗ ΚΟΙΝΩΝΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΟΛΙΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΔΙΕΘΝΩΝ ΣΧΕΣΕΩΝ

Πρόγραμμα Μεταπτυχιακών Σπουδών

«Παγκόσμιες Προκλήσεις και Συστήματα Αναλύσεων»

Συγκριτική ανάλυση των hashtags #RefugeesWelcome και #StopRefugees

Άντληση Tweets και εξέτασή τους με χρήση της ηλεκτρονικής πλατφόρμας analytics Knime

Μεταπτυχιακή Διπλωματική Εργασία

Βασίλειος Λαγός

Τριμελής Επιτροπή:

Καθηγητής Α. Κατσής

Αναπληρωτής Καθηγητής Ν.Σ. Κουτσούκης

Λέκτορας Σ. Βανδώρος

Τελική έκδοση

Κόρινθος, 2017

Φύλλο Αξιολόγησης

Η διπλωματική εργασία με τίτλο «Συγκριτική ανάλυση των hashtags #RefugeesWelcome και #StopRefugees» του Βασίλειου **Λαγού** αξιολογήθηκε από την τριμελή επιτροπή, τόσο ως προς την ποιότητα του κειμένου, όσο και ως προς την ποιότητα της προφορικής παρουσίασης και υπεράσπισης της διπλωματικής εργασίας ενώπιον ακροατηρίου.

Η διαδικασία αξιολόγησης της διπλωματικής εργασίας ολοκληρώθηκε την ___/___/___ με γενική επίδοση:

- Καλώς
- Λίαν Καλώς
- Άριστα

Τα μέλη της τριμελούς επιτροπής:

1. Καθηγητής Α. Κατσής
2. Αναπληρωτής Καθηγητής Ν.Σ. Κουτσούκης
3. Λέκτορας Σ. Βανδώρος

Περίληψη

Η παρούσα εργασία εξετάζει δεδομένα σχετικά με τα hashtags #RefugeesWelcome και #StopRefugees. Η άντληση των δεδομένων γίνεται από το Twitter με τη χρήση της πλατφόρμας analytics Knime και αφορά tweets που έχουν δημοσιευθεί την περίοδο 2007-2017. Αρχικά, γίνεται μια περιγραφή των αιτιών του προσφυγικού ζητήματος, καθώς και του αντίκτυπου που οι μαζικές προσφυγικές ροές είχαν στις χώρες της Ευρώπης και της Αμερικής. Στη συνέχεια γίνεται αναφορά στο ζήτημα της αξιοπιστίας των δημοσιεύσεων στο Twitter, καθώς και στο κοινό που ο κάθε χρήστης απευθύνεται. Ακολουθεί η παρουσίαση των ευρημάτων ανά κάθε hashtag ξεχωριστά. Τέλος, παρουσιάζονται τα συμπεράσματα που μπορούν να εξαχθούν από τα στοιχεία που καταγράφηκαν, γίνεται αναφορά στους περιορισμούς της καταγραφής καθώς και προτάσεις για μελλοντική έρευνα στο αντικείμενο.

Όροι Κλειδιά: Προσφυγικό ζήτημα, Twitter, #RefugeesWelcome, #StopRefugees, Knime

Abstract

The current dissertation examines data related to the #RefugeesWelcome and #StopRefugees hashtags. The Twitter data mining is held with the Knime Analytics Platform and regards tweets published during the 2007-2017 period. First, a description of the reasons that led to the refugee issue takes place, as well as of the impact that the massive refugee flows had to the countries of Europe and the US. Following this, there is a reference on the credibility of twitter data and on the twitter audience. Subsequently, there is a presentation of the research findings per each hashtag respectively. The dissertation concludes with a reference to the limitations of the data discovery and recording, as well as with suggestions for future research on the subject.

Keywords: Refugee issue, Twitter, #RefugeesWelcome, #StopRefugees, Knime

Contents

1	Introduction	1
2	The refugee issue	2
2.1	Introduction	2
2.2	Definition of terms	2
2.3	The Syrian political and diplomatic aspects.....	3
2.3.1	The opposition	4
2.4	Syrian refugee statistics.....	6
2.4.1	Europe.....	6
2.5	Humanitarian assistance	8
2.6	The refugee relocation system.....	10
2.7	United States of America	15
3	Twitter	17
3.1	Introduction	17
3.2	Information credibility on twitter	17
3.3	Twitter audience	20
4	Knime analytics platform tools.....	21
4.1	Introduction	21
4.2	Knime	21
4.2.1	Repository.....	22
4.2.2	Nodes	22
4.2.3	Meta nodes.....	22
4.2.4	Workflow management	23
5	#RefugeesWelcome and #StopRefugees search.....	25
5.1	Introduction	25
5.2	Creation of the twitter application.....	25
5.3	#Refugees Welcome and #Stop Refugees tweets mining	26
5.4	Exclusion of retweets	28
6	Significant terms extraction.....	31

6.1 Introduction	31
6.2 Text preprocessing	31
6.2.1 Reading text data	32
6.2.2 #RefugeesWelcome and #StopRefugees input data preprocessing	32
6.3 Significant terms extraction	33
6.3.1 #RefugeesWelcome significant terms' extraction	34
6.3.2 #Stop Refugees significant terms' extraction	37
6.4 Conclusions	38
7 Tweets' location extraction	40
7.1 Introduction	40
7.2 #RefugeesWelcome tweets' location extraction	41
7.3 #StopRefugees tweets' location extraction	42
8 Examination of the number of tweets per year published	43
8.1 Introduction	43
8.2 #RefugeesWelcome number of tweets per year published visualization	43
8.3 #StopRefugees number of tweets per year published visualization	44
9 Examination of the popular tweets	46
9.1 Introduction	46
9.2 #RefugeesWelcome popular tweets	46
9.3 #Stop Refugees popular tweets	49
9.4 Conclusions	50
10 #RefugeesWelcome and #StopRefugees centralized comparison	51
11 Conclusions	52
12 References	54

1 Introduction

Data analytics is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software. Data analytics technologies and techniques are widely used by scientists and international relations researchers to verify or disprove scientific models, theories and hypotheses.

The massive growth of data generated from social media sources have resulted in a growing interest on efficient and effective means of collecting, analyzing and querying large volumes of data. In particular, online social networking and microblogging platform Twitter has seen exponential growth in its user base since its inception in 2006 with over 200 million monthly active users producing 500 million tweets daily.

The current dissertation examines the context of two ostensibly contrary hashtags that relate to the Syrian refugee issue, #RefugeesWelcome and #StopRefugees. The research includes tweets published from 2007 to 2017.

There is a comparative description of the number of tweets each hashtag returns, the countries that twitter users tweet from, the most significant terms related to each hashtag, as well as of the number of tweets per year for each hashtag. Subsequently, the same comparison is being conducted for the most popular tweets related to each hashtag. The dissertation concludes with a reference to the limitations of the data discovery and recording, as well as with suggestion for future research on the subject.

The data analysis is conducted with the KNIME 3.3.1 Analytics Platform.

A weakness of the research is that twitter users do not always use their real location to geolocate their tweets, a fact that could affect the research's findings. Another weakness is that the tools of Knime do not mine all the tweets from Twitter. Last, the significant terms related to the hashtags, cannot confirm that users tweeting for #RefugeesWelcome are positively adjacent to the refugee issue or negatively adjacent when tweeting for #StopRefugees, respectively.

In the chapter to follow, a reference on the causes of the refugee issue and statistics related to it is conducted.

2 The refugee issue

2.1 Introduction

An estimated 11 million Syrians have fled their homes since the outbreak of the civil war in March 2011. After six years of war, a number of 13.5 million are in need of humanitarian assistance into the country. The majority of Syrians who chose and managed to escape the conflict sought refuge in neighboring countries or within Syria itself. According to the United Nations High Commissioner for Refugees (UNHCR), 4.8 million have fled to Turkey, Lebanon, Jordan, Egypt and Iraq, whereas 6.6 million are internally displaced within Syria. Meanwhile, about one million have requested asylum in Europe. Germany, with more than 300,000 cumulated applications, and Sweden with 100,000 are EU's top receiving countries. (syrianrefugees, 2016)

At the same time, the United States accepted a number of more than 10,000 Syrian refugees in 2016 aiming to surpass this number in 2017 and has spent \$364 million in humanitarian assistance since 2012. (Bureau of Population, Refugees, and Migration, 2016)

In this chapter the Syrian political and diplomatic aspects will be discussed and there will be a further description of Syrian related statistics in Europe and the US.

2.2 Definition of terms

In order to better realize the challenges the EU member states face, as well as the findings of this thesis, there has to be a definition of different terms regarding the refugee issue.

The article 2 par. D of the Directive 2011/25 EU of the European Parliament and of the Council of 13 December 2011 on standards for the qualification of third-country nationals or stateless persons as beneficiaries of international protection, for a uniform status for refugees or for persons eligible for subsidiary protection, and for the content of the protection granted, provides information on what meaning the EU gives to the regarded parts.

According to that

a) "international protection" means refugee status and subsidiary protection status,

b) “beneficiary of international protection” means a person who has been granted refugee status or subsidiary protection status

c) “refugee” means a third-country national who, owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, political opinion or membership of a particular social group, is outside the country of nationality and is unable or, owing to such fear, is unwilling to avail himself or herself of the protection of that country, or a stateless person, who, being outside of the country of former habitual residence for the same reasons as mentioned above, is unable or, owing to such fear, unwilling to return to it, and to whom Article 12 does not apply;

d) “refugee status” means the recognition by a Member State of a third-country national or a stateless person as a refugee;

e) “application for international protection” means a request made by a third-country national or a stateless person for protection from a Member State, who can be understood to seek refugee status or subsidiary protection status, and who does not explicitly request another kind of protection, outside the scope of this Directive, that can be applied for separately;

f) “family members” means, in so far as the family already existed in the country of origin, the following members of the family of the beneficiary of international protection who are present in the same Member State in relation to the application for international protection: — the spouse of the beneficiary of international protection or his or her unmarried partner in a stable relationship, where the law or practice of the Member State concerned treats unmarried couples in a way comparable to married couples under its law relating to third-country nationals, — the minor children of the couples referred to in the first indent or of the beneficiary of international protection, on condition that they are unmarried and regardless of whether they were born in or out of wedlock or adopted as defined under national law, — the father, mother or another adult responsible for the beneficiary of international protection whether by law or by the practice of the Member State concerned, when that beneficiary is a minor and unmarried;

g) “country of origin” means the country or countries of nationality or, for stateless persons, of former habitual residence. (Commission, Directive 2011/95/EU of the European Parliament and of the Council, 2011, p. 4)

2.3 The Syrian political and diplomatic aspects

The wave of the Arab uprisings that started in Tunisia on January 2011, reached Syria in mid-March, when residents of the small town Dara’a took the streets to protest the torture against a group of students who had put an anti-government graffiti.

The protest soon spread over other major cities of the country, like Damascus, Homs and Hama. Important distinguisher is that in these protests many were calling for something more modest than a regime change. They were calling for a release of political prisoners, an end to the half-century-old state of emergency, greater freedoms, and an end to corruption. These protests did not have the same results like Tunisia or Egypt, as Assad responded to protestors immediately and forcefully, in order to try to stop an insurrection.

The clash soon took the form of a civil war, with the Syrian army firing on unarmed protestors, heading to mass arrests, as well as cutting off basic needs like food, water and electricity to cities that kept on the protests. This upheaval led to the protesters taking arms as well. (CFR, 2016)

2.3.1 The opposition

The opposition, as the regime's opponents are used to be referred to us, is constituted by a wide range of parts, which have a number of different interests in the area.

The Free Syrian Army, which is constituted by defectors of the Syria's army, was created after the settling of the defectors in Turkey, in July 2011. The fact that the defectors were outgunned by the regime and that they chose not to remain as local militias and not to coordinate their operations on interests resulted in it to be a loose coalition. (CFR, 2016)

At the same time, the political counterpart of the Free Syrian Army was created in Istanbul. The United States, Turkey and the Gulf Cooperation Council Countries recognize the so called government in exile of Syria the "Syrian National Coalition", as the "legitimate representative of the Syrian People". The fact that many other parts of the opposition recognized only little legitimacy to the exiled government proves the complicated rival situation on the matter.

Groups of various other interests took advantage of the situation prevailing in the country. In January 2012, Habbat al-Nusra, the Syrian wing of al-Qaeda, spread a call for Sunnis from around the region to join a jihad against the regime. The fact that it enjoyed the support of al-Qaeda's Chief, in addition to the success it had in the battlefield, were a crucial factor for it to gain foreign and Syrian recruits.

One year later, in April 2013, another group emerged from the remnants of al-Qaeda in Iraq, called Islamic State which in several months gained control in an important territory in Syria and exceeded al-Nusra in brutality,. (CFR, 2016)

It is often said that the rise of extremists groups in Syria was supported in a great extent by Syria's regime, as Assad wanted to offer the international opinion two alternatives: the one of a secular regime and that of a jihadist rule. This view was stronger supported after the release of hundreds of Islamists militants from prisons to discredit the rebellion.

The coalition between Assad's regime and the opposition parties has brought a significant death toll to unarmed civilians. This toll was further increased by the use of chemical weapons by the regime, until the international world powers took action to dismantle the chemical weapons' use.

The use of both sieges and aerial bombardment serves dual purposes according to analysts. First, they raise the costs of resistance for civilians to pressure rebels to acquiesce and also prevent opposition forces from offering a viable alternative to the regime's governance.

It is notable that in February 2014 the regimes "systematic" attacks on civilians were characterized by the UN Human Rights Council as "crimes against humanity", and stressed that same crimes were also committed by war criminals.

The deepening of the Syrian civil war made pro and anti regime forces depend on external forces. Assad's Syria was backed by Russia and Iran. The second, a longtime ally in an attempt to protect a vital land route to Hezbollah, its proxy, has resorted to investments ranking up to millions to shore up the regime.

Russia, a power traditionally averse to regime change has provided both diplomatic and air force support to the regime. It has often been claimed though, that besides the strikes against the Islamic State, Moscow targeted other opposition groups as well, in a bid to back up Assad. (CFR, 2016)

This allies' game was soon joined by the opposition forces as well. The rebel groups started to coordinate under the Army of Conquest, which was formed in March 2015 by Saudi Arabia, Turkey and Qatar, in a bid to overcome the obstacles of coordination of rebel groups in the north. CIA as well reportedly provides covert training and arms to opposition forces. (CFR, 2016)

This puzzle of rivals comes to become further perplexed with the active participation of the Kurds, through the YPG, People Protection Unit, in the conflict. Kurds have always fought to consolidate a de facto autonomous territory in northern Syria and the Islamic State's siege of Kobani in the fall of 2014 was a turning point in their relation with the Arabs. (CFR, 2016)

The fact that the YPG proved very effective in the battle against the ISIS, was a solid reason for it to enjoy military support from the US. Turkey, a significant factor in the Syrian issue, considers YPG to be an extension of the Turkey-based Kurdistan Workers' Party, which Washington has designated a terrorist organization. The United States considers Turkey a vital partner in the war against the Islamic State and it faces the dilemma of trying not to alienate either partner. (CFR, 2016)

The diplomatic attempts for a solution in the Syrian crisis face serious difficulties as well, due to the above-mentioned alliances. Saudi Arabia, Qatar and Turkey aligned with the United States against the Assad regime, while Iran joined Russia in backing it. This perplexed form of alliances

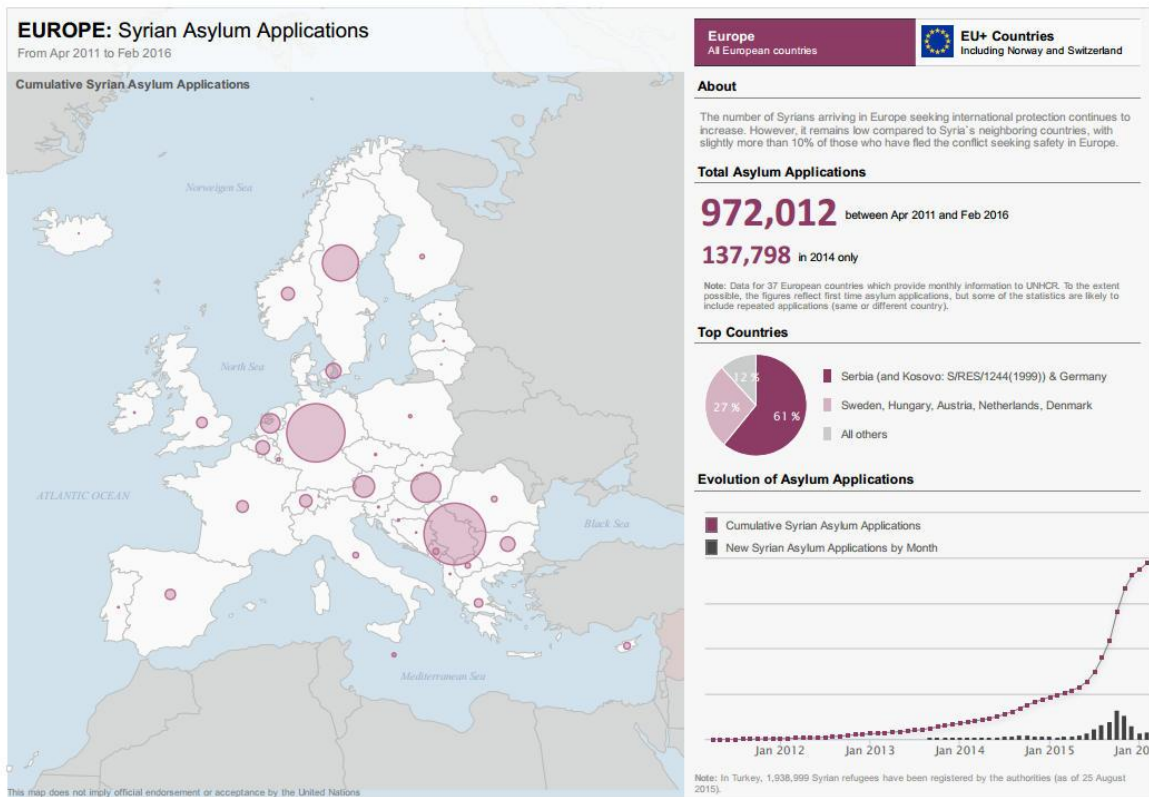
results in vetoes on resolutions of the Security Council that have to do with Syria, a fact that poses delays to any humanitarian and human rights measures. Despite this fact, the UN remains at the center of efforts to broker a peace agreement. (CFR, 2016)

The negotiations are being made in the basis of the Geneva Communiqué, issued in June 2012, a multilateral document that calls for “a Syrian-led political process”, beginning with the establishment of a transitional governing body “formed on the basis of mutual consent”. The peace talks have not come up to a result so far, due to a variety of reasons, one of which is Assad himself. There is not a question by his side in leaving power, and he enjoys the support of Russia and Iran, whereas a possibility of him staying in power in a transition would never enjoy the support of the opposition, as well as their supporters -Turkey, Saudi Arabia and the United States-, among them. Meanwhile, voices in the Security Council are in favor of a federal division of Syria, though both Assad regime and the main opposition bloc have rejected this. A partial cease-fire took effect in early March 2016. (CFR, 2016)

2.4 Syrian refugee statistics

2.4.1 Europe

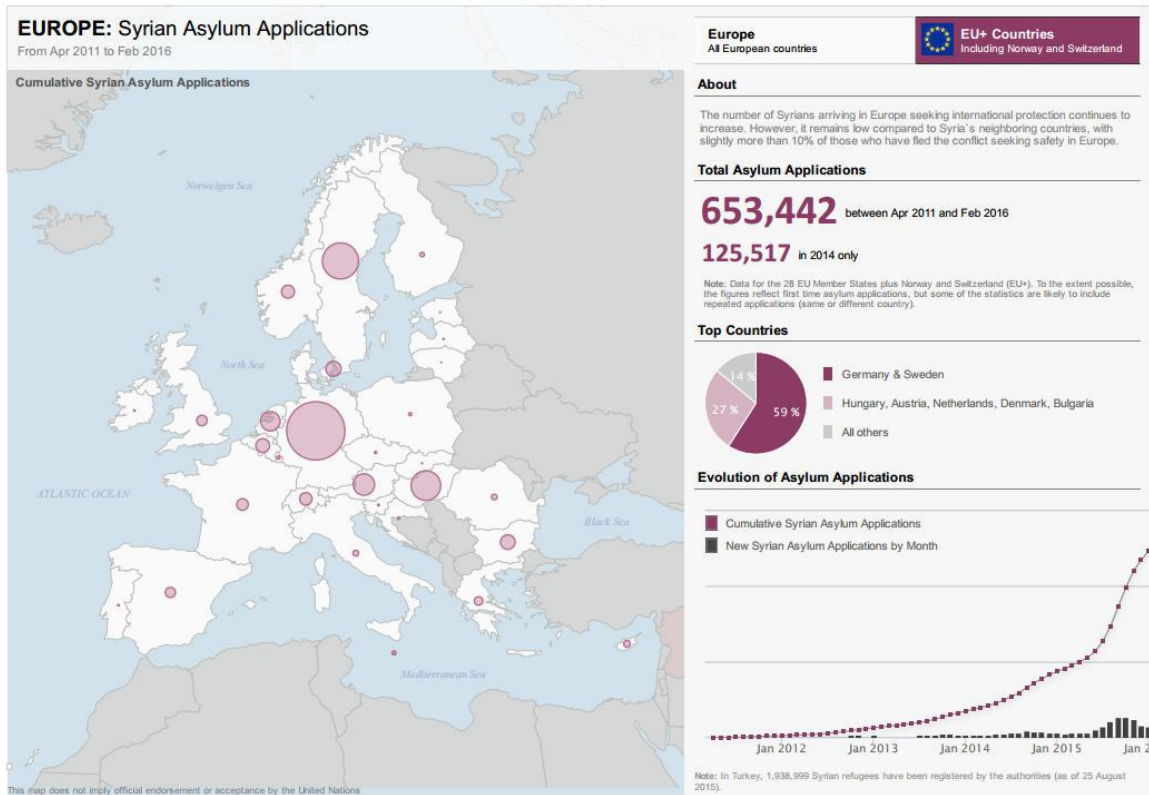
According to data from the UN Refugee Agency, the total first time Asylum applications from Syrians reached 1,177,914 between April 2011 and September 2016, of which the 137,798 of them took place only in 2014. Germany and Serbia received the biggest amount, having to evaluate the 65% of the asylum applications, whereas Sweden, Hungary, Austria, the Netherlands and Denmark share the 24% of the asylum applications in Europe. (UNHCR, 2016)



Picture 1: Europe: Syrian Asylum Applications (Source: UNHCR,2016¹)

The European Union received 857,299 out of the total Asylum Applications in Europe, in the same period of time. The 125,517 of them registered only in 2014. Germany and Sweden received the majority of them, with a percentage ranking up to 65%, whereas Hungary, Austria, the Netherlands, Denmark and Bulgaria received the rest 22%. (UNHCR, 2016)

¹ From UNHCR, The UN Refugee Agency, <http://data.unhcr.org/syrianrefugees/asylum.php>



Picture 2: EU Syrian Asylum Applications (Source: UNHCR, 2016²)

2.5 Humanitarian assistance

The European Union and its Member States are collectively lead providers of international aid. A sum of more than € 9.2 billion has been mobilized for relief and recovery assistance to Syrians inside the country from cross borders and cross lines entries and to refugees and their host communities in neighboring Lebanon, Jordan, Iraq, Turkey and Egypt. (Commission, Syria Crisis, 2016, p. 2)

The European Commission is funding humanitarian programs implemented by its humanitarian partners who provide millions of people with life-saving assistance including food and safe drinking water, non-food items, shelter, emergency medical treatments and protection.

The €445 million EU humanitarian assistance to the Syria crisis for 2016, as well as the allocation for 2017, is applied to address the needs inside Syria as well as the needs of Syrian refugees and host communities in neighboring countries.

² From UNHCR, The UN Refugee Agency, <http://data.unhcr.org/syrianrefugees/asylum.php>

The EU humanitarian assistance inside Syria responds to emergency crises throughout the country, working from all humanitarian hubs. Since the launch of the Humanitarian Initiative for Aleppo on October 2016, the EU has been working to ensure minimum conditions for humanitarian access and for the security and safety of medical staff and humanitarian workers. At the operational level, the EU has been working closely with humanitarian partners including the UN, ICRC and INGOs, to support the delivery of humanitarian assistance and urgent medical evacuations from Eastern Aleppo, allocating an emergency contribution of € 25 million.

In Lebanon, EU humanitarian funding has provided assistance to most vulnerable refugees, secondary health care for lifesaving cases, non-formal education and shelter, including water, hygiene and sanitation, to improve the living conditions of the vulnerable families mostly affected by the displacement. Since 2012, the EU's humanitarian aid to Lebanon reached about 750,000 Syrian refugees. (Commission, Syria Crisis, 2016, p. 2)

In Jordan, with 83% of the refugees living in urban settings, the European Commission considers cash assistance to be the most co-efficient and dignified method to support the most vulnerable refugees. In addition to this, the EU is working with international organizations in several sectors like protection health, winterization and basic needs. In 2016, the €53 million of EU humanitarian funding also targeted the emergency needs of more than 75,000 refugees stranded along the North-Eastern border. In 2017, the total allocation for Jordan is expected to reach €55 million.

In Egypt, the ongoing humanitarian operations amount to € 4 million and target the most in need out of the 115,000 Syrian registered refugees or awaiting registration in the country according to UNHCR. The European Union supports the most vulnerable refugees living in urban context through protection, health and education in emergency assistance.

In Turkey, the Commission's humanitarian funding supports vulnerable refugees who have fled violence in Syria as well as other countries, in particular refugees living outside of camps. The Commission's programs provide basic assistance to the most vulnerable refugees, including food assistance and emergency items, access to health care and to protection services. In 2016, the EU with Member States launched the Facility for Refugees in Turkey to deliver efficient and complementary support to Syrian and other refugees and host communities in close cooperation with Turkish authorities. The Facility provides a joint coordination mechanism for actors financed by the EU budget and national contributions made by the Member States. It is designed to ensure that the needs of refugees and host communities are addressed in a comprehensive and coordinate manner. The Facility has a €3 billion budget for 2016 and 2017 to support humanitarian interventions and long term assistance for refugees countrywide in Turkey. €1.45 billion has been contracted in humanitarian and long term assistance in Turkey until January 2017, whereas €517 million has been contracted for humanitarian operations. (Commission, Syria Crisis, 2016, p. 3)

2.6 The refugee relocation system

In order to ease the burden for the countries receiving refugees, and in the framework of the European solidarity, the European Commission introduced a refugee relocation system in September 2015. The relocation refers to the transfer of persons who are in need of international protection from one EU Member State to another EU Member State.

In this context, the Commission introduced two emergency proposals. In May 2016, the Commission proposed the relocation of 40,000 people from Italy and Greece over 2 years, a number that sums up to the 40% of the recent applicants.

In September of the same year, the Commission proposed to relocate 120,000 people from Italy, Greece and Hungary over 2 years, in an attempt to support the Member States most affected, Italy, Greece and Hungary.

According to the proposal, the number of the refugees EU Member States will receive, is based upon objective, quantifiable and verifiable criteria, that is the size of population, with a significance rate of 40%, the total GDP, with a significance rate of 40% as well, the average number of asylum application over the previous years, having a significance rate of 10% and the unemployment with a significance rate of 10%.

The Commission agreed to the appointment of Liaison Officers by every member state in order to match the destination country with refugees qualifications, language skills, family, cultural and social ties, in an attempt to facilitate the integration.

Receiving Member States get €6,000 for each person received, whereas Italy, Greece and Hungary receive €500 for each person relocated to cover transport costs. Relocation can only apply to applicants for which the average recognition rate of international protection at the EU lever is above 75%. Currently only three nationalities have such high recognition rates, the Syrians, Eritreans, and Iraqis. (Commission, European Solidarity: A Refugee Relocation System, 2016)

Following the EU Leaders' Summit with Turkey on 29 November 2015, the EU-Turkey Action Plan was adopted. The voluntary admission scheme proposed by the Commission on 15 December 2015 is a key element of the Plan, aimed at supporting Turkey in managing refugees and offering a safe and legal channel for persons in need of protection. (Commission, Press Release Database, 2016)

The EU Turkey Statement of 18 March 2016 provides that for every Syrian being returned to Turkey from the Greek islands, another Syrian will be resettled from Turkey to the EU. This

principle applies as of 4 April 2016. Priority is given to migrants who have not previously entered or tried to enter the EU irregularly. (Commission, Press Release Database, 2016)

Following the EU-Turkey Statement, the Commission adopted a proposal on 21 March to make 54,000 places not yet allocated out of the 160,000 foreseen for relocations available for the purpose of resettling Syrians from Turkey to the EU. (Commission, Press Release Database, 2016)

Table 1 shows the relocations from Greece to the Member States by 11 July 2016, according to the Communication from the Commission to the European Parliament, the European Council and the Council.

Table 1 Relocations from Greece to the Member States by July 2016

Member State	Formally pledged³	Effectively relocated	Commitment legally foreseen in the Council Decisions
Austria ⁴			1491
Belgium	200	90	2415
Bulgaria	210	6	831
Croatia	10		594
Cyprus	65	35	181
Czech Republic	30	4	1655
Estonia	93	27	2014
Finland	440	217	1299
France	2170	810	12599
Germany	240	37	17209
Hungary			988
Iceland			
Ireland	130	38	240

³ Transmitted via DublinNet under Article 5(2) of the Council Decision.

⁴ Council Implementing Decision (EU) 2016/408 of 10 March 2016 on the temporary suspension of the relocation of 30% of applicants allocated to Austria under Decision (EU) 2015/1601 establishing provisional measures to the area of international protection for the benefit of Italy and Greece.

Member State	Formally pledged⁵	Effectively relocated	Commitment legally foreseen in the Council Decisions
Latvia	129	39	295
Liechtenstein			
Lithuania	380	34	420
Luxemburg	100	71	309
Malta	24	24	78
Netherlands	450	242	3797
Norway			
Poland	65		4321
Portugal	730	302	1778
Romania	775	62	2572
Slovakia	10		652
Slovenia	60	28	349
Spain	350	147	6647
Sweden ⁵			2378
Switzerland	30		
TOTAL	6,691	2,213	63,302

Table 2 shows the relocations from Italy by July 11th 2016 to the Member States, according to the same Communication.

Table 2: Relocations from Italy to the Member States by July 11 2016

Member State	Formally pledged ⁶	Effectively Relocated	Commitment legally foreseen in the Council Decisions
Austria			462
Belgium	30	29	1397
Bulgaria	140		471
Croatia	10	4	374
Cyprus	15	10	139
Czech Republic	20		1036
Estonia	8		125
Finland	380	180	779
France	300	181	7115
Germany	10	20	10327
Hungary			306
Iceland			
Ireland	20		360
Latvia	30	2	186
Liechtenstein			
Lithuania	40		251
Luxemburg	20		248
Malta	17	17	53
Netherlands	175	125	2150
Norway			
Poland	35		1861
Portugal	388	150	1173
Romania	540	6	1608
Slovakia			250
Slovenia	20	6	218
Spain	50	40	1676
Sweden ⁷	50	39	1388

⁶ Transmitted via DubliNet under Article 5(2) of the Council Decision.

⁶ Council Implementing Decision (EU) 2016/408 of 10 March 2016 on the temporary suspension of the relocation of 30% of applicants allocated to Austria under Decision (EU) 2015/1601 establishing provisional measures to the area of international protection for the benefit of Italy and Greece.

⁶ Council Decision (EU) 2016/946 of 9 June 2016 establishing provisional measures in the areas of international protection for the benefit of Sweden in Accordance with Article 9 of Decision (EU) 2015/1523 and Article 9 of Decision (EU) 2015/1601 establishing provisional measures in the area of international protection for the benefit of Italy and Greece.

Member State	Formally pledged ⁶	Effectively Relocated	Commitment legally foreseen in the Council Decisions
Switzerland		34	
TOTAL	2,428	843	34,953

Table 3 shows the Resettlement State of Play as of 11 July 2016, under 20 July 2015 Conclusions and under the 1:1 mechanism with Turkey, which is applicable since April 4th 2016.

Table 3: Resettlement State of Play as of 11 July 2016

Member State/ Associated State	Pledges made under the 20 July 2015 Scheme	Total resettled under the 20 July 2015 scheme, including the 1:1 mechanism with Turkey	Third country from which the resettlement has taken place
Malta	14	0	
Netherlands	1,000	366	Lebanon: 219; Jordan: 7; Turkey: 61 (out of which 56 under the 1:1 mechanism); Morocco: 1; Ethiopia:8; Kenya 70
Norway	3,500	1,098	Lebanon
Poland	900	0	
Portugal	191	12 ⁸	Turkey: 12 under the 1:1 mechanism
Romania	80	0	
Slovakia	100 ⁹	0	
Slovenia	20	0	
Spain	1,449	118	Lebanon:61; Turkey 57 under the 1:1 mechanism
Sweden	491	380 ¹⁰	Sudan: 85; Iraq: 10 Kenya: 6; Egypt:3; Lebano:3; Jordan:1; Turkey:272:272 (out of which 264 under the

⁸ Portugal resettled 39 refugees from Egypt in 2015 under the national program, outside of the 20 July 2015 scheme.

⁹ Slovakia has resettled 149 Assyrians outside of the 20 July 2015 scheme

¹⁰ Sweden resettled 1,900 people in 2015 under its national program, outside of the 20 July 2015 scheme.

Member State/ Associated State	Pledges made under the 20 July 2015 Scheme	Total resettled under the 20 July 2015 scheme, including the 1:1 mechanism with Turkey	Third country from which the resettlement has taken place
			1:1 mechanism
Switzerland	519	519	Lebanon: 431 Syria: 88
United Kingdom	2,200	1,864 ¹¹	Jordan, Lebanon, Turkey, Egypt, Iraq and other countries based o humanitarian need.
TOTAL	22,504	8,268	A total of 802 people were resettled from Turkey under the 1:1 mechanism; 764 of whom through the scheme of July 2015

According to the latest press release by the European Commission on the Relocation and Resettlement process, published in July 13th 2016, and based on the information received by the participating States, 8,268 persons have been resettled by July 11th 2016 under the resettlement scheme of July 20th 2015, mainly from Turkey, Lebanon and Jordan. They have been received by 20 resettling states, Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Iceland, Ireland, Italy, Liechtenstein, Latvia, Lithuania, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom. (Commission, Press Release Database, 2016)

2.7 United States of America

The United States offer humanitarian assistance in an attempt to support the operations of the United Nations, other international organizations and non-governmental organizations. Through these Organizations the United States provide assistance in all 14 governorates of Syria.

The Bureau of Population, Refugees and Migration issued a report for the proposed refugee admissions for the fiscal Year 2017, in September 2016, in which it claims that in Fiscal Year 2017 it aims to surpass the number of 10,000 Syrian refugees that the United States admitted in 2016.

¹¹ Under the existing UK national resettlement schemes in 2015.

The United States is one of the 32 countries that have agreed to accept referrals from UNHCR as part of its ambitious international effort to secure permanent or temporary resettlement for up to 10 percent of Syrian refugees. (Bureau of Population, Refugees, and Migration, 2016, p. 5)

Table 4 shows the sums of humanitarian assistance offered by the United States to Organizations committed to the solution of the refugee issue.

Table 4: Humanitarian Assistance Funding Numbers by Organization

Organization	September 2016	Total Since FY 2012
NGOs	\$205 million	\$1.7 billion
UNHCR	\$51 million	1.5 billion
UNICEF	\$39 million	\$592 million
UNRWA	\$31 million	\$385 million
IOM	\$13 million	\$72 million
ICRC	\$9 million	\$221 million
UNFPA	\$ 8 million	\$47 million
WFP	\$6 million	\$1.3 billion
ILO	\$3 million	\$3 million
Other Organizations		\$106 million
Other (admin)	\$0.4 million	\$9 million
TOTAL	\$364 million	\$5.9 billion

As Table 4 shows the US has offered a sum of \$364 million of humanitarian assistance only in September 2016, whereas the total sum of donations since 2012 sums up to \$5.9 billion.

3 Twitter

3.1 Introduction

Microblogging is one of the most popular tools of internet connection today. Used by millions of users, microblogging sites such as Facebook, Twitter, Tumblr, etc, give the opportunity to their users to share opinions on a variety of topics. The developing technology in the use and accessibility of those platforms as well as their free format leads to the shift of the use of traditional communications tools, such as blogs, or mailing lists to microblogging services. (Pak & Paroubek)

Twitter is a microblogging service that counts 313 million monthly active users from all over the world, the 82% of which are active users on mobile. (Twitter, Twitter Web site, 2016)

It allows users to type messages up to 140 characters, which are also referred to as tweets. (Twitter, Twitter Support, 2016) A hashtag – written a # symbol – is used to index keywords or topics on Twitter. This function enables persons to easily follow topics they are interested in. Popular hashtags can as well become trending topics. (Twitter, Twitter Help Center, 2016) One can publish tweets via SMS, email-s, through the web and directly from smartphones, using a wide variety of web based applications. (Twitter, Twitter Help Center, 2016)

3.2 Information credibility on twitter

Twitter prompts users to answer the question “What are you doing?” creating a constantly-updated timeline, or stream of short messages that range from humor and musings on life to links and breaking news. Therefore, twitter makes possible the propagation of real time information to a large group of users. This makes it an ideal environment for the dissemination of breaking news directly from the news source and/or geographical location of events.

This fact raises the question of credibility in the tweets one publishes. In general, the perception of users with respect to the credibility of news seems to be positive. People trust the Internet as a news source as much as other media. Twitter is widely used as a news media and has often been used to track epidemics, detects news events, geolocate such events and find controversial emerging topics. The Information Architecture of Twitter, which allows real time propagation of

tweets, allows immediate publish during wildfires, hurricanes, floods and earthquakes. (Park, 2013)

Twitter has a special interest for the analysts of political communication in engagement. This is because of the “asymmetrical” model of relationships between users that Twitter enables, instead of the “symmetric” model that other social media like Facebook encourage. Instead of the two-way relationship of Facebook, Twitter allows a four-way relationship, referring to people one follows but they do not follow back, people that follow one person, but the person does not follow back, both sides follow each other, none of the persons follow each other.

As it can be pointed out, this asymmetric model gives the opportunity of more types of relationships.

The fact that Twitter provides its users the capability of asymmetrical networking, is the reason why Twitter is perhaps more conducive to political interaction than other social media that only allow a symmetrical interaction. (Porter, 2009)

Twitter can be a fruitful soil for opinion leadership formation. It could be said that opinion leadership in twitter does not abide by the innovation diffusion theory norms, that see opinion leaders more exposed to all forms of external communication, to be persons of a higher socioeconomic status, more innovative and at the middle of interpersonal communication networks. Opinion leaders on Twitter tend to depend more on their own expertise and perspectives rather than on their social positions. That means that the level of visible education or established social standing may not matter much in acting as an opinion leader on Twitter. What plays an important role in the influential power of political leaders in Twitter instead, is their informal status as highly connected individuals. According to the social media monitoring company Beevolve, Twitter users on average have 208 followers and follow 102 users themselves, which is amazing considering the everyday relationship boundary of ordinary people. (Park, 2013, p. 1642)

In addition to this, it is more likely that opinion leaders be involved in a ‘multi-step flow’ process, as opposed to the traditional ‘two-step flow’ process. While the two-step flow process is a process of information moving from the media to opinion leaders and influence moving from opinion leaders to their followers, a multi-step flow indicated that a message distributed through a myriad of intermediary channels. On Twitter those who are well-connected play a more potent role in creating and distributing information through a multi-step flow than those with less connection. (Park, 2013, p. 1642)

The question emerging is at which level one can trust tweets as a valuable source of information and how one could evaluate credible tweets.

Today, major search engines choose to display search results from real-time web, especially for trending topics. This may result in the exploitation of this function from spammers, in order to attract visitors in their websites. It has also increased the potential impact of orchestrated attacks spreading misinformation on purpose. In this way, Twitter, is often used as a means for political propaganda. There is also the factor of the unwilling spread of information, when hackers attack official websites in order to publish misled information. (Castillo, Mendoza, & Poblete, 2011)

It is possible to distinguish four different features in the evaluation of the credibility of the tweets.

One is the message-based feature, considering certain characteristics of the message. These features can be either twitter-independent, or twitter-dependent. The first ones include characteristics such as the message length and the number of positive or negative words in a message. (Castillo, Mendoza, & Poblete, 2011)

The user-based features include characteristics of the users who post tweets. These characteristics are registration age, the number of followers and the number of the tweets the user has authored in the past. (Castillo, Mendoza, & Poblete, 2011)

Topic-based features refer to aggregates computed from both of the two abovementioned features. These could be the number of tweets with hashtags, the number of tweets of positive and negative sentiment in a set. (Castillo, Mendoza, & Poblete, 2011)

The fourths feature is the propagation-based one that considers characteristics related to the propagation of tweets that can be built through retweets. (Castillo, Mendoza, & Poblete, 2011)

It can be said, that in the framework of the topic based feature, a tweet can be considered credible when it includes a URL. In addition to this, the sentiment-based features, like sentiments fraction seem to play an important role for the tweet's credibility. In general, tweets that include negative sentiment terms are related to credible news. (Castillo, Mendoza, & Poblete, 2011)

A way to evaluate the credibility of tweets in the user-based features, is the number of tweets that one user has propagated. A low number usually shows small credibility. The number of followers also plays a role in the feature. (Castillo, Mendoza, & Poblete, 2011)

Tweets with many retweets tend also to be more credible, with regards to the propagation feature. (Castillo, Mendoza, & Poblete, 2011)

3.3 Twitter audience

Social media technologies collapse multiple audiences into single contexts, making it difficult for people to use the same techniques online that they do to handle multiple face-to-face conversations. (Marwick & Boyd, 2016, p. 114)

In every mediated conversation, whether through instant messaging or comments, participants have a sense of audience. This audience is often imagined and constructed by the individuals in order to present themselves appropriately, based on technological affordances and immediate social context. The perplexity of the current social media, that allows the collapse of multiple contexts and the gathering of commonly distinct audiences, make it even more complicated for the individual to self present himself. (Marwick & Boyd, 2016)

It is understandable that Twitter or Facebook audience is limitless, but individuals often act as if it were bounded. As it is used to in microblogging sites, a tweet's actual audience may differ from the producer's imagined audience. One typical characteristic of Twitter is the capability of having a public or private account, which lengthens or strengthens the number of the audience it refers to. A second one is the capability of retweeting that gives the chance of multiplying one's tweet audience. The various ways that people can consume and spread tweets, makes it virtually impossible for Twitter users to account for their potential audience, let alone actual readers. In general, as it happens with all microblogging sites, the tweets are read by few people, whom the Twitterers do not know. This is the reason why they imagine it. (Marwick & Boyd, 2016)

A research conducted by "New Media and Society" in an attempt to find out how Twitter users imagine their audience, (Marwick & Boyd, 2016) proved that the majority of users conceptualized their audience focusing on abstract categories of people, whereas fewer ones indicated that their audience was articulated through the service itself. The first ones with a small number of followers, referred to their audience as friends, with friends being individuals they followed, or IRL friends, for friends they know outside of Twitter. (Marwick & Boyd, 2016, p. 118) Another approach some respondents took was to conceptualize their audience as the ideal person, who will presumably share their perspective and appreciate their work. (Marwick & Boyd, 2016, p. 120) The second ones define their audience as the overall between followers and following (Marwick & Boyd, 2016, p. 118) , or suggest that audience conceptions are tweet-dependent, thus meaning users write different tweets to target different people. (Marwick & Boyd, 2016)

The following chapter will examine a number of characteristics related to two hashtags which emerged after the Syrian refugee crisis and are considered to be pro or contra the refugees' acceptance.

4 Knime analytics platform tools

4.1 Introduction

This dissertation aims to examine two different hashtags, the “#RefugeesWelcome” and the “#StopRefugees. Data regarding the number of rows (tweets) each hashtag returns, the user’s locations, the terms more frequently used in those tweets and the number of tweets per year are retrieved and analyzed.

The examination is conducted with the use of the latest updated KNIME 3.3.1 Analytics Platform, released in December 6, 2016.

In the following chapter, a reference on the Knime’s tools used for the current research is held.

4.2 Knime

Knime is an open source data analytics, reporting and integration platform. It integrates various components for machine learning and data mining through its modular pipelining concept. A graphical user interface allows assembly of nodes for data preprocessing.

The Konstanz Information Miner, as the abbreviation reads, offers a large set of building blocks and third party tools. Files are stored in the workspace, which contains the workflows. Workflows might be programs or processes that describe the steps applied to load, visualize or transform the data.

The architecture of Knime is designed on the basis of three main principles.

The visual interactive framework allows data to be manipulated by single drag and drops from a variety of processing units. Customized applications may as well be modeled through individual data pipelines.

The modularity of the various processing units and data containers do not require a mutual dependence, enabling thus an easy distribution of computation and allowing an independent development of different algorithms. Data types are already embodied, that is, types are not

predefined, whereas new types can easily be added bringing along type specific renderers and comparators.

It is, as well, easy to add new processing nodes or views and distribute them through a simple plug&play principle without the need for complicated install/uninstall procedures. (Berthold, p. 1)

The above is achieved through a pipeline of nodes, connected by edges that transport data or models, in order for the data analysis to process. Each node processes the arriving data and produces results to its outputs. (Berthold, p. 2)

4.2.1 Repository

Knime offers a big variety of nodes; among them are nodes for various types of data Input/Output, manipulation and transformation, as well as data mining and machine learning and visualization components.

Workflows may as well contain nodes, metanodes, connections, workflow variables, workflow credentials and annotations. (Berthold, p. 10)

4.2.2 Nodes

Nodes in Knime are the most general processing unit and usually resemble one visual node in the workflow. The class Node wraps all functionality and makes use of user defined implementations of a NodeModel, a NodeDialog and one or more NodeView instances. The dialog and view settings are only implemented when demanded by the user settings. This schema follows the Model-View-Controller design pattern. Moreover, each node has a number of Input and Output instances, which can transport either data or models. (Berthold, p. 2)

The Node Repository of the KNIME Workflow offers a range of nodes in order to extract and process with the data manipulation. The Nodes used in the current dissertation include the Twitter API Connector, the Twitters Search, the Row Filter, the Group By, the Line Chart (JFreeChart), the Excel Writer (XLS), the Excel Reader (XLS), the String to Date/Time, the Strings to Document, the Punctuation Erasure, the Number Filter, the N Charts Filter, the Stop word Filter, the Case converter, the Snowball Stemmer, the Bag of Words Creator, the Term to String and the Row Filter Nodes.

4.2.3 Meta nodes

The ability to wrap a certain workflow into an encapsulated node provides an abstraction mechanism. A sub workflow can be included as a single component of another workflow, namely a meta-node. (Berthold, p. 5)

A metanode used in the current dissertation can be seen in Figure 3.

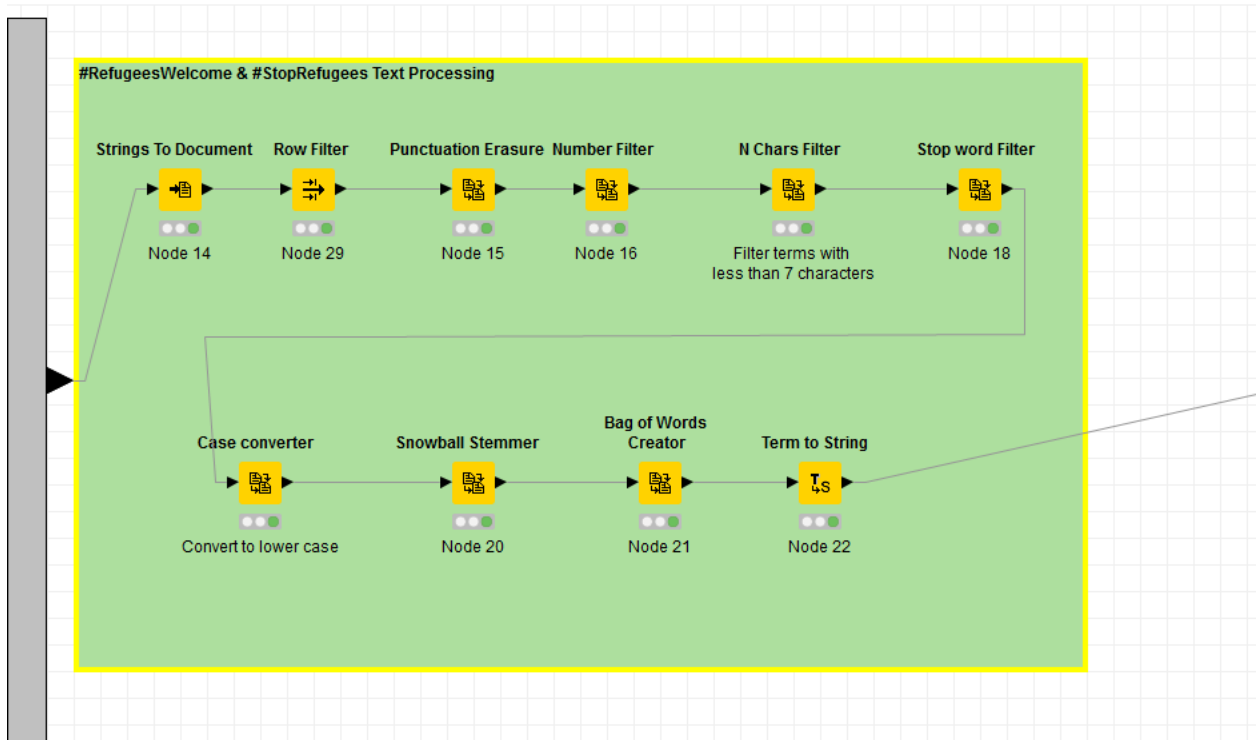


Figure 3: Nodes consisting the Metanode “#RefugeesWelcome & #StopRefugees Text Processing”

4.2.4 Workflow management

Workflows in Knime are graphs connecting nodes, or a direct acyclic graph (DAG), as it is referred to in the computing language. The Workflow Manager allows the insertion of new nodes and the addition of connections between two nodes. It keeps track of the status of nodes (configured or executed) and returns a pool of executable nodes on demand. The underlying graph structure allows the workflow manager to determine all nodes required to be executed along the paths leading to the node the user actually wants to execute.

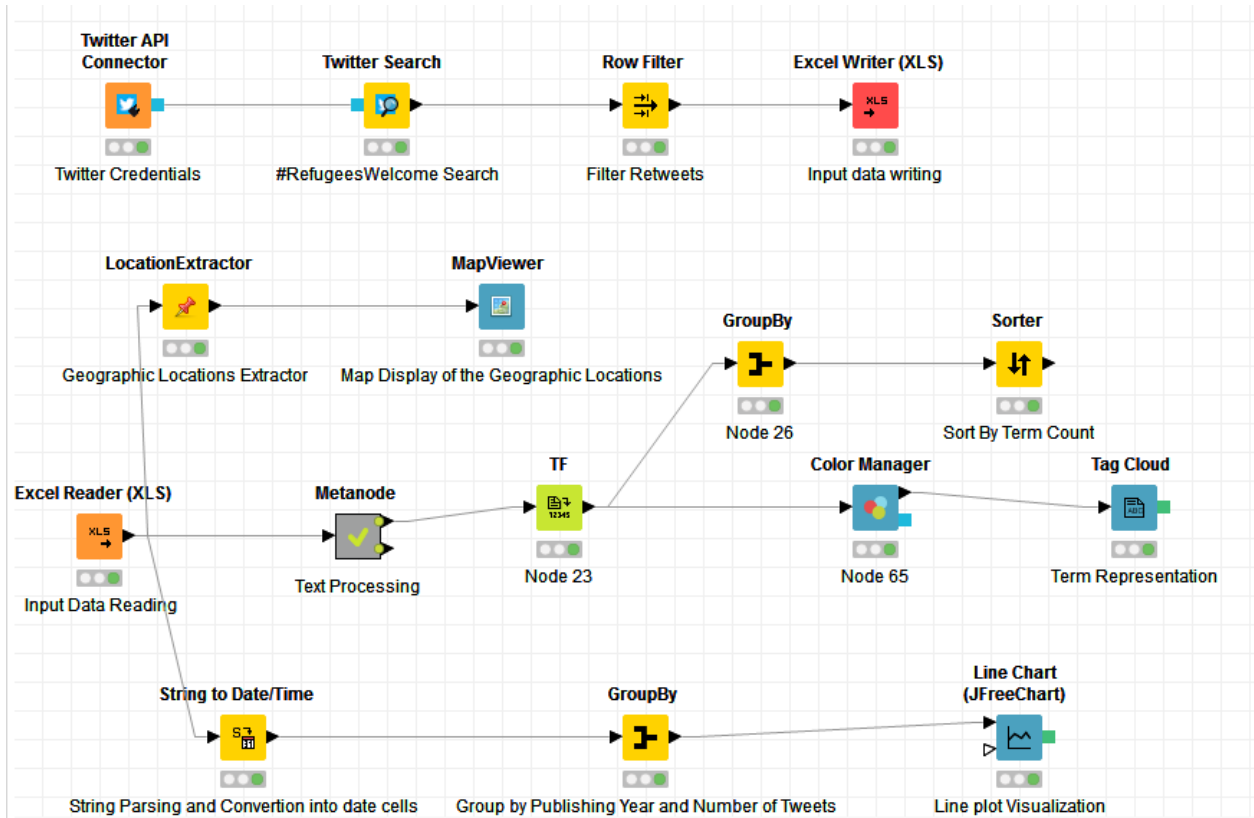


Figure 4: #RefugeesWelcome Workflow

Figure 4 shows an example of the #RefugeesWelcome workflow used for the current research. The Workflow includes all nodes used for the data mining and visualization. A description at the bottom of every node describes the demand it is set to implement.

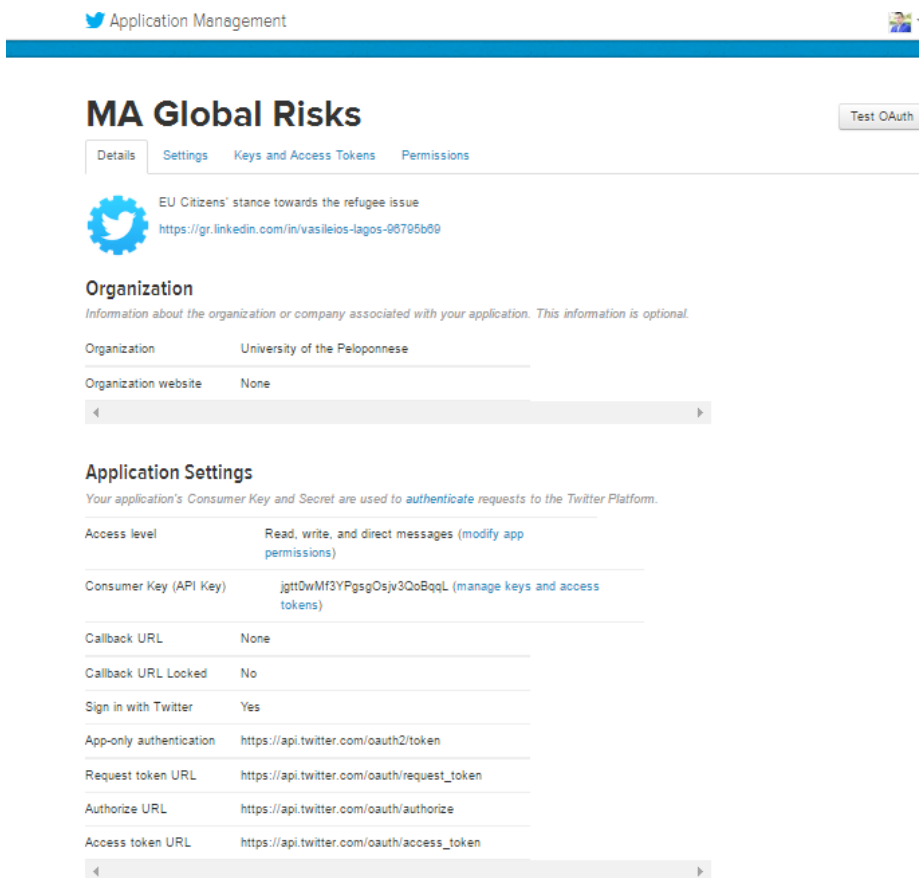
5 #RefugeesWelcome and #StopRefugees search

5.1 Introduction

The following chapter examines the procedure followed for the retrieval of tweets. The process for the twitter application management, the twitter search and the exclusion of retweets is explained. A reference on the number of tweets mined is being held as well.

5.2 Creation of the twitter application

In order to get access to twitter's public data stream, there is a need for a developer account, which can be retrieved through the <https://dev.twitter.com> website. Twitter offers its users the capability of creating and manipulating apps.



The screenshot shows the 'Application Management' interface for a Twitter application named 'MA Global Risks'. The page is divided into several sections: 'Details', 'Settings', 'Keys and Access Tokens', and 'Permissions'. The 'Details' section shows the application name, a profile picture, and the organization 'EU Citizens' stance towards the refugee issue'. The 'Organization' section shows the organization name 'University of the Peloponnese' and its website 'None'. The 'Application Settings' section shows the access level 'Read, write, and direct messages (modify app permissions)', the consumer key 'jgtt0wMf3YPgsgOsJv3QoBqQL', and various URLs for authentication and token management.

Access level	Read, write, and direct messages (modify app permissions)
Consumer Key (API Key)	jgtt0wMf3YPgsgOsJv3QoBqQL (manage keys and access tokens)
Callback URL	None
Callback URL Locked	No
Sign in with Twitter	Yes
App-only authentication	https://api.twitter.com/oauth2/token
Request token URL	https://api.twitter.com/oauth/request_token
Authorize URL	https://api.twitter.com/oauth/authorize
Access token URL	https://api.twitter.com/oauth/access_token

Figure 5: Application Credentials

Figure 5 shows the app created for the current thesis, under the name “MA in Global Risks”.

Having completed this step, the application is ensured Twitter credentials.

The second step is the setup of the Twitter API Connector. This Node creates a connection to access Twitter’s API. In this step, the API keys and Access tokens provided by the developer account are filled in, as seen in Figure 6 in order to give the Knime Workklow the capability to connect with the Twitter account.

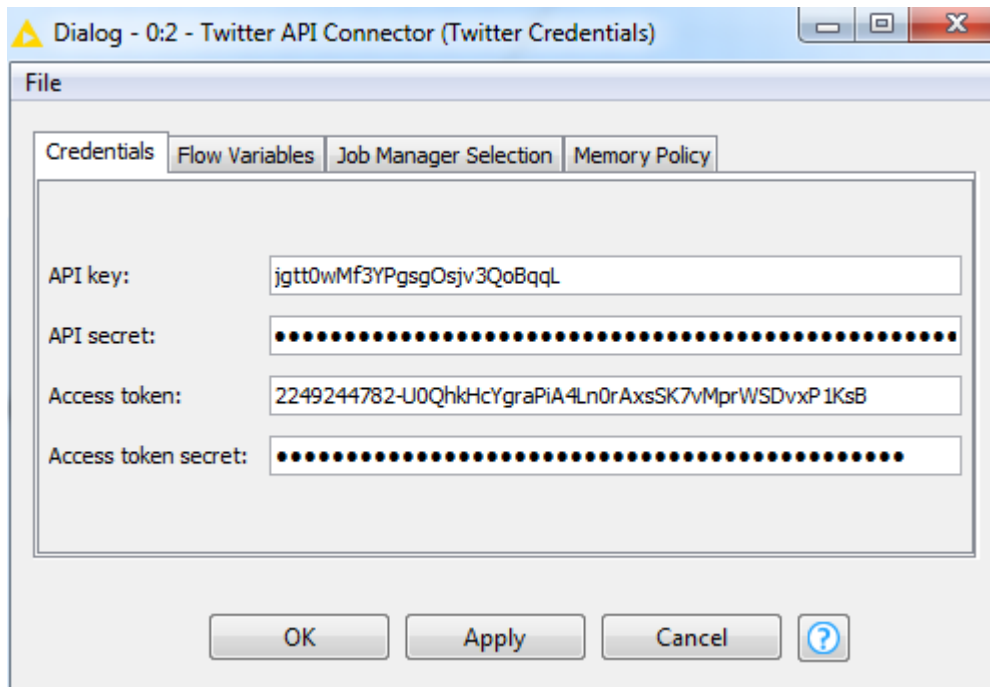


Figure 6: Twitter Credentials

The third step includes the search of twitter data. Knime offers different ways to access Twitter data, either by looking for past Tweets matching the search terms, or live tweets via Twitter’s Streaming API.

Due to the needs of the research, the “Twitter Search” Node is used, as it offers the change of searching past tweets.

5.3 #Refugees Welcome and #Stop Refugees tweets mining

In order to extract rows (tweets) from Twitter, the “Twitter Search” Node has to be connected to the “Twitter API Connector” Node.

The Twitter API accepts a determined amount of requests in 15 minute workflow. The amount of request the Twitter API accepts is limited 450 regarding application authentication in a 15 minute window. Since one request can return a maximum of 100 tweets, the maximum amount of tweets that can be pulled in a 15 minute window is 45,000.

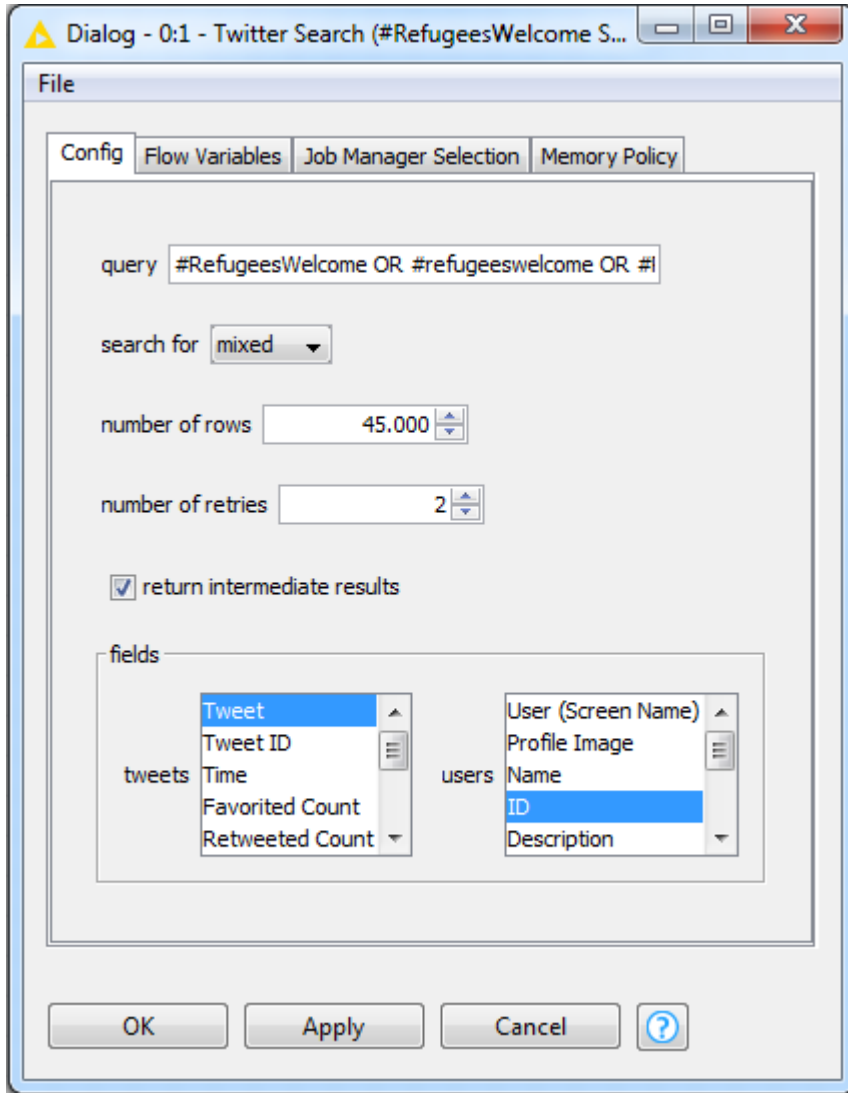


Figure 7: Twitter Search Dialog

For the research to be as even as possible, different workflows have been used for the tweets' search regarding each hashtag. That means, that in total four workflows are used; one for the #RefugeesWelcome hashtag a second one for the popular tweets regarding the same hashtag, a third one for the #StopRefugees hashtag and a fourth one for the popular tweets of the same hashtag.

In order to receive the biggest amount possible of tweets related to the #RefugeesWelcome hashtag, the query includes a number of different variations of the phrase. The hashtags

#RefugeesWelcome OR #refugeeswelcome OR #REFUGEESWELCOME OR #Refugees_Welcome OR #refugees_welcome are used in the query.

The same procedure is held for the query regarding the #Stop Refugees hashtag, where the #stopRefugges OR #StopRefugees OR #stoprefugees OR stop_refugees variations are used.

The dialog offers a variety of information to be retrieved regarding the tweets' and users' fields.

With regards to the "tweets" field, the information "Tweets", "Is Retweet", "Location" and "Country" are requested. With regards to the users' field, the information "ID", "Creation Time", "Language", "Location" and "Time Zone" are requested.

The number of rows (tweets) requested, for each query are 45,000. The number of rows (tweets) returned for the #RefugeesWelcome hashtag are 10,330, whereas the number of rows returned for the #StopRefugees hashtag are 4063.

5.4 Exclusion of retweets

For the research to be as accurate as possible, the retweets need to be excluded from the search. It is noticed that, even when the retweets are not selected in the Tweets information dialog, the "Twitter Search" Node's Table returns them.

For this reason the "Row Filter" Node is used, which is set to exclude the rows containing retweets.

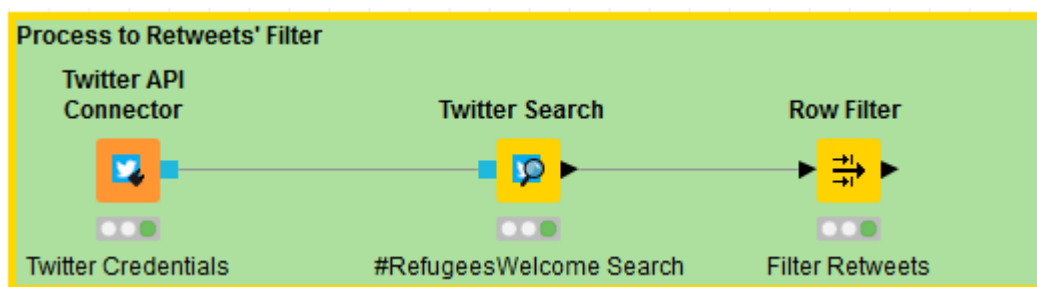


Figure 8: Filter Retweets Process

In order for the retweets to be excluded, the retweets column will be the one to be tested in the Node dialog, as seen in Figure 9.

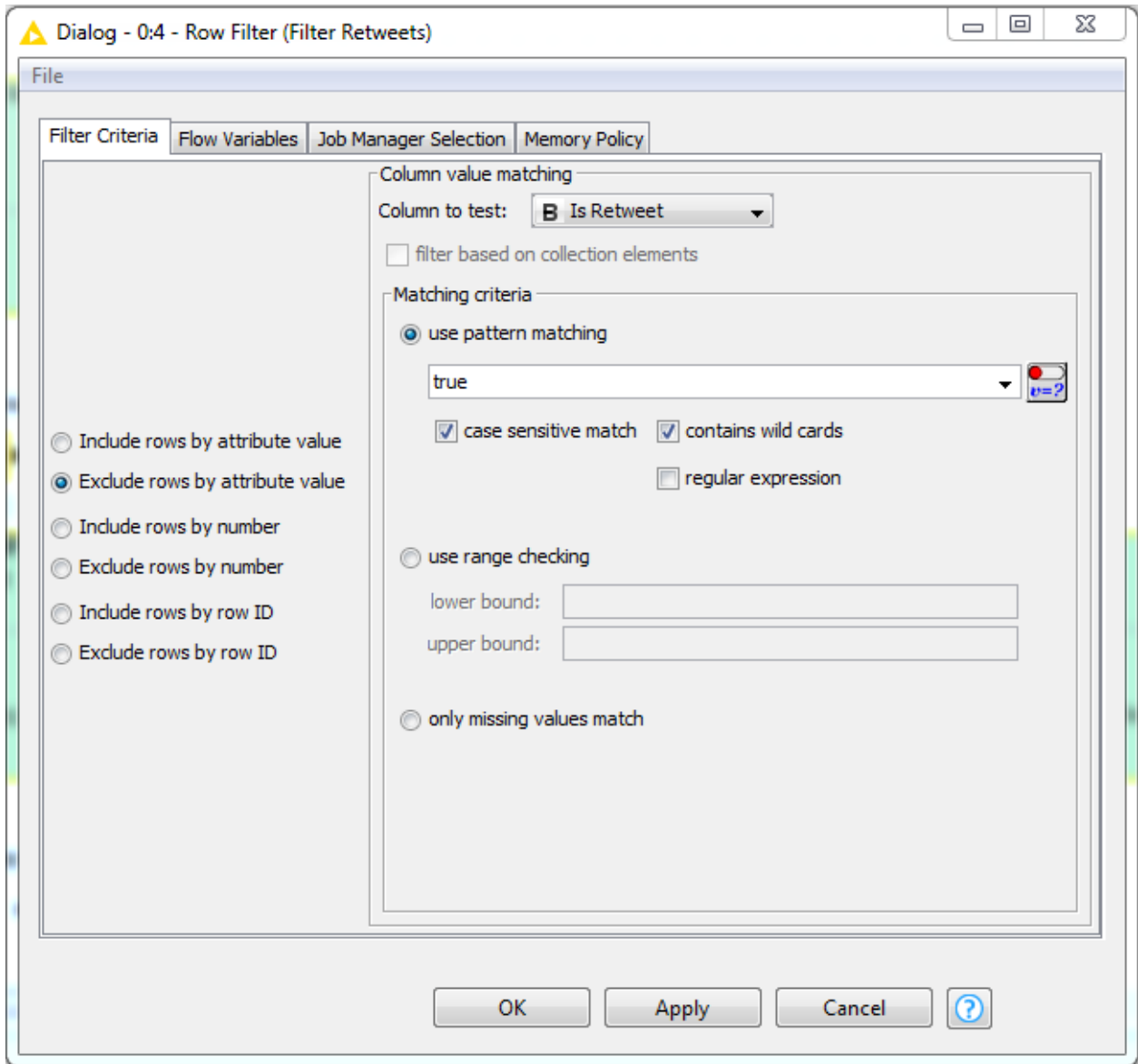


Figure 9: Filter Retweets Dialog

As Figure 9 shows, the column to be tested in the “Row Filter” Node is the “Is Retweet” one, whereas the use pattern matching is “true”, as the Filter Criteria shall exclude rows when the retweet value is true. For the exclusion to be as accurate as possible, the case sensitive match and the wild cards are also summoned.

After the Retweets are excluded, the remaining rows (tweets) for the hashtag “#RefugeesWelcome” are 3,083. This means that 10,027 from the initial returned rows are retweets.

The remaining rows (tweets) for the “#StopRefugees” Hashtag are 1,494. This means that 2,569 from the initial returned rows are retweets.

From the above, it can be pointed out that more people tweet about #RefugeesWelcome than #StopRefugees.

In the following chapter the most significant terms related to each hashtag will be retrieved and examined.

6 Significant terms extraction

6.1 Introduction

The hypothesis of the dissertation is that users tweeting for the #RefugeesWelcome hashtag would be more positive adjacent towards the refugees, whereas the ones who tweet for the #StopRefugees hashtag would be negatively adjacent towards the issue.

The significant terms extraction of both hashtags will be useful in order to examine the terms used when referring to each hashtag. The hypothesis is that terms mostly used by the users of each hashtag, will be an important helpful hand in realizing the users' sentiments towards the refugee issue.

In order for the terms to be extracted, the text preprocessing feature needs to be applied.

The following chapter will describe the process applied for the terms extraction.

6.2 Text preprocessing

The Knime text preprocessing feature is designed and developed to read and process textual data and transform it into numerical data in order for regular KNIME data mining nodes to be applied. The feature allows the parsing of texts available in various formats (e.g. Xml, Microsoft Word or PDF and the internal representation of documents and terms) as KNIME data cells stored in data table. The documents can be filtered, stemmed and preprocessed in various ways. In addition to these, words can be computed, keywords can be extracted and documents can be visualized. (Thiel & Berthold, 2012, p. 3)

It is important that the right order of the nodes is followed, since they require specific structural specifications of the input data tables.

The Text Processing feature will be applied in order for the most important terms of every tweet to be extracted.

6.2.1 Reading text data

The data table retrieved by the “Twitter Search” Node needs to be extracted and written in a file that can be parsed and manipulated by the Nodes.

The Node selected is the “Excel Writer (XLS)” which writes the input data table into a spreadsheet of a workbook that contains the “Tweets”, “Is Retweet”, “Location”, “Country”, “ID”, “Creation Time”, “Language”, and “Time Zone” information.

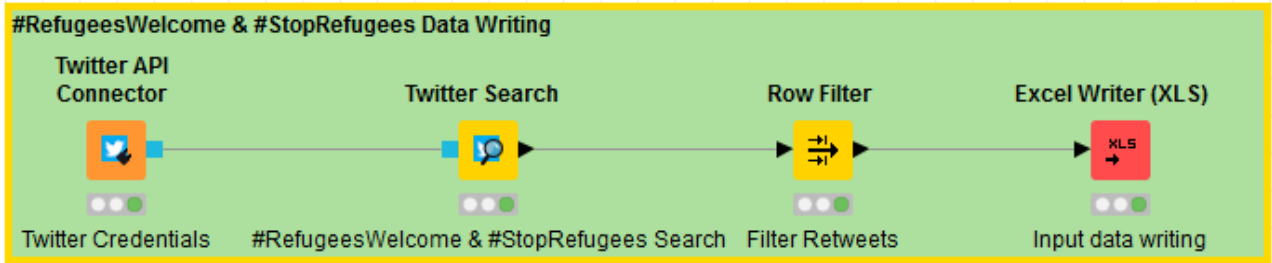


Figure 10: #RefugeesWelcome and #StopRefugees Hastags Input Data Writing

6.2.2 #RefugeesWelcome and #StopRefugees input data preprocessing

The textual data is preprocessed by various nodes provided by the Knime Text Preprocessing extension. All preprocessing steps are applied in the “Preprocessing” Metanode as shown in Figure 11.

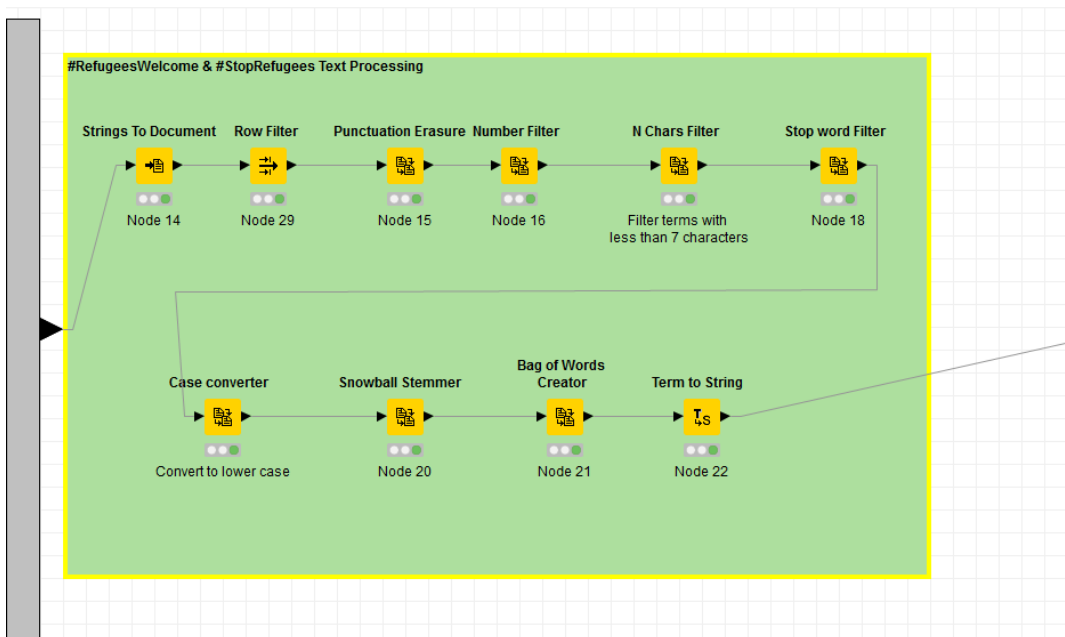


Figure 11: #RefugeesWelcome and #StopRefugees Hashtags Text Preprocessing

The node “Strings to Document” needs to be applied, in order for a Document Cell for each row to be created. Each column that contains the data for each field of the document is specified in the dialog of the node.

In order to process the data manipulation, the terms need to be preprocessed. The preprocessing step includes filter of the terms in order to get rid of terms that do not contain content. Terms such as stop words, number, punctuation marks, or very small words are filtered. Terms are also filtered to remove endings, based on declination or conjugation by applying stemming.

The “Punctuation Erasure” Node removes all punctuation characters of terms contained in the input document. For the preprocessing procedure to be clearer, terms with unmodifiable flags are preprocessed as well.

Terms contained in the input document that consist of digits, containing decimal separators and possible leadings are being removed by the “Number Filter” Node, whereas the “N Charts Filter” Node is set in order to filter terms that contain less than 7 characters.

The “Stop Word” Node filters stop words included in the input document as specified in the built-in stop word list and the “Case Converter” Node is set to convert all terms contained in the input document to lower case.

The “Snowball Stemmer” Node stems terms contained in the input document, in accordance with the Snowball Stemming library.

After the document is preprocessed, the terms are converted into strings, so that they be further manipulated. At this point, two different “Row Filter” Nodes are used, in order for the data to be further cleaned up. The first node excludes terms containing URLs, whereas the second one terms containing replies.

Having applied these changes to the input data, the preprocessing procedure is finished. In the following chapters this procedure will be the starting point for the mining of the most important terms related to each hashtag, as well as the location, the dates published and the number of tweets per year extraction.

6.3 Significant terms extraction

In order for the mostly used terms to be extracted, the “Term Frequency” Node computes the term frequency of each term according to the document and adds a column containing the tf value. This

term value is computed by dividing the absolute frequency of a term according to the document by the number of all terms of that document.

The Term Frequency measures how frequently a term occurs in a document. Since every document is different in length, it is possible that a term would appear more times in a long document than in shorter ones. The term frequency is thus divided by the document length, the total number of terms in the document, as a way for normalization of the search's results.

6.3.1 #RefugeesWelcome significant terms' extraction

In order for the terms most frequently used in both hashtags to be counted, they have to be grouped by the "Term as String" and "Term" columns. Subsequently, the "Sorter" Node sorts the rows according to the "Term as String" Node in a descending order.

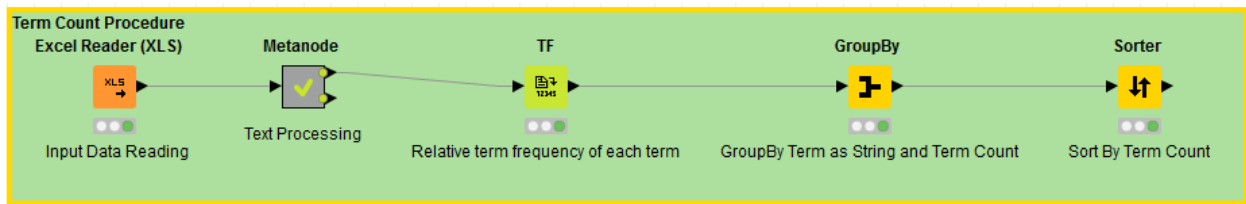


Figure 12: Term Count Procedure

Terms most frequently used related to the #RefugeesWelcome hashtag are shown in Table 6.

Table 6: Refugees Welcome Hashtag's most frequently used term count

Term	Term Count
Refugees	2170
Welcome	1841
#refugeeswelcome	737
Bluehand	381
Britain first	380
Refugee crisis	71
Support	67
Womens' march	60
Refugee action	47
Migrants	75
Treated	43
Shame on Europe	40
children	38
no borders	28
Brexit	25
Germany	25
with refugees	23
America	22

Term	Term Count
Berlin attack	22
Country	22
Remain	21
Socialism	21
Islamophobia	21
Terrorist attack	21
Borders	20
Muslims	20
religion of peace	20
Resettlement	19
Bridges not walls	17
Community	17
Refugees chief	17
Solidarity	17
Helping	15
No nazis	15
Families	14
Justice	14
Campaign	13
Fortress Europe	13
Freezing	13
Genocide	13
Human rights	13
Protest	13
Education	12
Here to stay	12
Students	12
Demonstrate	11
Remember	10
Tomorrow	10
Why I march	10

6.3.1.1 “Refugees Welcome” hashtags’ tag cloud visualization

The Tag Cloud will facilitate the visualization of the most important terms related to the hashtag. Figure 12 depicts the procedure to the Tag Cloud Visualization.

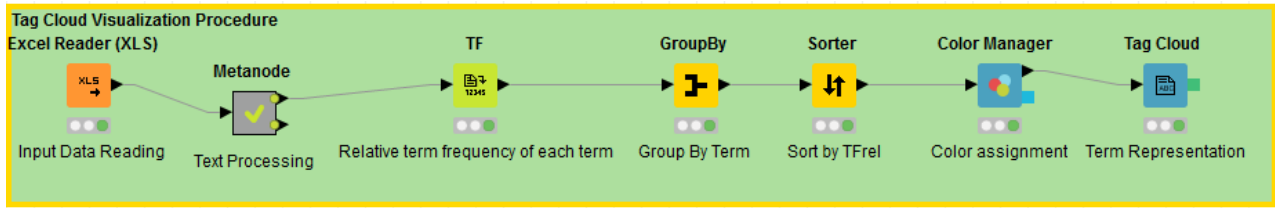


Figure 13: Tag Cloud Visualization Procedure

The “Group by” Node is set to group the input data by term, in order for distinct terms to be extracted. In order for the “Tag Cloud” Node to visualize the terms, they need to be sorted by their term frequency.

The Tag Cloud has been configured to display 200 rows out of the 7,111 output by the sorted table. Figure 14 depicts the Tag Cloud of the #RefugeesWelcome hashtag. The terms less frequently used are the more transparent ones, whereas the ones in bolder characters are the ones more frequently used.



Figure 14: #Refugees Welcome Tag Cloud Visualization

The “Color Manager” Node identifies the minimum term value to be equal to 0.083 and the maximum value equal to 1.0. The color assigned for the terms of the minimum value is the grey one, whereas the maximum value has been assigned a black color. The terms are accredited a gradient color, meaning that different hues of both colors are accredited, depending on the term frequency.

6.3.2 #Stop Refugees significant terms’ extraction

The procedure applied for the #RefugeesWelcome hashtag’s terms extraction, is also applied to the ones related to the #StopRefugees one.

Table 6 shows the terms most frequently used in the tweets, after the Term Frequency Node has computed the relative term frequency.

Table 6: Stop Refugees Hashtag’s most frequently used term count

Term as String	Term Count
Refugees	931
Children	340
Weather	331
Freezing	189
Country	74
Illegal	76
Migrants	32
Stop refugees	27
Terrorists	21
America	26
Immigrants	48
Muslims	19
Funding	17
Accepting	15
Amnesty	15
Welcome	15
Germany	14
Stop islam	12
Supporting	11
Borders	10
Importing	10
Threatening	9
Islamic	8

6.3.1.2 “Stop Refugees” hashtags’ tag cloud visualization

The Tag Cloud has been configured to display 200 rows out of the 2,300 output by the sorted table. Figure 15 depicts the Tag Cloud of the #RefugeesWelcome hashtag.

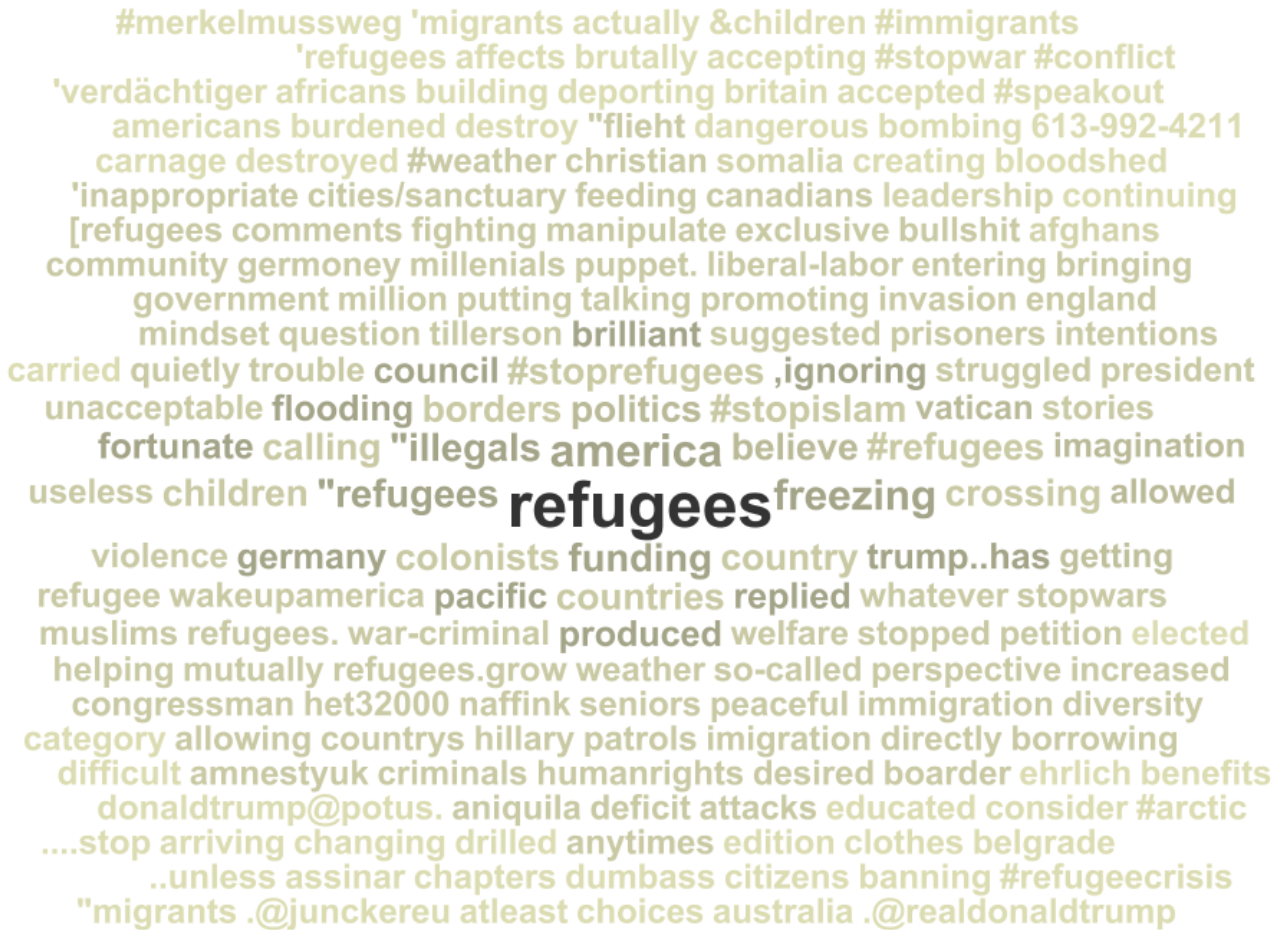


Figure 15: #Stop Refugees Tag Cloud Visualization

The “Color Manager” Node is set to attribute a grey color to the minimum value, which has a rank of 0.090 and a black color to the maximum value ranking at 1.0.

6.4 Conclusions

It is difficult to ensure or dismiss the initial hypothesis that users tweeting about #RefugeesWelcome would be more positively adjacent towards the refugee issue compared to the ones tweeting about #StopRefugees.

There are several terms in common in tweets of both hashtags, and the users’ disposition cannot be determined just by examining the terms themselves. A future sentiment analysis of the tweets’

content would better examine the users' disposition towards the refugee issue, tweeting either for #RefugeesWelcome or #StopRefugees.

7 Tweets' location extraction

7.1 Introduction

Another aspect that this dissertation attempts to examine is the location where the Twitter users tweet from. The results of this query will show the geographical locations where the #RefugeesWelcome and the #StopRefugees related tweets are located, and will be a helpful hand in realizing how users from around the world are disposed towards the refugees.

Twitter gives the choice to its users to add a location of their tweets using Twitter for Android, Twitter for iOS, twitter.com or other mobile applications.

In order for a user to tweet with a location, one can choose the one proposed by Twitter according to his current coordinates or he can select one from the ones listed in a dropdown menu. Twitter users may as well add additional location context to their Tweets, such as a general location label, or a name of a specific business, landmark or other point of interest. (Twitter, Twitter Help Center, 2017)

It has to be pointed out that, due to these facts, the locations which users have registered their tweets at, may not be precise or true.

The "Location Extractor" Node allows the extraction of geographic locations from unstructured text, using Palladian's location extraction mechanism. In order to use the LocationExtractor, a Location Source must be configured.

Since the Twitter search provides already the User-Location source, a geonames-based Location Source needs to be set up. Knime currently offers a location source for the GeoNames API with the Palladian community contribution. In order to get access to the Location Source, there is a need for a GeoNames account, which one can get following the [KNIME → Palladian Location Extractor preferences page](#).

The GeoNames API provides a database with real world locations and meta-information such as alternate names, population figures, coordinates, and hierarchical relations. The location extraction algorithm performs various steps for recognizing potential locations within a given text, followed

by a disambiguation. The disambiguation step checks hierarchical relations and identifies locations by their proximity to other given locations in the text.

Each identified location in the text is returned, multiple occurrences are returned as often as they occur, whereas extracted locations are classified into different categories.

Figure 16 shows the process for the visualization of the geographic locations.

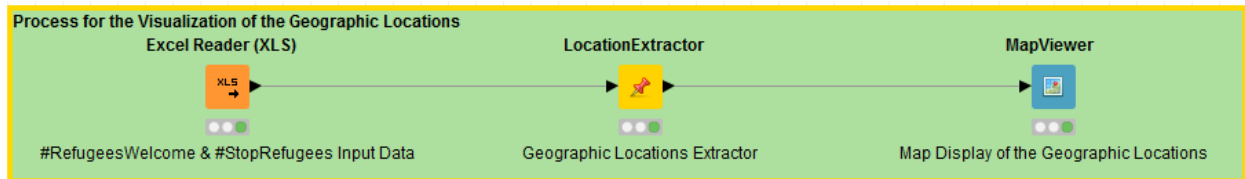


Figure 16: Process for the Visualization of the Geographic Locations

7.2 #RefugeesWelcome tweets' location extraction

Figure 17 depicts the distribution of the tweets under the #RefugeesWelcome hashtag by the countries they were published.



Figure 17: Refugees Welcome Hashtag Location Distribution

It can be seen that the majority of the tweets is located in Northern Europe, as well as in the United States of America.

7.3 #StopRefugees tweets' location extraction

Figure 18 depicts the map with the location of the tweets under the #StopRefugees hashtag.



Figure 18: Stop Refugees Hashtag Location Distribution

As it can be pointed out, the majority of the tweets are located in Northern Europe and the United States, whereas the distribution is sparser than that of the #RefugeesWelcome hashtag. It can also be pointed out that the distribution includes more countries in Europe.

8 Examination of the number of tweets per year published

8.1 Introduction

In this chapter the procedure for the visualization of the tweets per year published is described and examined. The visualization of the number of tweets per year is a significant part of the research, because it will depict the annual changes in the variation.

The process followed is depicted at Figure 19.

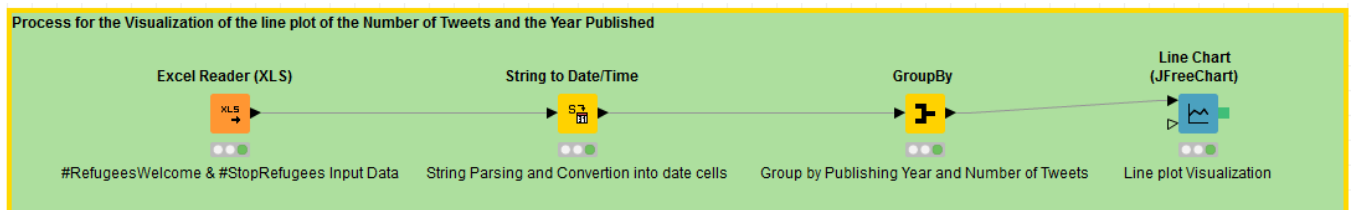
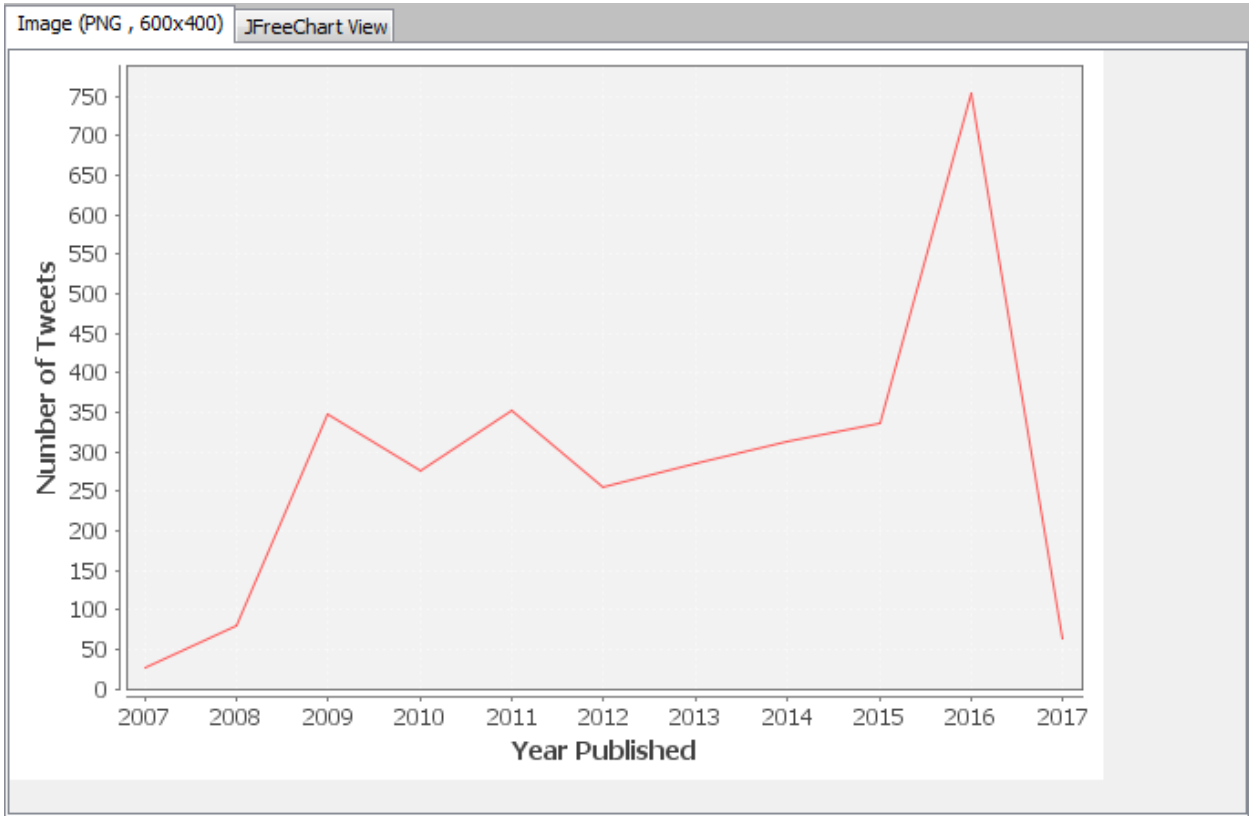


Figure 19: Process for the Visualization of the line plot of the Number of Tweets and the Year Published

The “Excel Reader” Node which reads the file with the input data regarding the two hashtags in question, is connected with the “String to Date/time” Node, which parses the strings in the column with the date and time information by using the entered format pattern, which in this case is the “yyyy”, since the year of the publication date is in concern. Subsequently, the “Group by” Node is configured to run according to the date and number of tweets. The x axis in the Line Chart Node represents the year of the publication, whereas the y axis, the number of tweets.

8.2 #RefugeesWelcome number of tweets per year published visualization

Graph 1 shows the line plot of the #RefugeesWelcome hashtag. The horizontal axis represents the Year the tweets were published and the vertical axis the Number of Tweets.



Graph 1: Refugees Welcome Hashtag line plot

It can be seen that there is a constant flow of tweets during the years 2008 to 2015, which increases dramatically in 2016.

Tweets present an increase in 2009, which declined slightly in 2010, to recover again in 2011. In 2012 the number of tweets falls slightly again and presents a steady increase until 2015, when they rise sharply to reach the highest number in 2016.

8.3 #StopRefugees number of tweets per year published visualization

Graph 2 shows the line plot of the Number of Tweets according to the Date Published.



Graph 2: Stop Refugees Hahstag Line Plot

The plot presents an increase in the number of tweets during 2008, reaching high levels in 2009. Tweets decline in 2010, to return in the 2008 numbers in 2011. They decline slightly in 2013, to recover again in 2014. They decline significantly in 2015, to increase during 2015 and reach the highest numbers in 2016.

9 Examination of the popular tweets

9.1 Introduction

Popular tweets are the ones most favorited and retweeted by Twitter users. By liking and reproducing the tweets, users express their support to the user. The “Twitter Search” Node’s dialog offers the chance of retrieving the most popular tweets regarding the query,

It will be interesting to examine the most popular tweets the twitter search returns, regarding the two hashtags in question.

9.2 #RefugeesWelcome popular tweets

To retrieve the popular hastags related to the #RefugeesWelcome hashtag, a new Search has to be executed through the “Twitter Search Dialog”.

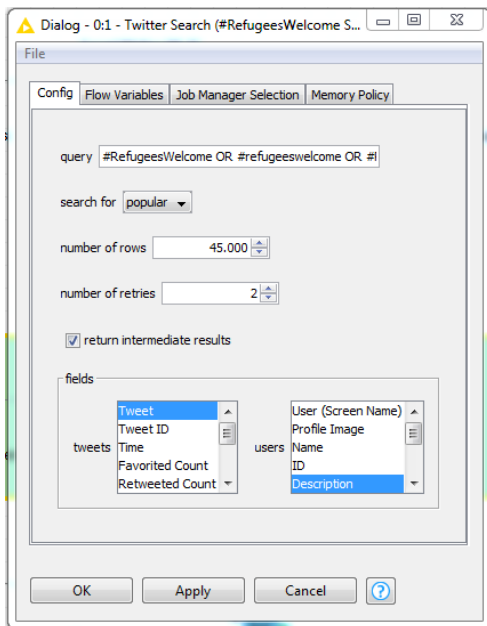


Figure 20: Popular Hashtag Search Dialog

The search includes the same tweets' and users' fields as the #RefugeesWelcome and #StopRefugees hashtag search.

The number of rows requested is 45,000 and the output table returns 10 rows (tweets), of which all are first-time tweets.

Table 7 shows the most frequently terms by descending order.

Table 7: "Refugees Welcome" Hashtag's most frequently used terms count

A/A	Term
1	Welcome
2	Bridgesnotwalls
3	Humanitarian
4	Assistance
5	Standing

The Tag Cloud of the #RefugeesWelcome hashtag's popular tweets search can be seen at Figure 21.



Figure 21: Tag Cloud of the Refugees Welcome Hashtags' Popular Terms

A pink color has been attributed to the minimum value ranking at 0.14 and a black color to the maximum value ranking at 0.33. Figure 21 depicts the gradients of the colors depending on the term frequency.

Figure 22 shows the locations where the most popular tweets are published from.

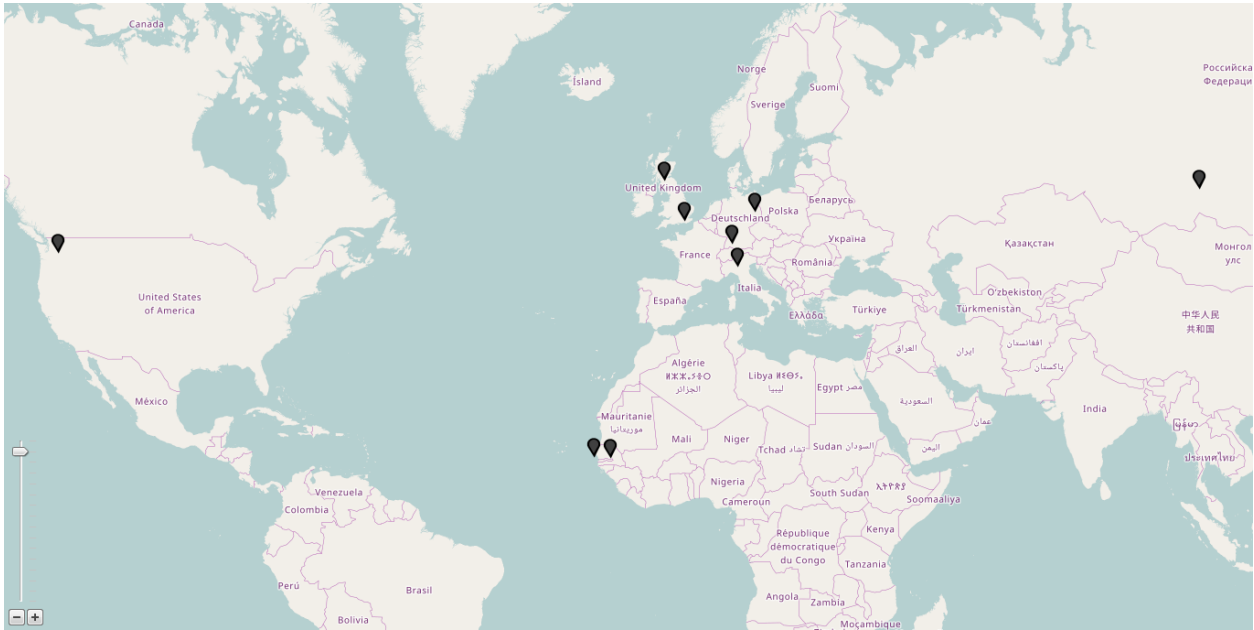
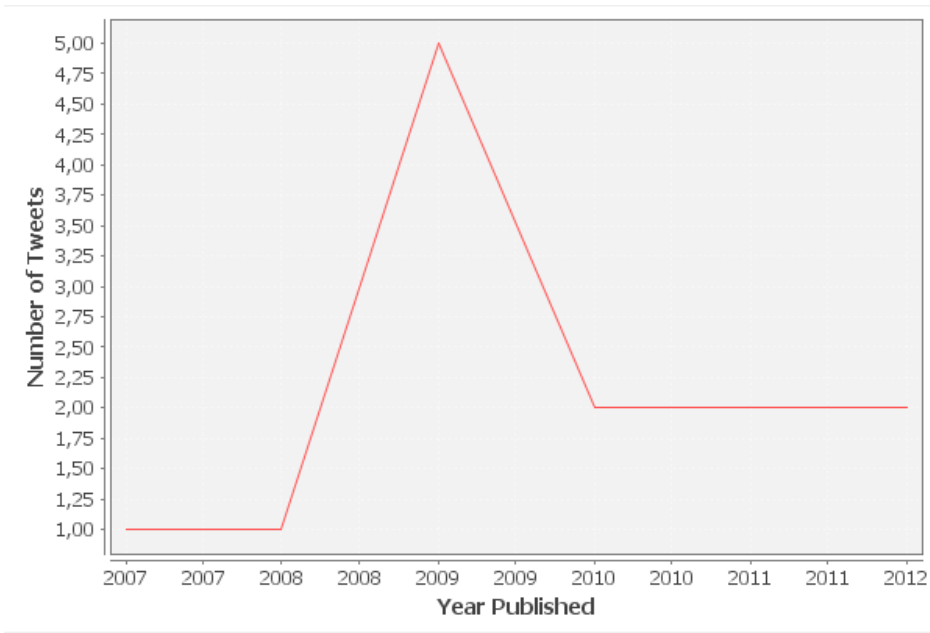


Figure 22: Location of the Refugees Welcome Hashtag Popular Tweets

The popular tweets were mostly published from users in Northern Europe, and specifically United Kingdom and Germany, as seen in Figure 22.

Graph 3 depicts the line plot of the number of tweets per year published.



Graph 3: Number of Refugees Welcome Hashtag Popular Tweets per Year

The tweets present a gradual vertical increase during 2008, to reach the peak in 2009, whereas a gradual reduction occurs again until the year 2010, when they remain stable until 2012.

9.3 #Stop Refugees popular tweets

The #StopRefugees hashtag’s popular search returns only four rows (tweets).

Terms more frequently used are shown in Table 8, in a descending order by the term frequency.

Table 8: “Refugees Welcome” Hashtag's most frequently used terms count

A/A	Term
1	Refugees
2	Chobani
3	Employing
4	Failing
5	European
6	Freezing
7	Unacceptable

Figure 22 depicts the hashtag’s tag cloud.



Figure 23: Stop Refugees Hashtag Popular Terms

The minimum value ranking at 0.2 has been attributed a pink color, whereas a black color has been attributed to the maximum value ranking at 0.333.

Figure 24 shows the map of the tweets locations, all of which are published in London, United Kingdom.

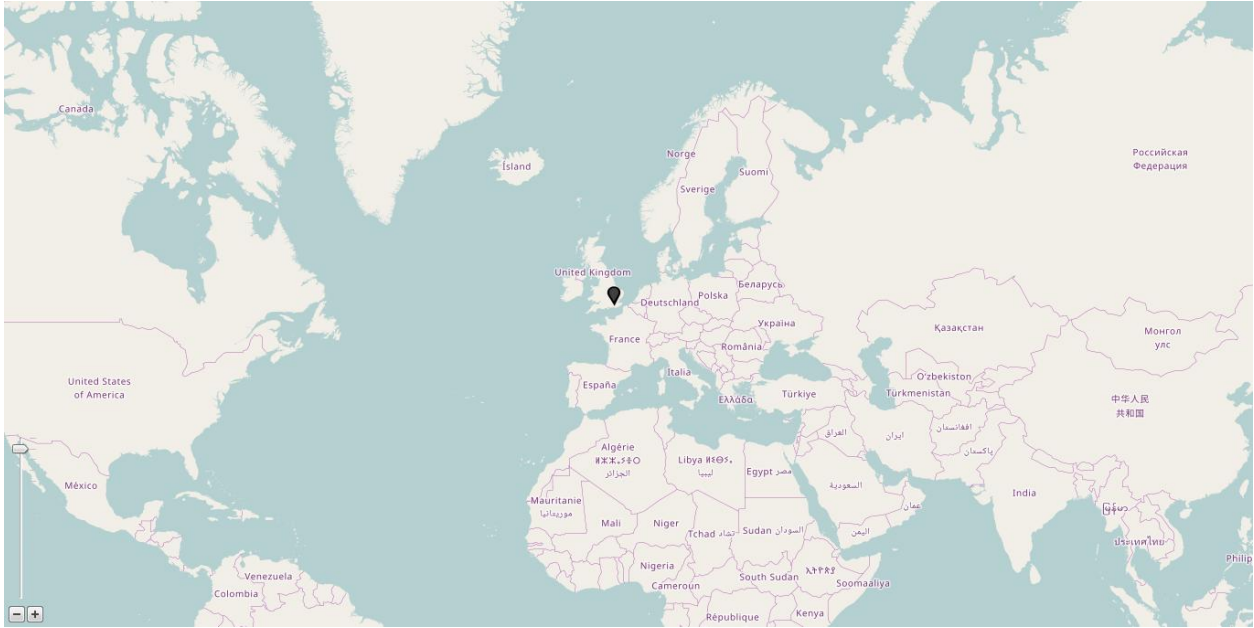


Figure 24: Stop Refugees Popular Tweets' Location

All tweets were as well published in 2008.

9.4 Conclusions

The research of the popular tweets related to the two hashtags in question, present some differences comparing to the overall tweets. The vast majority of the tweets are located in Europe whereas the ones from the United States are significantly less. With regards to the tweets related to the #StopRefugees hashtag, all of them are published in Europe.

There is also a difference in the years the popular tweets were published. Tweets related to the #WelcomeRefugees hashtag mount in 2009, whereas the ones related to the #StopRefugees hashtag are all published in 2008.

The conclusions related to the significant terms seem to be the same with the ones of the overall tweets. It is not easy to distinguish whether the terms used in the tweets for each hashtag are positively or negatively adjacent to the refugees.

10 #RefugeesWelcome and #StopRefugees centralized comparison

As mentioned above, there are some differences between the output data of the search relating to the #RefugeesWelcome and #StopRefugees.

The search about the #RefugeesWelcome hashtag returned 10,330 rows (tweets) of which 3083 are first-time tweets and 10027 are retweets., whereas the search about the #StopRefugees Hashtag returned 4063 rows (tweets) of which 1494 are first-time tweets and 2569 are retweets.

From the above, it can be pointed out that the more users tweet about #RefugeesWelcome than #StopRefugees.

The ten terms most frequently used in the tweets regarding the #RefugeesWelcome hashtag, excluding the hashtag itself, are “welcome”. “bluehand”, “Britain first”, “refugees”, “crisis”, “support”, “women’s march”, immigrants”, “no borders”, “shame on Europe” and “children”.

The ten terms most frequently used in the tweets regarding the #StopRefugees hashtag are “children”, “weather”, “amnesty”, illegal”, “migrants”, America”, “Muslims”, funding”, freezing” and “country”.

The majority of the tweets of both fields are located in Northern Europe and the United States of America. There is a wider distribution in more European countries with regards to the #StopRefugees hashtag.

Tweets regarding the #RefugeesWelcome hashtag present a stable flow during the years 2008 to 2015 and they show an important increase in 2016, whereas tweets of the #StopRefugees hashtag show a stable flow during all years from 2008 to 2016.

The popular tweets of both hashtags are mostly located in Europe. The majority of the tweets related to the #WelcomeRefugees Hashtag are published in 2009, whereas the ones related to the #StopRefugees hashtag are all published in 2008.

11 Conclusions

The dissertation examined two different hashtags regarding the refugee issue, #RefugeesWelcome and #StopRefugees. The first hypothesis was that users pro refugees would tweet using the #RefugeesWelcome hashtag, whereas the ones contra refugees would tweet under #StopRefugees.

The research pointed out that the majority of users tweeting about the refugee issue use the #RefugeesWelcome hashtag. It is important to point out that the Ktime Analytics Platform does not return all tweets related to each hashtag.

Users from Northern Europe and the United States seem to tweet more about the refugee issue, as shown by the analysis. The fact that Europe faces a number of important challenges with regards to the Syrians arriving at its territory is probably the main reason for this. Another fact is that users from Northern Europe countries seem to push pressure to their governments to act pro or contra the refugee issue.

The big amount of tweets in the United States could be attributed to the US elections and Donald Trump's political campaign, a big part of which relied on the "No to Refugees" slogan. The fact that "Women's March", a movement that was initiated by women against Trump's rhetoric at January 21st, is also an important term in the tweets, supports this hypothesis.

A weakness of the research is that Twitter offers users the chance to choose a location of their own choice that may not be their real one thus changing the results of the research.

The flow of tweets regarding the "RefugeesWelcome" hashtag presents an increase during 2016. This year a massive influx of refugees reached the European coasts and the issue of the management of this number of refugees emerged.

The same flow of tweets seems less turbulent with regards to the "Stop Refugees" hashtag, with a small increase during 2016 as well.

The majority of the popular tweets of the hashtags are published between 2008 and 2009, and they locate mostly in Europe.

The terms prevailing in both hashtags cannot provide a clear view on whether the hashtags "Refugees Welcome" or "Stop Refugees" are associated with positive or negative sentiments

respectively. There are both positive and negative terms in the tweets associated with both hashtags. A further sentiment analysis of the tweets would be a helpful hand in figuring out whether the hashtags from the tweets have a positive or negative sentiment.

12 References

- Berthold, M. R. (n.d.). *KNIME: The Konstanz Information Miner*. Konstanz: Konstanz University, Department of Computer and Information Science.
- blog, K. (2016, 12 28). *KNIME* . Retrieved from KNIME Web site: <http://www.knime.org/blog/knime-twitter-nodes>
- Bureau of Population, Refugees, and Migration. (2016). *Proposed Refugee Admissions for Fiscal Year 2017*. United States Department of State, United States Department of Homeland Security, United States Department of Health and Human Services.
- Castillo, C., Mendoza, M., & Poblete, B. (2011). Information Credibility on Twitter. *WWW 2011-Session: Information Credibility* (pp. 675-684). Hyderabad, India: International World Wide Web Conference Committee.
- CFR, C. o. (2016, 12 28). *Council on Foreign Relations*. Retrieved from Council on Foreign Relations Web site: <http://www.cfr.org/syria/s-syrias-civil-war/p38607>
- Commission, E. (2011). Directive 2011/95/EU of the European Parliament and of the Council. *Official Journal of the European Union*.
- Commission, E. (2016, 12 23). *European Solidarity: A Refugee Relocation System*. Brussels: European Commission. Retrieved from European Commission Web site.
- Commission, E. (2016, 7 13). *Press Release Database*. Retrieved from European Commission Web site: http://europa.eu/rapid/press-release_IP-16-2435_en.htm
- Commission, E. (2016). *Syria Crisis*. Brussels: Humanitarian Aid and Civil Protection.
- KNIME. (2016, 12 28). *KNIME Organization*. Retrieved from Knime Web site: <http://www.knime.org/blog/knime-twitter-nodes>
- Marwick, A. E., & Boyd, D. (2016). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *new media & society*, 114-133.

- Pak, A., & Paroubek, P. (n.d.). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. Paris: Universite de Paris-Sud.
- Park, C. S. (2013). Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. *Computers in Human Behavior*, 1641-1648.
- Porter, J. (2009). Relationship Symmetry in Social Networks: Why Facebook will go Fully Asymmetric. *bokardo*.
- syrianrefugees*. (2016, 09). Retrieved from syrianrefugees: <http://syrianrefugees.eu/>
- Thiel, K., & Berthold, M. (2012). *The KNIME Text Processing Feature: An Introduction*. KNIME.
- Twitter. (2016, 12 21). Retrieved from Twitter Web site: <https://about.twitter.com/company>
- Twitter. (2016, 12 21). Retrieved from Twitter Help Center: <https://support.twitter.com/articles/49309%20>
- Twitter. (2016, 21 12). *Twitter Help Center*.
- Twitter. (2016, 12 21). *Twitter Support*. Retrieved from Twitter Support: <https://support.twitter.com/articles/15367>
- Twitter. (2017, 1 6). Retrieved from Twitter Help Center: <https://support.twitter.com/articles/122236>
- UNHCR. (2016, 12 23). *United Nations*. Retrieved from UNHCR: <http://data.unhcr.org/syrianrefugees/asylum.php>