

# **Tendências no Twitter**

**Tiago Manuel Rodrigues Martins**

**Dissertação para obtenção do Grau de Mestre em  
Engenharia Informática, Área de Especialização em  
Tecnologias do Conhecimento e Decisão**

**Orientadora: Dr.ª Isabel Azevedo**

**Júri:**

Presidente:

Doutora Maria de Fátima Coutinho Rodrigues, ISEP

Vogais:

Doutor Paulo Jorge Machado Oliveira, ISEP

Doutora Isabel de Fátima Silva Azevedo, ISEP

Porto, Outubro de 2014



# Dedicatória

Dedico este trabalho aos meus pais, José Manuel e Maria Armandina, os verdadeiros pilares da minha vida, por me terem proporcionado a melhor educação e por todos os seus esforços, fazendo de mim a pessoa que sou hoje.



# Resumo

Atualmente as redes sociais fazem parte do dia a dia das pessoas, sendo que para muitas delas essas plataformas tornaram-se imprescindíveis. Quer seja por mera curiosidade ou por uma necessidade imperiosa de partilhar com o mundo as suas ideias ou pensamentos, a verdade é que as pessoas utilizam cada vez mais as redes sociais.

O trabalho descrito neste documento pretende explorar uma dessas redes sociais, o Twitter, mais propriamente através do estudo das suas tendências do momento ou *trending topics* que são expressões referidas pelos utilizadores a uma taxa superior em comparação com outras. São descritas algumas das características do Twitter e dos recursos disponíveis para programadores.

Para além do estudo detalhado do Twitter, são extraídos dados de diversas fontes, destacando-se o próprio Twitter, mas também outras plataformas e jornais *online*. Posteriormente é feita a análise desses dados através da verificação da evolução das tendências ao longo do tempo nos diversos locais em estudo e ainda através do cruzamento de dados das várias fontes de dados, com análise dos resultados obtidos. Na análise efetuada são analisadas as tendências considerando algumas métricas, a sua globalidade, a relação das tendências com tópicos dos jornais *online* e ainda a sua periodicidade.

**Palavras-chave:** Redes Sociais, Twitter, Tendências, Extração e Análise de Dados



# Abstract

Currently social networks are a part of everyday people's life, and for many of them these platforms became an essential part of their lives. By curiosity or willingness to share their ideas or thoughts, people increasingly use social networks.

The work described in this document intends to explore one of these social networks, Twitter, more specifically by studying their trending topics which are expressions that are mentioned by users at a higher rate compared with others. Some of the characteristics of Twitter are described and some of the resources available for developers.

Besides the detailed study of Twitter, data is extracted from different sources, especially data from Twitter itself, as well as data from other platforms and online newspapers. Later on is performed a data analysis by checking the evolution of trends over time in various locations and also through the intersection of information between the various data sources, with analysis of the results. In the performed analysis some metrics are considered, the relation between trends and online newspapers and also their periodicity.

**Keywords:** Social Networks, Twitter, Trends, Data Extraction and Analysis





# Agradecimentos

À professora Isabel Azevedo, pelo seu incansável apoio e por me ter orientado durante este trabalho de mestrado.

Aos meus pais, por estarem sempre a meu lado e por me terem incentivado a voltar a estudar após uma longa paragem de cinco anos depois de terminada a licenciatura.

À Elsa, por me ter acompanhado e apoiado, tanto nos bons como nos maus momentos, deste projeto.



# Índice

Lista de Figuras .....	xiii
Lista de Tabelas .....	xv
Lista de Códigos .....	xvii
<b>1 Introdução .....</b>	<b>1</b>
1.1 Motivação .....	1
1.2 Objetivos .....	2
1.3 Metodologia .....	2
1.4 Estrutura do Documento .....	3
<b>2 Estado da Arte .....</b>	<b>5</b>
2.1 Twitter .....	5
2.1.1 Acontecimentos no Twitter .....	9
2.1.2 Noções Sobre o Funcionamento .....	14
2.1.3 Hashtag .....	15
2.1.4 Tipos de Contas no Twitter .....	16
2.2 Twitter API .....	18
2.2.1 Versão 1.1 .....	19
2.2.2 Limites .....	21
2.2.3 Ligação à Twitter API .....	22
2.2.4 Mecanismo de Autenticação .....	22
2.3 Tendências do Momento .....	24
2.3.1 Extração das Dez Tendências de um Local .....	26
2.3.2 Tendências Comuns em Diferentes Localizações .....	28
2.3.3 Verificar Termos em Páginas Web .....	29
2.4 Identificador WOEID .....	30
2.4.1 Estrutura .....	30
2.4.2 Tipos de Locais .....	31
2.4.3 Língua .....	32
2.4.4 Consistência Posicional .....	33
<b>3 Extração de Dados .....</b>	<b>35</b>
3.1 Tecnologias Utilizadas .....	35
3.1.1 Python .....	36
3.1.2 SQL .....	36
3.1.3 Semantic Web e SPARQL .....	37
3.2 Extração das Tendências do Momento .....	38
3.2.1 Tendências em Números .....	38
3.2.2 Tarefas Criadas no Sistema Operativo .....	39
3.2.3 Extratores das Tendências .....	40

3.2.4	Sistema Implementado .....	41
3.3	Outras Fontes de Dados Consideradas .....	42
3.3.1	Jornais Diários .....	42
3.3.2	DBpedia .....	42
3.3.3	Extração de Dados dos Jornais Diários.....	43
3.3.4	Extração de Dados da DBpedia.....	45
3.3.5	Consultas SQL com Dados Oriundos da DBpedia .....	49
<b>4</b>	<b>Análise de Dados.....</b>	<b>53</b>
4.1	Análise das Tendências .....	53
4.1.1	Tendências Diferentes.....	54
4.1.2	Repetição de Tendências .....	55
4.1.3	Número Médio de Caracteres .....	55
4.1.4	Número Médio de Palavras.....	56
4.1.5	Número de Palavras por Tópico.....	57
4.1.6	Análise de Hashtags.....	57
4.1.7	Periodicidade de Tópicos .....	60
4.2	Análise das Tendências - Outras Fontes de Dados .....	62
<b>5</b>	<b>Conclusões e Trabalho Futuro .....</b>	<b>75</b>
5.1	Análise de Resultados.....	75
5.2	Limitações .....	76
5.3	Possíveis Melhorias .....	77
	<b>Referências.....</b>	<b>79</b>
	<b>Anexo A .....</b>	<b>83</b>
	Tendências Repetidas .....	83
	<b>Anexo B .....</b>	<b>87</b>
	Hashtags Repetidas .....	87

# Lista de Figuras

Figura 1 – Distribuição geográfica dos utilizadores do Twitter em 2012.....	7
Figura 2 – Distribuição etária dos utilizadores do Twitter em 2012 .....	8
Figura 3 – O <i>tweet</i> com mais <i>retweets</i> da história no Twitter ( <i>@TheEllenShow</i> ).....	10
Figura 4 – Alcance do <i>tweet</i> com mais <i>retweets</i> da história no Twitter.....	10
Figura 5 – Jogadores mais mencionados na fase de grupos do mundial.....	12
Figura 6 – Evolução dos <i>tweets</i> durante a marcação de penaltis.....	12
Figura 7 – Jogadores mais mencionados no mundial .....	13
Figura 8 – O mundial à escala global.....	14
Figura 9 – Botão “Seguir” o antes e o depois.....	14
Figura 10 – Exemplo de um <i>tweet</i> que contém o símbolo “#” .....	15
Figura 11 – Alguns dos <i>tweets</i> com a <i>hashtag</i> #MSW_RADICAL .....	15
Figura 12 – Contas oficiais e verificadas do Twitter (Personalidades).....	16
Figura 13 – Contas oficiais e verificadas do Twitter (Desportistas) .....	16
Figura 14 – Contas oficiais e verificadas do Twitter (Músicos).....	17
Figura 15 – Contas oficiais e verificadas do Twitter (Eventos).....	17
Figura 16 – Contas oficiais e verificadas do Twitter (Notícias) .....	17
Figura 17 – Twitter REST API.....	19
Figura 18 – Twitter <i>Streaming</i> API .....	20
Figura 19 – Funcionamento do protocolo <i>OAuth</i> para aceder à Twitter API .....	23
Figura 20 – Elipsóide WGS 84.....	33
Figura 21 – As camadas da <i>Semantic Web</i> .....	37
Figura 22 – Sistema de extração das tendências .....	41
Figura 23 – Sistema de extração de dados com <i>queries</i> SPARQL .....	43
Figura 24 – Total de tendências extraídas .....	54
Figura 25 – Comparação de totais extraídos e totais de tendências diferentes por local.....	54
Figura 26 – Número médio de caracteres por tópico .....	56
Figura 27 – Número médio de palavras por tópico .....	56
Figura 28 – Número de palavras por tópico.....	57
Figura 29 – Comparação de totais extraídos e totais de <i>hashtags</i> .....	58
Figura 30 – Percentagem de <i>hashtags</i> face ao total de tendências .....	58
Figura 31 – Número médio de caracteres por <i>hashtag</i> .....	59
Figura 32 – Tendências do festival Rock in Rio .....	61
Figura 33 – Tendências de artistas presentes no festival Rock in Rio.....	62
Figura 34 – Tendências nos jornais em 03/07/2014.....	63
Figura 35 – Tendências nos jornais em 04/07/2014.....	64
Figura 36 – Tendências nos jornais em 05/07/2014.....	64
Figura 37 – Tendências nos jornais em 06/07/2014.....	65
Figura 38 – Tendências nos jornais em 07/07/2014.....	65
Figura 39 – Tendências nos jornais em 08/07/2014.....	66
Figura 40 – Tendências nos jornais em 09/07/2014.....	66

Figura 41 – Tendências nos jornais em 10/07/2014 .....	67
Figura 42 – Tendências nos jornais em 11/07/2014 .....	67
Figura 43 – Tendências nos jornais em 12/07/2014 .....	68
Figura 44 – Tendências nos jornais em 13/07/2014 .....	69
Figura 45 – Tendências nos jornais em 14/07/2014 .....	69
Figura 46 – Tendências nos jornais em 15/07/2014 .....	70
Figura 47 – Tendências nos jornais em 16/07/2014 .....	70
Figura 48 – Tendências nos jornais em 17/07/2014 .....	71
Figura 49 – Tendências nos jornais em 18/07/2014 .....	71
Figura 50 – Tendências nos jornais em 19/07/2014 .....	72
Figura 51 – Tendências nos jornais em 20/07/2014 .....	72
Figura 52 – Tendências nos jornais em 21/07/2014 .....	73

# Lista de Tabelas

Tabela 1 – Popularidade do Twitter em vários locais .....	6
Tabela 2 – Línguas faladas no Twitter em 2013.....	7
Tabela 3 – Jogos da fase de grupos com mais <i>tweets</i> .....	11
Tabela 4 – Jogos do mundial com mais <i>tweets</i> .....	13
Tabela 5 – Limites definidos pelo Twitter .....	21
Tabela 6 – Códigos WOEID usados para a obtenção das tendências .....	25
Tabela 7 – Tipos de locais.....	31
Tabela 8 – Códigos de continentes e oceanos.....	32
Tabela 9 – Línguas suportadas pelo <i>Yahoo! Geoplanet</i> .....	32
Tabela 10 – <i>Packages</i> instalados no <i>IPython</i> .....	36
Tabela 11 – Tabelas criadas posteriormente na base de dados .....	37
Tabela 12 – Número de tendências extraídas na 1ª fase (por dia).....	39
Tabela 13 – Número de tendências extraídas na 2ª fase (por dia).....	39
Tabela 14 – Jornais extraídos e respetivos países de origem .....	42
Tabela 15 – Classes que suportaram a extração de dados da <i>DBpedia</i> .....	46
Tabela 16 – Classes usadas para extrair dados sobre desportistas .....	47
Tabela 17 – Recursos usados para extrair dados sobre festivais portugueses da <i>DBpedia</i> .....	47
Tabela 18 – Combinação de dados das tabelas <i>trends</i> e <i>countries</i> .....	50
Tabela 19 – Combinação de dados das tabelas <i>trends</i> e <i>european_capital_cities</i> .....	50
Tabela 20 – Combinação de dados das tabelas <i>trends</i> e <i>pt_cities</i> .....	50
Tabela 21 – Combinação de dados das tabelas <i>trends</i> e <i>players</i> .....	51
Tabela 22 – Combinação de dados das tabelas <i>trends</i> e <i>festivals</i> .....	51
Tabela 23 – Número de tendências repetidas (cinco ou mais repetições).....	55
Tabela 24 – Número de <i>hashtags</i> repetidas (três ou mais repetições).....	59
Tabela 25 – Tendências repetidas (quinze ou mais repetições) .....	60
Tabela 26 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (03/07/2014).....	63
Tabela 27 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (04/07/2014).....	63
Tabela 28 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (05/07/2014).....	64
Tabela 29 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (06/07/2014).....	65
Tabela 30 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (07/07/2014).....	65
Tabela 31 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (08/07/2014).....	66
Tabela 32 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (09/07/2014).....	66
Tabela 33 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (10/07/2014).....	67
Tabela 34 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (11/07/2014).....	67
Tabela 35 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (12/07/2014).....	68
Tabela 36 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (13/07/2014).....	68
Tabela 37 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (14/07/2014).....	69
Tabela 38 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (15/07/2014).....	70
Tabela 39 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (16/07/2014).....	70
Tabela 40 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (17/07/2014).....	71

Tabela 41 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (18/07/2014) .....	71
Tabela 42 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (19/07/2014) .....	72
Tabela 43 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (20/07/2014) .....	72
Tabela 44 – Comparação de dados do Twitter, <i>DBpedia</i> e jornais <i>online</i> (21/07/2014) .....	73



# Lista de Códigos

Código 1 – Autorização de uma aplicação para aceder aos dados da conta do Twitter .....	22
Código 2 – Extração das tendências de Portugal (WOEID=23424925) .....	26
Código 3 – Dez tendências de Portugal em 13/03/14 – 20h37m .....	28
Código 4 – Intersecção das tendências em Espanha e no Mundo em 04/05/14 – 10h55m .....	28
Código 5 – Verificação de termos em páginas <i>web</i> .....	29
Código 6 – <i>Script</i> executado pela tarefa “Trends_10H” .....	40
Código 7 – Extração das tendências em Portugal .....	40
Código 8 – <i>Script</i> de inserção das tendências na base de dados .....	41
Código 9 – <i>Script</i> executado pela tarefa “jornais” .....	44
Código 10 – Extração diária do <i>Jornal de Notícias</i> (Portugal) .....	44
Código 11 – Extração diária do <i>El Mundo</i> (Espanha) .....	44
Código 12 – Extração diária do <i>The New York Times</i> (Estados Unidos) .....	44
Código 13 – Extração diária do <i>Folha de São Paulo</i> (Brasil) .....	45
Código 14 – Extração diária do <i>La Nación</i> (Argentina) .....	45
Código 15 – Extração diária do <i>The Guardian</i> (Reino Unido) .....	45
Código 16 – Extração da lista de países membros das Nações Unidas da <i>DBpedia</i> .....	46
Código 17 – Extração da lista de capitais europeias da <i>DBpedia</i> .....	46
Código 18 – Extração de dados da <i>DBpedia</i> (jogadores de futebol) .....	47
Código 19 – <i>Script</i> de inserção dos países na base de dados .....	48
Código 20 – <i>Script</i> de inserção das capitais europeias na base de dados .....	48
Código 21 – <i>Script</i> de inserção das cidades portuguesas na base de dados .....	48
Código 22 – <i>Script</i> de inserção de desportistas de várias modalidades na base de dados .....	48
Código 23 – <i>Script</i> de inserção de eventos de música na base de dados .....	49
Código 24 – Consulta SQL com a combinação de tendências com países .....	49
Código 25 – Consulta SQL com a combinação de tendências com cidades capitais europeias .....	50
Código 26 – Consulta SQL com a combinação de tendências com cidades portuguesas .....	50
Código 27 – Consulta SQL com a combinação de tendências com desportistas .....	50
Código 28 – Consulta SQL com a combinação de tendências com festivais de música .....	51



# Acrónimos e Símbolos

## Lista de Acrónimos

<b>API</b>	<i>Application Program Interface</i>
<b>GPS</b>	<i>Global Positioning System</i>
<b>IERS</b>	<i>International Earth Rotation and Reference Systems Service</i>
<b>IRM</b>	<i>IERS Reference Meridian</i>
<b>IRP</b>	<i>IERS Reference Pole</i>
<b>JSON</b>	<i>JavaScript Object Notation</i>
<b>LOD</b>	<i>Linking Open Data</i>
<b>OAuth</b>	<i>Open Authentication</i>
<b>RDF</b>	<i>Resource Description Framework</i>
<b>REST</b>	<i>Representational State Transfer</i>
<b>RIR</b>	<i>Rock in Rio</i>
<b>SMS</b>	<i>Short Message Service</i>
<b>SPARQL</b>	<i>SPARQL Protocol and RDF Query Language</i>
<b>SQL</b>	<i>Structured Query Language</i>
<b>TPM</b>	<i>Tweets per Minute</i>
<b>UNOOSA</b>	<i>United Nations Office for Outer Space Affairs</i>
<b>URI</b>	<i>Uniform Resource Identifier</i>
<b>URL</b>	<i>Uniform Resource Locator</i>
<b>UTF-8</b>	<i>8bit Unicode Transformation Format</i>
<b>WGS</b>	<i>World Geodetic System</i>
<b>WOEID</b>	<i>Where On Earth ID</i>



# 1 Introdução

*“O homem é por natureza um ser social.”*

Aristóteles

## 1.1 Motivação

Desde a antiguidade que o homem vive ou anseia viver em sociedade. Na obra literária “Política” [Aristóteles, 350AC] é referido que aquele que não consegue viver em sociedade ou não sente essa necessidade é um animal ou um deus. O homem possui por natureza um instinto social que o leva a relacionar-se. O conceito de rede social remonta assim à antiguidade no que diz respeito ao privilégio do contacto humano. No entanto, nos dias de hoje, com o avanço da tecnologia, com a informatização generalizada e o aparecimento de *sites* de redes sociais, o paradigma das relações entre pessoas foi alterado passando muitas relações a terem um carácter virtual.

A forma como as pessoas comunicam e partilham informação, alterou-se com o aparecimento das redes sociais na primeira década de 2000, atraindo milhões de utilizadores em todo o mundo, tornando-se rapidamente parte do seu dia a dia. As redes sociais assumem assim um lugar de destaque cada vez mais acentuado nas sociedades modernas, estando a tornar-se na escolha generalizada dos utilizadores quando se fala na disseminação e partilha de informação.

Das redes sociais mais utilizadas, o Twitter consegue destacar-se de todas as outras quando se fala em velocidade de propagação de notícias ou divulgação de eventos que estejam a decorrer num determinado momento ou localização.

Assim, com toda esta dinâmica que caracteriza o Twitter, tornou-se óbvia a escolha deste tema relacionado com o estudo desta rede social e das suas tendências do momento (*trending topics*). O facto de se poder observar a evolução das tendências em vários locais,

consoante alguns eventos que iam decorrendo ao longo do tempo, foi também bastante estimulante na altura de escolher o tema e assim poder constatar que a informação que circula no Twitter é muitas das vezes um reflexo do que está a acontecer no mundo.

## **1.2 Objetivos**

De forma sucinta, os objetivos principais deste trabalho foram os seguintes:

- Analisar os tópicos mais mencionados no Twitter em relação a características simples do texto, como o número médio de caracteres, número médio de palavras, utilização de *hashtags* e número de palavras existentes por tópico;
- Verificar se os assuntos mais discutidos em alguns países através do Twitter são comuns aos discutidos noutros ou seja se têm um carácter global ou não;
- Analisar o que as pessoas discutem concretamente, nomeadamente a relação das tendências com tópicos abordados nos noticiários;
- Verificar se os tópicos que são mais discutidos têm geralmente um carácter único ou tendem a repetir-se durante algum tempo ou com alguma periodicidade.

## **1.3 Metodologia**

Os objetivos principais foram concretizados através da execução das seguintes tarefas:

- Estudo do Twitter onde se subentende a análise das suas características, tipos de conta, *hashtags* entre outras;
- Análise de alguns acontecimentos/eventos;
- Estudo das tendências do momento do Twitter;
- Estudo do sistema de referência geográfica WOEID (*Where On Earth ID*);
- Desenvolvimento de uma aplicação para extração de dados da Twitter API;
- Escolha dos países/locais alvo de análise e criação das tarefas automáticas para efetuar a extração desejada;
- Análise dos dados recolhidos.

## 1.4 Estrutura do Documento

Este trabalho encontra-se estruturado em cinco capítulos que são descritos de seguida.

O primeiro capítulo é introdutório e contém quatro tópicos nos quais se abordam as motivações para ter sido escolhido este tema em particular, os objetivos que se pretendem atingir juntamente com as tarefas desenvolvidas para atingir esses mesmos objetivos (metodologia) e por último a estrutura do documento elaborado.

O segundo capítulo começa por abordar o Twitter e a sua importância a nível mundial, sendo também enumeradas algumas das suas características e explicado o seu funcionamento propriamente dito. Neste capítulo inicia-se o estudo do interface de programação do Twitter (Twitter API), das tendências do momento (*trending topics*) e são também abordados alguns conceitos relacionados com o sistema de referência geográfica usado pelo Twitter.

O terceiro capítulo debruça-se sobre o processo de extração de dados. São apresentadas as tecnologias utilizadas neste percurso, as tarefas criadas para a extração das tendências durante um intervalo de tempo previamente definido e são também apresentados os extratores propriamente ditos. São ainda apresentadas e analisadas outras fontes de dados que foram também consideradas em todo este processo.

O quarto capítulo apresenta toda a análise efetuada aos dados extraídos das diversas fontes de dados usadas.

No último capítulo apresentam-se algumas reflexões e as conclusões finais. Foram também incluídos neste capítulo, as limitações do trabalho e os melhoramentos que podem vir a ser efetuados.





## 2 Estado da Arte

Neste capítulo são abordados diversos conceitos relacionados com a rede social Twitter, sendo enumerados alguns acontecimentos que ocorreram em 2014 e que tiveram grande repercussão no Twitter. São também apresentadas algumas noções sobre o funcionamento desta rede social, das *hashtags*, assim como uma explicação sobre os tipos de contas existentes.

Ainda neste capítulo é explorada a Twitter API, onde são descritas algumas das suas características, ao nível da sua arquitetura, limites de utilização e mecanismos de autenticação.

São também apresentados alguns conceitos relacionados com as tendências do momento (*trending topics*), explicando o seu funcionamento e em que consistem.

Por último e não menos importante há uma secção dedicada ao estudo dos identificadores de locais com importância para o trabalho realizado, nomeadamente os WOEID (*Where on Earth ID*).

### 2.1 Twitter

O Twitter, criado em 2006, é uma rede social extremamente popular, otimizada para comunicações instantâneas. O Twitter é assim habitualmente descrito como um serviço de *microblogging* que permite às pessoas comunicar através de mensagens de até 140 caracteres, denominadas de *tweets*, sendo considerado um serviço global e gratuito de mensagens de texto. Essas mensagens, transmitidas em tempo real, correspondem quase sempre a ideias ou pensamentos muitas vezes reflexo de algo que se está a passar no momento, podendo ser um evento local ou mesmo um evento à escala mundial.

A dimensão dos *tweets* (140 caracteres) foi definida para ser compatível em termos de tamanho com as mensagens de texto (SMS), visto inicialmente ter sido direcionado para o mercado das comunicações móveis. Foi já catalogado como uma espécie de “SMS da Internet” [Dijck, 2011].

Para além de texto, o utilizador tem também a possibilidade de inserir imagens, vídeos e até hiperligações para outros *sites*, tudo isto no *tweet* que está a criar. Nestes casos o número de caracteres disponíveis (140) diminui consideravelmente, pois estes elementos (imagens, vídeos) também ocupam espaço no *tweet*.

A escolha do nome Twitter passou por um processo evolutivo no qual foram ponderadas outras designações como, por exemplo, *Friendstalker* e *Twwtr* [Stampler, 2013].

Segundo o serviço Alexa, o Twitter ocupava no dia 27 de Julho de 2014 o 7º lugar nos Estados Unidos da América e o 8º lugar a nível mundial quanto ao nível de popularidade no que diz respeito a *sites* [Alexa.com, 2014]. Esta classificação é obtida combinando a média dos visitantes diários e as visualizações da página nos últimos três meses.

Na Tabela 1 apresentam-se alguns dados sobre as classificações do Twitter nos vários países escolhidos para a extração dos dados e que mais tarde vão ser abordados neste trabalho. As respetivas posições refletem a popularidade do Twitter nestes países.

Tabela 1 – Popularidade do Twitter em vários locais

<i>Pais</i>	<i>Posição</i>
<i>Portugal</i>	17
<i>Espanha</i>	5
<i>Estados Unidos</i>	7
<i>Brasil</i>	12
<i>Argentina</i>	9
<i>Reino Unido</i>	10

Fonte: <https://www.alexa.com> (27 de Julho de 2014)

A Tabela 2 apresenta as percentagens relativas às línguas mais faladas no Twitter reportando-se esta informação a Setembro de 2013. Nota-se claramente um predomínio da língua inglesa, influenciada sobretudo pela difusão do Twitter nos Estados Unidos, assim como uma forte presença do Espanhol motivado não só pela popularidade do Twitter em Espanha mas também em países da América Latina. De referir também a ascensão do Português como uma das línguas mais faladas, justificada essencialmente pela afirmação do Brasil como um dos países com maior número de utilizadores no Twitter.

Tabela 2 – Línguas faladas no Twitter em 2013

<i>Línguas</i>	<i>Percentagem (%)</i>
<i>Inglês</i>	34
<i>Japonês</i>	16
<i>Espanhol</i>	12
<i>Malaio</i>	8
<i>Português</i>	6
<i>Árabe</i>	6
<i>Francês</i>	2
<i>Turco</i>	2
<i>Tailandês</i>	1
<i>Coreano</i>	1
<i>Outras</i>	12

Fonte: <https://twitter.com/mashable/status/413601039053762560/photo/1>

A Figura 1 mostra a distribuição dos utilizadores do Twitter a nível mundial. Os Estados Unidos, o Brasil, o Reino Unido, a Índia e a Austrália surgem como os países com o maior número de utilizadores.

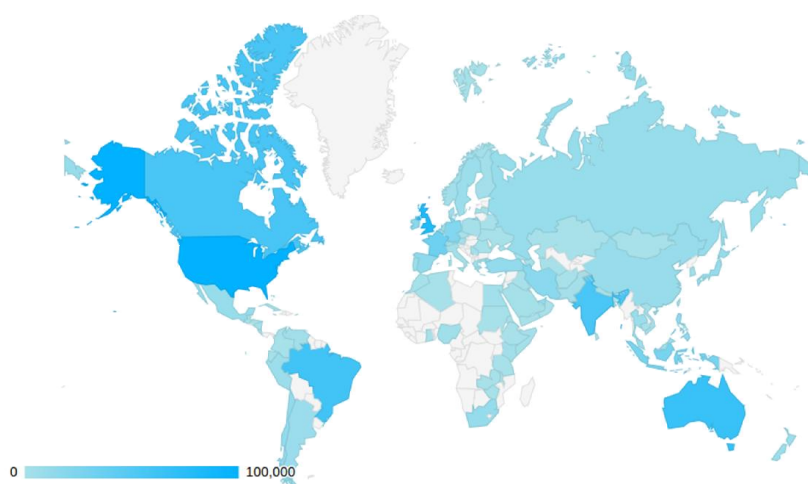


Figura 1 – Distribuição geográfica dos utilizadores do Twitter em 2012

Fonte: <http://www.beevolve.com/twitter-statistics/>

Estudos recentes apontam para um crescimento acelerado nos próximos anos em particular no continente asiático, onde se prevê que por volta de 2018, mais de 40% dos utilizadores do Twitter estejam na zona asiática [Hockenson, 2014].

A Figura 2 apresenta a distribuição dos utilizadores do Twitter respetivamente por faixa etária. De referir o peso de 73,7% que os utilizadores dos 15-25 anos têm na rede social, sendo que 88,6% têm até 35 anos de idade, apontando claramente os jovens como público-alvo do Twitter.

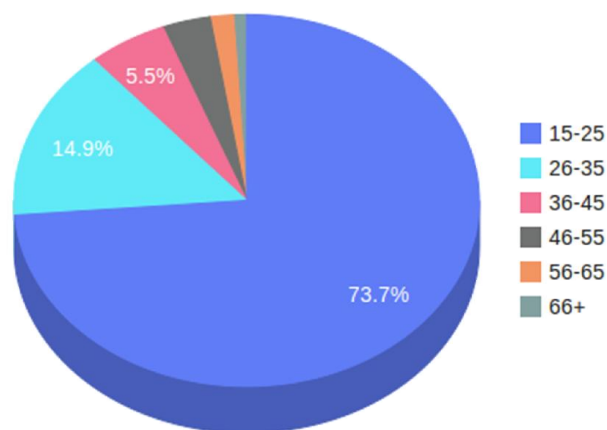


Figura 2 – Distribuição etária dos utilizadores do Twitter em 2012

Fonte: <http://www.beevolve.com/twitter-statistics/>

Foram necessários três anos, dois meses e um dia para se atingir um bilião de *tweets* enviados, sendo que atualmente o mesmo número de *tweets* é atingido em apenas 48 horas. O Twitter tem maior expressão em países tão díspares como os Estados Unidos da América, Índia, Japão, Reino Unido e Espanha.

A elevada velocidade de propagação e a facilidade de publicação tornaram o Twitter num importante meio de comunicação para todos os seus utilizadores. De referir o papel importante do Twitter em alguns eventos recentes a nível mundial, como foi o caso da “Primavera Árabe” em finais de 2010 e princípios de 2011 [Kassim, 2012] [Manfreda, 2013].

Os investigadores da *San Diego State University* efetuaram, entre Junho e Dezembro de 2012, um estudo no qual analisaram através de um algoritmo, 161.821 *tweets* que continham a palavra “*flu*” e 6.174 *tweets* que continham a palavra “*influenza*”. Os *tweets* relacionados com a doença foram usados para identificar um pico no surto de gripe em 9 de cada 11 cidades possíveis [Griffiths, 2013]. Uma análise detalhada nestes casos pode servir para preparar as autoridades responsáveis, neste caso os hospitais, sobre uma situação iminente e assim tentar minimizar todos os aspetos negativos inerentes a uma situação de maior afluência aos hospitais.

Nos Estados Unidos da América, o Twitter é frequentemente usado com o propósito de alertar as populações para eventuais tempestades em formação e por vezes para informar sobre os efeitos devastadores de desastres naturais, como foi o caso do furacão *Sandy* em 2012 [Guskin and Hitlin, 2012].

Dados presentes num estudo recente (Janeiro de 2014), mostram que o Twitter ultrapassou a barreira dos 600 milhões de utilizadores registados [Statisticbrain.com, 2014], dos quais aproximadamente 200 milhões são utilizadores ativos mensalmente. O número de *tweets* ultrapassa em média os 500 milhões por dia.

Outra demonstração do poder do Twitter no que diz respeito à propagação de informação ocorreu em 2013, quando um *tweet* falso sobre explosões na Casa Branca que teriam ferido o presidente Obama, semeou o pânico no mercado de *stocks* S&P 500 (*Standard & Poor's 500 Index*). Em apenas dois minutos o mercado de *stocks* registou uma quebra em que desapareceram 136 bilhões de dólares. A situação foi reposta momentos depois de ter sido confirmado que se tratava de um *tweet* falso [Wang, Kisling and Lam, 2013].

### **2.1.1 Acontecimentos no Twitter**

Alguns acontecimentos adquirem um grande destaque no Twitter. São acontecimentos que pelas suas características e pelo interesse despertado nas pessoas, atingem níveis de notoriedade mais elevados do que os restantes eventos. Seguidamente são apresentados dois dos eventos que ocorreram em 2014 e que tiveram grande repercussão no Twitter. No entanto, estes dois eventos são bastante distintos quanto à sua duração, já que a cerimónia de entrega dos óscares teve apenas um dia de duração e o mundial de futebol no Brasil durou um mês.

#### **2.1.1.1 Óscares 2014**

No dia 2 de Março de 2014, a cerimónia dos óscares foi o tema mais falado a nível mundial na rede social Twitter. O impacto que esta cerimónia teve encontra-se descrito nos seguintes números que se referem ao período de 48 horas entre as 17h de domingo (Hora ET – Costa Leste) e as 17h de terça-feira (Hora ET – Costa Leste) [Fleischman, 2014].

- Foram enviados 19,1 milhões de *tweets* no referido período;
- Esses *tweets* foram enviados por mais de 5 milhões de pessoas;
- Mais de 37 milhões de pessoas viram esses *tweets* em '*twitter.com*', em aplicações para telemóveis e para *desktop*, aproximando-se do número de pessoas que viram a cerimónia (43 milhões);
- Os 19,1 milhões de *tweets* foram visualizados 3,3 bilhões de vezes a nível mundial.

O número de visualizações foi medido usando uma métrica que analisava o número de vezes que os *tweets* apareciam aos utilizadores. Os *tweets* ganharam ainda maior exposição a partir do momento em que começaram a ser incorporados em páginas de notícias passando nesse momento a chegar a um maior número de utilizadores.

Analisando o *tweet* com mais *retweets* na história do Twitter (Figura 3), publicado pela apresentadora dos óscares de 2014, Ellen DeGeneres na sua conta *@TheEllenShow* fica-se com a ideia de como um único *tweet* consegue alcançar milhões de utilizadores.

Ellen DeGeneres @TheEllenShow · 3 de mar  
If only Bradley's arm was longer. Best photo ever. #oscars



Figura 3 – O *tweet* com mais *retweets* da história no Twitter (@TheEllenShow)  
Fonte: <https://twitter.com/TheEllenShow/status/440322224407314432/photo/1>

Os dados sobre esse *tweet* num período de 12 horas foram os seguintes (Figura 4):

- Foi visto em 'twitter.com', em aplicações para telemóveis e para desktop por 8,1 milhões de pessoas num total de 26 milhões de vezes;
- Foi embebido em 13.711 páginas *web*. Esses *tweets* embebidos foram vistos 6,8 milhões de vezes;
- O total de visualizações combinando estes dois pontos ascende a 32,8 milhões de visualizações;
- Foi "retweetado" 2,4 milhões de vezes.

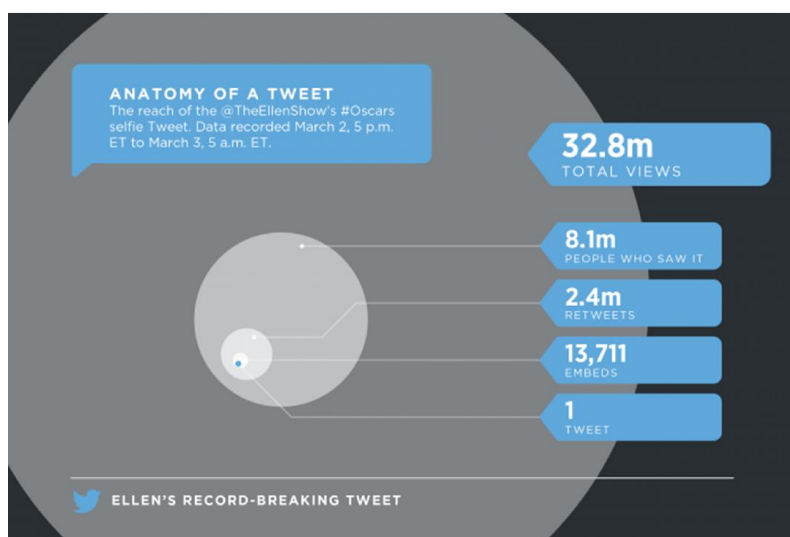


Figura 4 – Alcance do *tweet* com mais *retweets* da história no Twitter  
Fonte: [https://g.twimg.com/blog/blog/image/ellenimpressions\\_0.png](https://g.twimg.com/blog/blog/image/ellenimpressions_0.png)

### 2.1.1.2 Mundial 2014

Um dos grandes eventos desportivos de 2014 teve uma grande repercussão no Twitter. O mundial de futebol no Brasil decorreu entre 12 de Junho e 13 de Julho e durante os 64 jogos no Brasil, as pessoas manifestaram-se nas redes sociais, inclusive no Twitter, à medida que a competição se ia desenrolando.

Apesar do mundial de futebol do Brasil ter sido o evento com o maior número de *tweets* registados pelo Twitter, com cerca de 672 milhões de *tweets*, torna-se difícil comparar um evento de um mês de duração e 64 jogos, com eventos de curta duração como por exemplo, a final do *Super Bowl* americano, a final da liga dos campeões europeus ou ainda com uma cerimónia dos óscares.

Ao longo do mundial, os aficionados do futebol foram discutindo todos os temas e situações que foram surgindo à medida que se desenrolava a competição. Os picos de atividade que se verificaram ocorreram sempre durante os jogos, quer fossem relacionados com uma goleada estrondosa, uma vitória inesperada, um jogo decidido através do recurso à marcação de grandes penalidades ou simplesmente um lance polémico.

Durante a fase de grupos, foram registados cerca de 300 milhões de *tweets* sobre o mundial, sendo nessa altura previsível que este se tornasse o evento mais falado de sempre no Twitter. Para se ter um termo de comparação, os Jogos Olímpicos de 2012 em Londres tiveram 150 milhões de *tweets* [Rogers, 2014a]. Na Tabela 3 encontram-se os jogos da fase de grupos que tiveram mais *tweets*.

Tabela 3 – Jogos da fase de grupos com mais *tweets*

<i>Jogo</i>	<i>Nº Tweets (Milhões)</i>
<i>Brasil vs Croácia *</i>	12,2
<i>Brasil vs México</i>	8,95
<i>Alemanha vs Portugal</i>	8,9
<i>Espanha vs Holanda</i>	8,3
<i>Espanha vs Chile</i>	8,2

\* - *Jogo inicial do torneio*

Outra medida que o Twitter utiliza para efetuar medições são os “*Tweets per Minute*” ou TPM. Durante esta fase inicial os momentos que tiveram mais *tweets*, deram-se essencialmente no jogo de abertura do torneio e estiveram relacionados com um autogolo do jogador Marcelo (378.085 TPM) e com os dois golos dos jogadores Neymar (280.265 TPM) e Óscar (256.953 TPM). Um dos momentos que registou mais *tweets* na fase de grupos (261.026 TPM), registou-se no jogo entre a Alemanha e Portugal, aquando da expulsão do jogador Pepe.

Quanto aos jogadores mais mencionados na fase de grupos, os jogadores Messi e Neymar ambos com quatro golos ocuparam as duas primeiras posições logo seguidos pelo jogador uruguaio Suárez, que protagonizou um dos momentos do mundial ao morder um adversário (Figura 5).

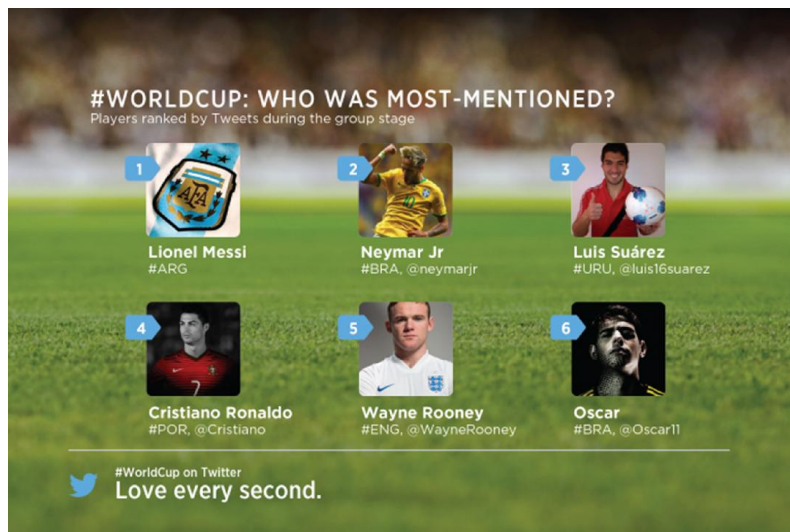


Figura 5 – Jogadores mais mencionados na fase de grupos do mundial

Fonte: [https://g.twimg.com/blog/blog/image/WCMENTIONStotal\\_Most\\_mentions-1.png](https://g.twimg.com/blog/blog/image/WCMENTIONStotal_Most_mentions-1.png)

Terminada a fase de grupos, surgiram vários jogos a serem decididos através do recurso à marcação de grandes penalidades. A Figura 6 demonstra o comportamento verificado pelo Twitter, que tem como característica mais curiosa o facto de se verificar um silêncio ou quebra no número de *tweets* antes do jogador efetuar o remate, disparando esse número de *tweets* depois do remate ser efetuado [Rios, 2014].

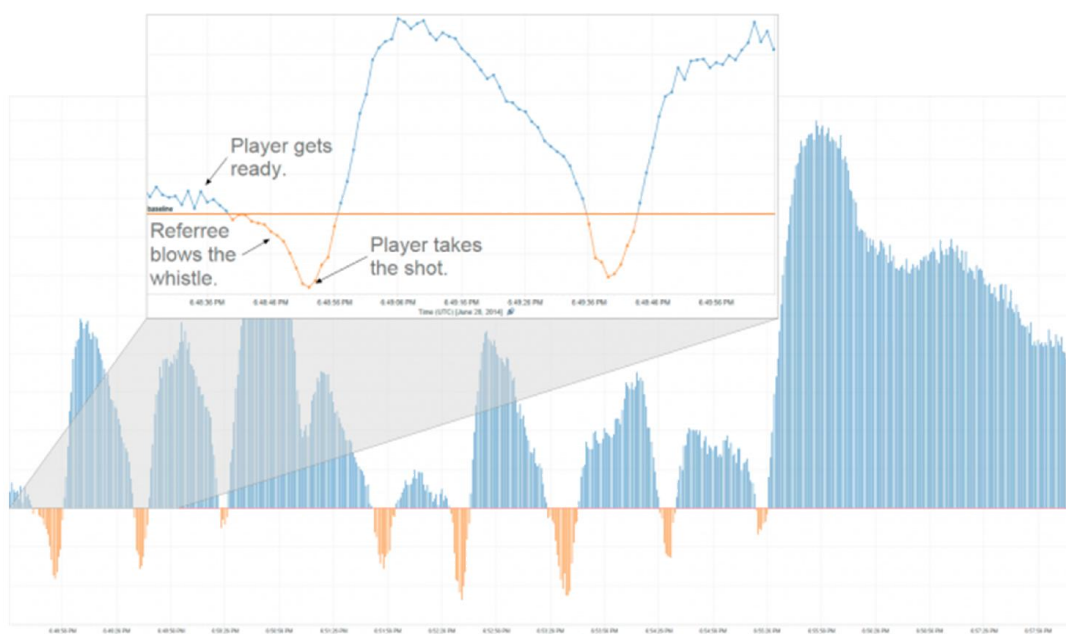


Figura 6 – Evolução dos *tweets* durante a marcação de penaltis

Fonte: [https://g.twimg.com/blog/blog/image/penalty\\_snapshot.png](https://g.twimg.com/blog/blog/image/penalty_snapshot.png)

Terminado o mundial, os números que o Twitter publicou, demonstraram bem a dimensão deste evento, tendo sido batidos todos os recordes para um evento desportivo. Na Tabela 4



encontram-se os jogos que durante a prova tiveram o maior número de *tweets* [Rogers, 2014b].

Tabela 4 – Jogos do mundial com mais *tweets*

Jogo	Nº Tweets (Milhões)
Brasil vs Alemanha *	35,6
Alemanha vs Argentina **	32,1
Brasil vs Chile	16,4
Holanda vs Argentina	14,2
Brasil vs Colômbia	12,4

\* - Recorde de tweets

\*\* - Final do Mundial

No que diz respeito aos cinco momentos que registaram mais *tweets* por minuto durante a competição, três deles decorreram no jogo entre o Brasil e a Alemanha (1-7). O momento que mais se destacou foi precisamente o final do último jogo do mundial que sagrou a Alemanha como campeã mundial de futebol, tendo registado 618.725 TPM. Quanto aos jogadores mais mencionados durante toda a competição, os jogadores Neymar e Messi ocuparam as duas primeiras posições logo seguidos pelo jogador uruguaio Suárez (Figura 7).

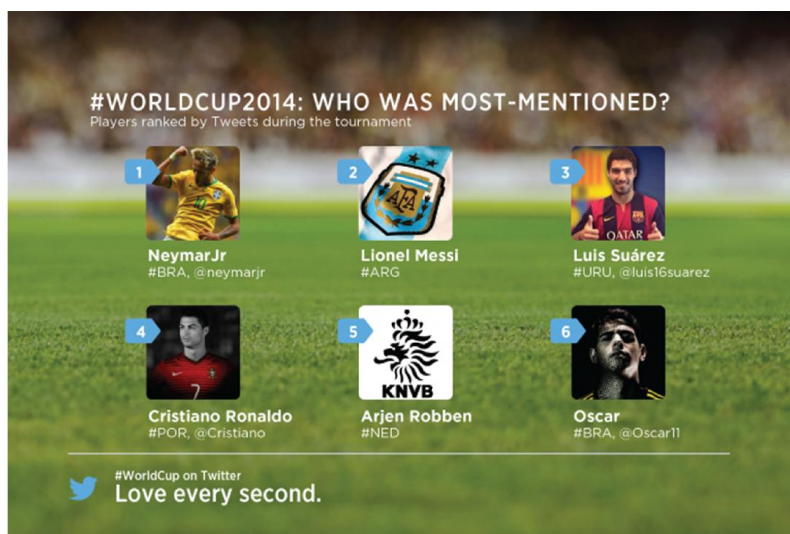


Figura 7 – Jogadores mais mencionados no mundial

Fonte: [https://g.twimg.com/blog/blog/image/wctotalplayers\\_Most\\_mentions.png](https://g.twimg.com/blog/blog/image/wctotalplayers_Most_mentions.png)

O mundial de futebol no Brasil foi sem dúvida um evento à escala global, que ultrapassou fronteiras e alcançou quase todos os países do mundo. A Figura 8 refere-se à final da competição e demonstra bem a magnitude deste evento através da análise de milhões de *tweets* relacionados com o evento. As cores apresentadas na Figura 8 vão desde o vermelho até ao branco que simboliza um maior fluxo no número de *tweets* a circular na rede.

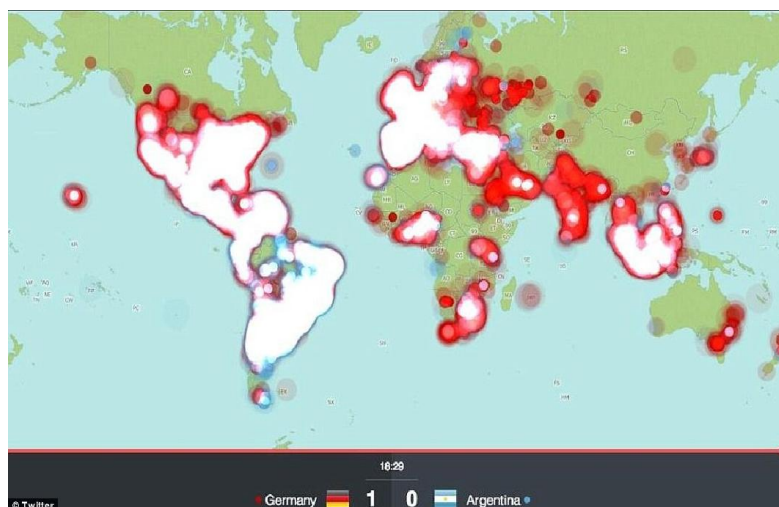


Figura 8 – O mundial à escala global

Fonte: <https://twitter.com/TwitterData/status/488483433849438208/photo/1>

### 2.1.2 Noções Sobre o Funcionamento

Ao contrário de outras redes sociais como o Facebook e o LinkedIn que promovem os conceitos de amizade e de contacto respetivamente, e que obrigam um utilizador a decidir se deve ou não aceitar determinado convite, o Twitter defende um conceito de seguidores ou *followers*. Por outras palavras no Twitter é possível seguir alguém sem que para isso seja necessário que essa pessoa faça o mesmo. Este mecanismo explora um aspeto fundamental da condição humana, a curiosidade. Essa curiosidade pode-se manifestar sobre uma celebridade, sobre política, desporto ou apenas sobre alguém novo em particular.



Figura 9 – Botão “Seguir” o antes e o depois

Quando um utilizador pretende mencionar outro utilizador ou mesmo responder a um *post* basta que seja usado o símbolo (@) seguido do nome do utilizador que se pretende mencionar. Para responder a uma mensagem de outro utilizador do Twitter e partilhar com os próprios seguidores pode ser usada a função de *retweet* bastando para isso ser identificada com o símbolo “RT”.

Os utilizadores podem agrupar publicações ou *posts* por tópicos através do uso das *hashtags* (#) que são utilizadas para marcar palavras-chave ou tópicos num determinado *tweet*.

### 2.1.3 Hashtag

As *hashtags* foram criadas fundamentalmente com o intuito de categorizar mensagens [Support.twitter.com, 2014a]. Essa categorização surge bem patente quando num *tweet*, um utilizador clica na palavra que se encontra precedida do símbolo #, surgindo de seguida todos os outros *tweets* marcados com essa palavra-chave. Os marcadores (#) podem aparecer em qualquer parte de um *tweet* (início, meio ou fim). As palavras marcadas que passam a ser muito populares são, muitas vezes, tendências (*trending topics*) sendo apresentadas mais à frente. Para se ter uma noção da dimensão e do nível de adoção das *hashtags*, o próprio termo “*hashtag*” passou recentemente a fazer parte do dicionário de inglês de *Oxford* [OxfordWords blog, 2014] [Oed.com, 2014].

Por exemplo no *tweet* da Figura 10, @eddie incluiu o marcador #FF. Os utilizadores criaram essa abreviação para "Follow Friday" (Seguir Sexta-feira), uma tradição semanal em que os utilizadores recomendam pessoas que outros deveriam seguir no Twitter.

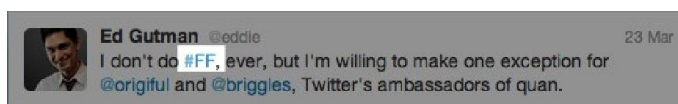



Figura 10 – Exemplo de um *tweet* que contém o símbolo “#”

O uso da *hashtag* tornou-se numa forma convencionada de marcação de mensagens que tem como objetivo a associação dessas mesmas mensagens do Twitter com certos eventos ou contextos [Chang, 2010]. Por exemplo na Figura 11, surge uma lista com alguns *tweets* que usam a *hashtag* #MSW\_RADICAL sugerida em Agosto de 2014 pela Sic Radical, canal de televisão que transmitiu o festival Meo Sudoeste. Os utilizadores ao escreverem essa *hashtag* conseguem depois visualizar tudo o que se está a falar no Twitter sobre esse evento.



Figura 11 – Alguns dos *tweets* com a *hashtag* #MSW\_RADICAL  
Fonte: [https://twitter.com/hashtag/msw\\_radical?f=realtime&src=hash](https://twitter.com/hashtag/msw_radical?f=realtime&src=hash)

### 2.1.4 Tipos de Contas no Twitter

No Twitter existem diversos tipos de contas de utilizador, sendo que a maioria são de utilizadores comuns, ou seja, do público em geral. No entanto existem também contas oficiais de utilizadores com outra expressão a nível local e mundial. Esse tipo de contas apresentam-se geralmente acompanhadas do símbolo , que assinala que a conta se trata de uma “conta verificada”. O Twitter concede este reconhecimento a contas que tenham elevados níveis de procura por parte de utilizadores em áreas tão distintas como a música, desporto, política, religião, notícias, entre outras [Support.twitter.com, 2014b].

Seguidamente são apresentados alguns exemplos de contas verificadas.



Figura 12 – Contas oficiais e verificadas do Twitter (Personalidades)



Figura 13 – Contas oficiais e verificadas do Twitter (Desportistas)



Figura 14 – Contas oficiais e verificadas do Twitter (Músicos)



Figura 15 – Contas oficiais e verificadas do Twitter (Eventos)



Figura 16 – Contas oficiais e verificadas do Twitter (Notícias)

O Twitter efetua uma verificação regular de forma a facilitar a localização das pessoas ou entidades que os utilizadores comuns procuram. De referir ainda que o Twitter não aceita solicitações de verificação do público em geral.

## 2.2 Twitter API

Como já foi dito anteriormente o Twitter é considerado uma rede social e ao mesmo tempo um serviço de *microblogging* em tempo real, que permite aos seus utilizadores partilhar os seus estados através dos *tweets* que aparecem nas suas *timelines*. Esses *tweets* podem conter nos seus 140 caracteres, referências a entidades e/ou locais.

É vital ter um conhecimento profundo de todos estes elementos (utilizadores, *tweets* e *timelines*), para assim tornar mais intuitivo e eficiente o processo de comunicação com a Twitter API (*Application Program Interface*) e posteriormente o próprio processo de extração de dados.

Os *tweets* são a essência do Twitter e apesar de estarem diretamente relacionados com a mudança de um estado de um utilizador em particular e de terem apenas os já mencionados 140 caracteres de dimensão máxima, contêm mais informação do que aquela que é apresentada ao utilizador comum. Assim, para cada *tweet*, tem-se o respetivo texto associado a mais dois elementos fulcrais do Twitter, as entidades (*entities*) e locais (*places*). As entidades são habitualmente referências de utilizadores, *hashtags* (#) e URL (*Uniform Resource Locator*). Os locais são localizações do mundo real, como por exemplo, Portugal ou outro país qualquer. De referir que um local tanto pode ser relativo ao sítio onde o *tweet* foi criado ou apenas uma simples referência feita no próprio *tweet* a determinado local.

As *timelines* são os conjuntos de *tweets* ordenados por ordem cronológica, sendo importante referir que um utilizador consegue ver a sua própria *timeline*, assim como a *timeline* de todos os utilizadores que siga nesse momento.

Citando informação proveniente da documentação da área de desenvolvimento do Twitter existem quatro objetos fundamentais na Twitter API [Dev.twitter.com, 2014a]:

- *Tweets* – São o elemento base no Twitter. Os utilizadores escrevem *tweets* e este processo é genericamente denominado de *status update*. Os *tweets* podem ser incorporados ou embebidos, respondidos, classificados como favoritos ou não favoritos e apagados.
- *Utilizadores* – Os utilizadores podem ser qualquer pessoa ou qualquer coisa. Podem escrever *tweets*, seguir outros utilizadores, criar listas, ter uma *timeline*, podem ser mencionados e podem ser procurados através das pesquisas.
- *Entidades* – Fornecem metadados (dados sobre dados) e informação adicional sobre conteúdo publicado no Twitter. As entidades nunca se dissociam do conteúdo que descrevem. Na Twitter API 1.1 as entidades são retornadas juntamente com os *tweets*.
- *Locais* – São localizações específicas que tem associadas coordenadas georreferenciáveis. Podem ser associados aos *tweets* ao especificar o *place\_id* quando

se publica o *tweet*. Os locais podem ser pesquisados e os *tweets* podem ser encontrados por *place\_id*.

### 2.2.1 Versão 1.1

A versão atual da Twitter API é a 1.1 tendo sido lançada em Setembro de 2012 funcionando nessa altura em paralelo com a versão 1 [Costa, 2012]. No entanto, desde Junho de 2013 a versão 1.1 é a única versão em funcionamento [Singletary, 2013].

A forma da API aceder aos dados do Twitter pode ser classificada segundo dois tipos distintos de arquitetura [Kumar, Morstatter and Liu, 2014]:

1. REST API – Baseada na arquitetura REST (*Representational State Transfer*) segue uma estratégia *pull* para retornar dados, sendo que o utilizador necessita de especificar bem os dados que pretende obter. É usada por muitas aplicações para por exemplo publicar *tweets*, “seguir” contas, criar listas entre outras.
2. Streaming API – Neste tipo de API é fornecido um *stream* contínuo de dados do Twitter. Estas API’s usam uma estratégia *push* para retornar dados. Assim que um pedido de acesso a dados é efetuado, a *Streaming API* fornece um *stream* contínuo de atualizações sem que para isso o utilizador realize um novo pedido. É usada por utilizadores, que pretendam receber em tempo real, um *stream* dos *tweets* públicos do Twitter. A *Streaming API* possui três tipos distintos de *streams*:
  - a. *Public Streams* – Contêm os *tweets* públicos no Twitter.
  - b. *User Streams* – Contêm os *tweets* do utilizador.
  - c. *Site Streams* – São *streams* multiutilizador e são direcionados para aplicações que acedem aos *tweets* de múltiplos utilizadores.

As Figuras 17 e 18 apresentam algumas das diferenças entre a REST API e a *Streaming API* [Dev.twitter.com, 2014b].

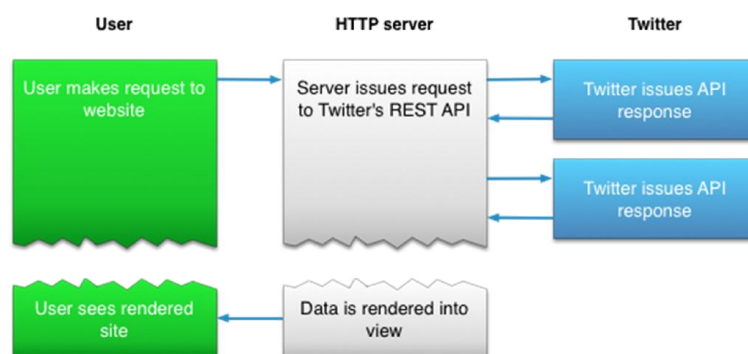


Figura 17 – Twitter REST API

Fonte: [https://g.twimg.com/dev/documentation/image/streaming-intro-1\\_1.png](https://g.twimg.com/dev/documentation/image/streaming-intro-1_1.png)

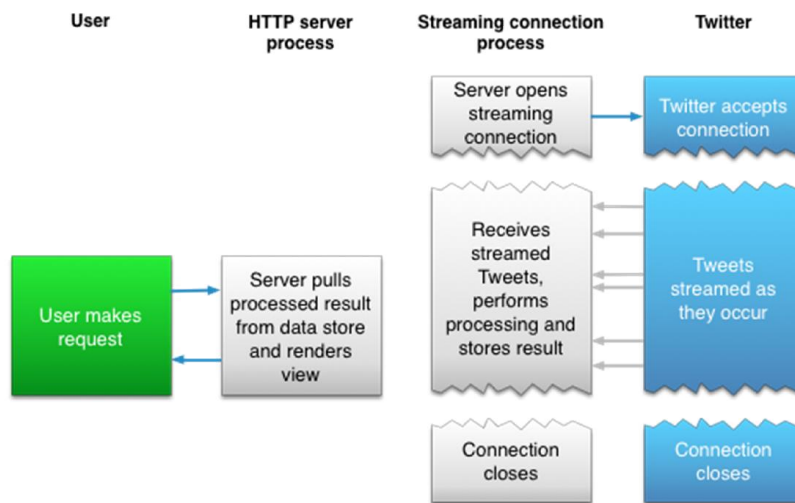


Figura 18 – Twitter *Streaming* API

Fonte: [https://g.twimg.com/dev/sites/default/files/images\\_documentation/streaming-intro-2\\_1.png](https://g.twimg.com/dev/sites/default/files/images_documentation/streaming-intro-2_1.png)

Os pedidos à Twitter API contêm parâmetros que podem incluir *hashtags*, palavras-chave, zonas geográficas e *userID's* do Twitter.

Por seu lado as respostas da Twitter API são retornadas no formato JSON (*JavaScript Object Notation*). Trata-se de um formato de texto que permite a troca de dados estruturados entre diversas linguagens de programação. Apresenta um pequeno conjunto de regras para a representação estruturada dos dados. Um texto em JSON possui uma sequência de símbolos assim como texto e números. Os *símbolos* existentes são:  $[ \{ \} ] : , true false null$ . Os valores encapsulados e permitidos no JSON podem ser objetos, *arrays*, números, *strings*, booleanos ou *null*.

A Twitter API só pode ser acedida através de pedidos autenticados. O Twitter usa o *OAuth* (*Open Authentication*) descrito seguidamente e cada pedido terá de incorporar credenciais de utilizador válidas. De referir também que a API encontra-se limitada quanto ao número de pedidos (*rate limit*).

As aplicações e páginas *web* que usam o Twitter são construídas usando a Twitter API. Quando estes pressupostos estão reunidos o utilizador pode deparar-se com uma limitação estabelecida pela própria API, que surge sob a forma de um aviso que informa o utilizador de que foram efetuadas demasiadas tentativas ao tentar extrair dados do Twitter num determinado período de tempo [Support.twitter.com, 2014c].



## 2.2.2 Limites

Na primeira versão da Twitter API uma aplicação autenticada podia efetuar 350 pedidos/hora por *access token* ou 150 pedidos/hora sem autenticação. No entanto, na versão atual, é permitido um maior leque de pedidos distintos dentro das restrições existentes. A saber, os métodos *GET* são de 15 e 180 chamadas a cada 15 minutos, e as aplicações tem de estar sempre autenticadas [Dev.twitter.com, 2014c].

Outro aspeto importante refere-se à duração da janela temporal, que é de 15 minutos na versão atual face aos 60 minutos da versão 1.

Os pedidos que são contabilizados para que se atinja o limite são aqueles que tem como intuito retornar informação do Twitter para que o utilizador a possa ler.

Se uma aplicação que está a usar a REST API transmite ao utilizador que o limite foi atingido torna-se visível que algumas partes do Twitter não são atualizadas. Usando a linha de comandos para efetuar pedidos à Twitter API as mensagens de erro geradas quando se atingem os limites de utilização podem ter o seguinte aspeto:

```
TwitterHTTPError: Twitter sent status 429 for URL: 1.1/trends/place.json
using parameters:
(id=23424925&oauth_consumer_key=JRkkGcFsQT0Kr3KpkSg&oauth_nonce=11061017341
342763461&oauth_signature_method=HMAC-
SHA1&oauth_timestamp=1401211325&oauth_token=579429366-
ZxEM2MKBDCMm4kpFuVfzKjr7h230q54KrRHfpE1&oauth_version=1.0&oauth_signature=
0d5xAck%2BJVw2GIhdWQJKDGFavtc%3D)
details: {"errors":[{"message":"Rate limit exceeded","code":88}]}
```

As pesquisas têm como limite as 180 consultas por cada 15 minutos tendo de seguir sempre a norma de autenticação por parte da aplicação.

Apesar de todas as capacidades e potencialidades do Twitter, existem limitações na sua utilização, de forma a prevenir o sobrecarregamento da plataforma.

Os limites definidos pela Twitter API não se aplicam nas ações de publicação de *tweets* nem nos processos de “seguir” novas contas de utilizador. Estas ações são controladas por limites do Twitter e não da Twitter API [Support.twitter.com, 2014d], encontrando-se descritos na Tabela 5.

Tabela 5 – Limites definidos pelo Twitter

<i>Características</i>	<i>Limitações</i>
<i>Mensagens Diretas</i>	<i>250/dia</i>
<i>Tweets ou Retweets</i>	<i>2.400/dia</i>
<i>Alterações de conta de e-mail</i>	<i>4/hora</i>
<i>A seguir (diariamente)</i>	<i>1.000/dia</i>
<i>A seguir (com base na conta)</i>	<i>Depois de seguir 2.000 contas existe um maior controlo no processo limitando as tentativas subsequentes de “seguir”</i>

### 2.2.3 Ligação à Twitter API

Inicialmente e para se poderem fazer pedidos à Twitter API, é necessária a criação de uma aplicação no *site* de desenvolvimento de aplicações do Twitter (<https://dev.twitter.com/apps>). Este passo é vital neste processo pois é a forma mais recorrente que o utilizador/programador tem, para conseguir o acesso à API e assim começar a fazer consultas com o intuito de extrair dados que considere relevante.

A aplicação *standard* a ser criada apenas tem de ser do tipo *read-only* para se conseguir aceder com sucesso à API. Neste contexto aquilo que se está a fazer ao criar uma aplicação com a própria conta é autorizar o acesso aos dados da conta. Ao ser criada com sucesso são geradas quatro credenciais *OAuth* (*consumer key*, *consumer secret*, *access token* e *access token secret*) que são recorrentemente usadas nos vários pedidos efetuados através da Twitter API sendo essas credenciais pessoais e intransmissíveis.

Após serem geradas as credenciais *OAuth*, podem-se começar a efetuar as consultas pretendidas à Twitter API.

```
import twitter
CONSUMER_KEY = 'xxxxxxxxxx'
CONSUMER_SECRET = 'xxxxxxxxxx'
OAUTH_TOKEN = 'xxxxxxxxxx'
OAUTH_TOKEN_SECRET = 'xxxxxxxxxx'
auth = twitter.oauth.OAuth(OAUTH_TOKEN, OAUTH_TOKEN_SECRET, CONSUMER_KEY,
CONSUMER_SECRET)
twitter_api = twitter.Twitter(auth=auth)
print twitter_api
```

Código 1 – Autorização de uma aplicação para aceder aos dados da conta do Twitter

O resultado da execução deste código dá origem a um objeto da Twitter API.

```
<twitter.api.Twitter object at 0x00000000517D518>
```

Com este resultado, pode-se concluir que as credenciais *OAuth* foram usadas com sucesso para obter autorização e desta forma começar a fazer consultas à Twitter API [Russell, 2013].

### 2.2.4 Mecanismo de Autenticação

O *OAuth* é um protocolo de autenticação que permite ao utilizador deixar que uma determinada aplicação aja em seu nome sem ter de partilhar qualquer tipo de *password*.

Trata-se de um *standard* a nível de autenticação, usado pelo Twitter para fornecer acesso a informação protegida. As *passwords* são altamente vulneráveis e este protocolo fornece uma alternativa viável a nível de segurança com melhoramentos face a outros protocolos. Também aumenta a confiança por parte do utilizador na aplicação já que a sua *password* do Twitter nunca é partilhada com aplicações de terceiros. A autenticação dos pedidos à Twitter API é sempre efetuada usando este protocolo.

Seguidamente apresenta-se o *workflow* do protocolo *OAuth* para aceder à Twitter API dividido em quatro passos.

- 1- As aplicações são vistas como consumidores, e todas as aplicações devem estar registadas no Twitter. No decorrer deste processo, é gerada uma chave “*consumer key*” e um “*consumer secret*” que tem de ser usados pela aplicação para se autenticar no Twitter.
- 2- A aplicação usa o “*consumer key*” e o “*consumer secret*” para criar uma ligação única ao Twitter para o qual o utilizador é direccionado para se autenticar. O utilizador autoriza a aplicação a autenticar-se em seu nome e o Twitter verifica a sua identificação e lança automaticamente um verificador em forma de código PIN.
- 3- O utilizador visualiza e fornece esse código PIN à aplicação. A aplicação usa esse código para pedir um “*Access Token*” e um “*Access Secret*” ao utilizador.
- 4- Ao usar o “*Access Token*” e o “*Access Secret*”, a aplicação autentica o utilizador no Twitter e começa a efetuar chamadas à API em nome do utilizador.

O “*Access Token*” e o “*Access Secret*” não mudam para um determinado utilizador e podem ser armazenados em cache pela aplicação para serem usados em pedidos futuros.

Este processo encontra-se também descrito na Figura 19 [Kumar, Morstatter and Liu, 2014].

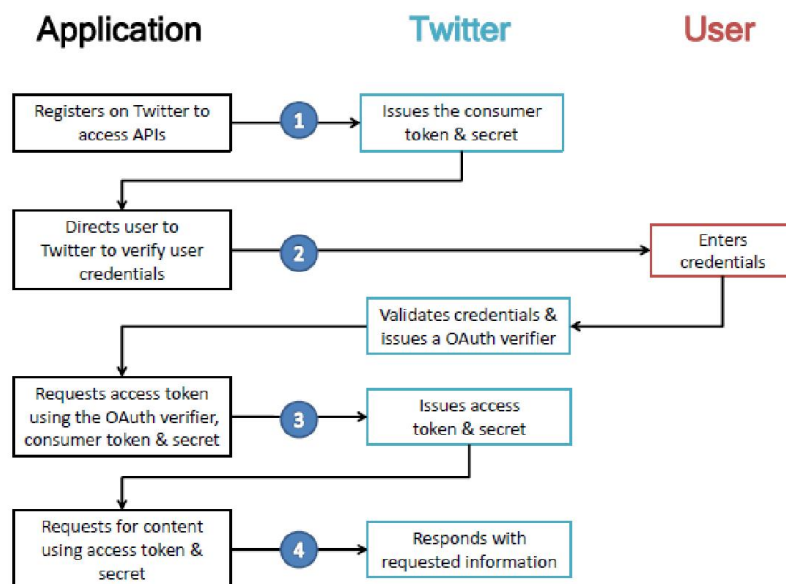


Figura 19 – Funcionamento do protocolo *OAuth* para aceder à Twitter API

## 2.3 Tendências do Momento

O Twitter como rede social usada mundialmente, tem a capacidade de rapidamente lançar e espalhar tendências. Vários autores realizaram diversos estudos relacionados com a importância das tendências do Twitter numa tentativa de prever outro tipo de tendências quer sejam de ordem cultural, tendências nos mercados bolsistas ou quaisquer outras tendências.

Os autores Asur e Huberman, [Asur and Huberman, 2010], assumem claramente que as redes sociais, neste caso o Twitter, podem assumir um papel preponderante na previsão de tendências de qualquer tipo. Neste caso os autores efetuaram uma análise a três milhões de *tweets* construindo um modelo de regressão linear que lhes permitiu prever as receitas que determinados filmes poderiam vir a ter mesmo antes das suas estreias no cinema.

Os autores Bollen, Mao e Zeng [Bollen, Mao and Zeng, 2011], analisaram o conteúdo de milhões de *tweets* e através do uso de ferramentas de deteção de estados de humor ou comportamentais como por exemplo o *OpinionFinder* ou o *Google-Profile of Mood States*, concluíram que mudanças nesses estados podem estar eventualmente correlacionados ou mesmo prever alterações registadas nos valores dos índices dos mercados de *stocks* neste caso de estudo, o índice *Dow Jones Industrial Average* ou *Dow Jones*.

Uma tendência ou *trending topic* do Twitter é uma expressão que é marcada pelos utilizadores a uma taxa superior em comparação com outras. Essa expressão tanto pode ser uma frase como apenas uma palavra. Por vezes os *trending topics* tornam-se populares através de uma ação concertada de diversos utilizadores para conseguir que determinado evento atinja o desejado nível de popularidade. Os *trending topics* são por vezes o resultado de esforços conjuntos e manipulações de fãs de certas celebridades e fenómenos culturais. Por outro lado um evento relacionado com desporto, música ou política por exemplo, pode eventualmente chegar a ser um *trending topic* consoante o número de utilizadores que o referirem no Twitter. Resumindo, os *trending topics* ajudam muitas das vezes a entender o que está a acontecer no momento.

O Twitter dá especial enfoque às tendências do momento através da apresentação de uma secção na página inicial dedicada para esse fim. As dez tendências do momento são geradas através de um algoritmo proprietário do próprio Twitter [Cheong and Lee, 2009].

Os mesmos autores [Cheong and Lee, 2009] classificaram os tópicos/tendências em três tipos distintos quanto à sua duração (longa, média e curta).

Assim os tópicos de longa duração são pouco discutidos sobretudo devido a serem termos pouco claros e mesmo possuindo um pico (*spike*) de notoriedade, esse pico decai com alguma velocidade. Estes tópicos são pesquisáveis durante vinte dias através da funcionalidade de pesquisa da Twitter API e raramente o resultado dessa pesquisa ultrapassa os 1.500 *tweets*. Os tópicos de média duração podem ser termos genéricos que são habitualmente

mencionados mas esse facto não garante que surjam num elevado número de *tweets*. Por último, os tópicos de curta duração surgem em grande quantidade sendo bastante mencionados e não apresentam picos de relevância.

Um dado curioso relacionado com as tendências no Twitter é que estas possuem habitualmente uma duração curta. Por outras palavras, a maior tendência registada num determinado dia raramente assume contornos de tendência nos dias seguintes, sendo claramente um indicador da efemeridade das tendências nesta rede social.

Num estudo recente foram analisadas as tendências mais fortes registadas entre dia 01/12/2013 e dia 17/12/2013 na cidade de Londres [Lambrecht, Tucker and Wiertz, 2014]. Durante este intervalo de tempo a tendência principal nunca se repetiu, tendo sido apenas verificado a presença em duas ocasiões de uma determinada tendência nas dez tendências dos três seguintes dias. Outro dado intrigante deveu-se ao facto de nenhuma das tendências principais ter aparecido nas dez tendências mais fortes do dia anterior, o que vem comprovar o rápido aparecimento das tendências no Twitter e a sua rápida propagação.

O Twitter faz a distinção entre dois tipos de tendências, as que são patrocinadas por empresas e as que emergem organicamente. A *hashtag* #MSW\_RADICAL referida anteriormente na secção 2.1.3, foi um exemplo de uma tendência promovida por um canal de televisão privado aquando da realização de um evento de música.

Neste trabalho de mestrado foram escolhidos sete locais para análise de tendências. Seis dizem respeito a países e a sétima escolha refere-se às tendências a nível mundial.

Após ser garantido o acesso à Twitter API, é possível começar a explorar as tendências do momento nos locais escolhidos. No entanto a localização das tendências pode ser facilmente parametrizável através do ajuste do respetivo código WOEID proprietário da empresa *Yahoo*.

Na Tabela 6, estão presentes os códigos WOEID usados nas consultas à API para obter as tendências em várias localizações [Developer.yahoo.com, n.d.]. O próprio conceito WOEID é explicado na secção 2.4.

Tabela 6 – Códigos WOEID usados para a obtenção das tendências

Localização	WOEID
Mundial	1
Portugal	23424925
Espanha	23424950
Estados Unidos	23424977
Brasil	23424768
Argentina	23424747
Reino Unido	23424975

### 2.3.1 Extração das Dez Tendências de um Local

O método `GET trends/place` devolve as dez tendências ou *trending topics* para um determinado local, sendo para isso necessário fornecer o identificador WOEID do local que se pretende analisar.

A resposta ao pedido vem na forma de um *array* de tendências contendo ainda outros elementos como o parâmetro da consulta (*query parameter*) usado para efetuar pesquisas de tópicos no *Twitter Search*, assim como o URL dessa pesquisa [Dev.twitter.com, 2013].

Seguidamente é apresentado código em linguagem Python, que se encontra estruturado de forma a garantir em primeiro lugar, a correta declaração das credenciais *OAuth* para estabelecer ligação com a Twitter API. Em segundo lugar define-se o local a analisar, neste caso Portugal (WOEID=23424925) e por último apresentam-se os resultados no formato JSON.

```
import twitter
import json
CONSUMER_KEY = 'xxxxxxxxxx'
CONSUMER_SECRET = 'xxxxxxxxxx'
OAUTH_TOKEN = 'xxxxxxxxxx'
OAUTH_TOKEN_SECRET = 'xxxxxxxxxx'
auth = twitter.oauth.OAuth(OAUTH_TOKEN, OAUTH_TOKEN_SECRET, CONSUMER_KEY,
CONSUMER_SECRET)
twitter_api = twitter.Twitter(auth=auth)
PT_WOE_ID = 23424925
pt_trends = twitter_api.trends.place(_id=PT_WOE_ID)
print json.dumps(pt_trends, indent=1)
```

Código 2 – Extração das tendências de Portugal (WOEID=23424925)

O resultado em JSON apresenta os tópicos mais revelantes em Portugal no dia 13 de Março de 2014 às 20h37m em que se destacam os tópicos relacionados com o jogo da Liga Europa entre o Tottenham e o Benfica com os termos “*#CarregaBenfica*”, “*White Hart Lane*” e “*Tottenham*”.

```
[
  {
    "created_at": "2014-03-13T20:37:06Z",
    "trends": [
      {
        "url":
"http://twitter.com/search?q=%2323AnosDeOrgulhoLSNossoEternoGurizinho",
        "query": "%2323AnosDeOrgulhoLSNossoEternoGurizinho",
        "name": "#23AnosDeOrgulhoLSNossoEternoGurizinho",
        "promoted_content": null,
        "events": null
      },
      {
        "url": "http://twitter.com/search?q=%2310fatosSobreMim",
        "query": "%2310fatosSobreMim",
        "name": "#10fatosSobreMim",
        "promoted_content": null,
        "events": null
      }
    ]
  }
]
```

```

    },
    {
      "url": "http://twitter.com/search?q=%23ZAwards",
      "query": "%23ZAwards",
      "name": "#ZAwards",
      "promoted_content": null,
      "events": null
    },
    {
      "url": "http://twitter.com/search?q=%23CeTopaLuanSantana",
      "query": "%23CeTopaLuanSantana",
      "name": "#CeTopaLuanSantana",
      "promoted_content": null,
      "events": null
    },
    {
      "url": "http://twitter.com/search?q=%23CarregaBenfica",
      "query": "%23CarregaBenfica",
      "name": "#CarregaBenfica",
      "promoted_content": null,
      "events": null
    },
    {
      "url": "http://twitter.com/search?q=%22Hi+Portugal%22",
      "query": "%22Hi+Portugal%22",
      "name": "Hi Portugal",
      "promoted_content": null,
      "events": null
    },
    {
      "url": "http://twitter.com/search?q=Fly",
      "query": "Fly",
      "name": "Fly",
      "promoted_content": null,
      "events": null
    },
    {
      "url": "http://twitter.com/search?q=%22White+Hart+Lane%22",
      "query": "%22White+Hart+Lane%22",
      "name": "White Hart Lane",
      "promoted_content": null,
      "events": null
    },
    {
      "url": "http://twitter.com/search?q=Tic",
      "query": "Tic",
      "name": "Tic",
      "promoted_content": null,
      "events": null
    },
    {
      "url": "http://twitter.com/search?q=Tottenham",
      "query": "Tottenham",
      "name": "Tottenham",
      "promoted_content": null,
      "events": null
    }
  ],
  "as_of": "2014-03-13T20:40:46Z",

```

```

"locations": [
  {
    "woeid": 23424925,
    "name": "Portugal"
  }
]
}
]

```

Código 3 – Dez tendências de Portugal em 13/03/14 – 20h37m

Apesar da consulta (código 2) ter sido executada às 20h40m o resultado obtido (código 3) reflete as tendências das 20h37m em Portugal. Isto deve-se ao facto de estas tendências ficarem em *cache* durante alguns minutos, ou seja, por mais chamadas que se se façam à API nesse período apenas se está a fazer com que os limites de utilização da API 1.1 (15 tentativas) sejam atingidos na respetiva janela temporal impossibilitando outro tipo de pedidos à API e limitando assim a atividade do utilizador.

### 2.3.2 Tendências Comuns em Diferentes Localizações

É possível efetuar a intersecção de dois conjuntos de tendências, para assim identificar possíveis tendências com maior expressão em partes diferentes do mundo.

Neste caso escolheu-se efetuar a extração das tendências a nível mundial (WOEID=1) e em Espanha (WOEID=23424950) às 10h55 do dia 4 de Maio de 2014.

O código mostrado em seguida cria dois conjuntos de tendências e depois identifica elementos comuns entre eles.

```

world_trends_set = set([trend['name']
  for trend in world_trends[0]['trends']])
es_trends_set = set([trend['name'] for trend in es_trends[0]['trends']])
common_trends = world_trends_set.intersection(es_trends_set)
print common_trends

```

Código 4 – Intersecção das tendências em Espanha e no Mundo em 04/05/14 – 10h55m

O resultado apresentado foi o seguinte:

```

set([u'#FelizDiaDeLaMadre'])

```

Este resultado é demonstrativo da importância que o dia da mãe possui em algumas zonas do planeta fazendo com que inclusive se consiga tornar uma tendência a nível mundial.

Os países que tem mais facilidade em tornar tendências locais em tendências mundiais, são como é óbvio, aqueles em que o Twitter atinge uma maior dimensão, ou seja, aqueles que possuem um maior número de utilizadores.



### 2.3.3 Verificar Termos em Páginas Web

É possível fazer a verificação do número de vezes que um determinado tópico aparece num determinado *site*. No seguinte exemplo escolheu-se o termo “*Real Madrid*”, o *site* desportivo espanhol AS e o *site* generalista português *Jornal de Notícias*.

```
import urllib2
import re
html_content_as = urllib2.urlopen('http://www.as.com').read()
html_content_jn = urllib2.urlopen('http://www.jn.pt').read()
matches_as = re.findall('Real Madrid', html_content_as);
matches_jn = re.findall('Real Madrid', html_content_jn);

if len(matches_as or matches_jn) == 0:
    print 'O termo escolhido não existe nas páginas em questão!'
else:
    print matches_as, matches_jn
```

Código 5 – Verificação de termos em páginas web

O resultado obtido foi de 41 referências em (“*www.as.com*”) e apenas uma referência em (“*www.jn.pt*”) como demonstrado no seguinte conjunto.

```
['Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid',
'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid',
'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid',
'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid',
'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid',
'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid',
'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid',
'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid', 'Real Madrid',
'Real Madrid'] ['Real Madrid']
```

## 2.4 Identificador WOEID

Em termos de referência geográfica de dados, o serviço *Yahoo! GeoPlanet* funciona como um elo de ligação entre o mundo real e o mundo virtual. Este serviço foi criado para efetuar a gestão de todos os locais no planeta [Developer.yahoo.com, n.d.] através da atribuição de identificadores georreferenciáveis WOEID (*Where On Earth ID*).

Um identificador WOEID é constituído por 32 bits, e é criado e fornecido pelo serviço *Yahoo! GeoPlanet*. Uma vez atribuído, este identificador nunca pode ser alterado ou mesmo reciclado. Se um WOEID se tornar obsoleto, o mapeamento será efetuado para o seu sucessor ou *parent ID* (identificador pai).

Sabendo que a geografia mundial não se trata de algo estático, torna-se necessária a implementação de sistemas que permitam uma atualização constante e que garantam precisão e exatidão nos dados a nível geográfico. Por exemplo, cidades que surjam através da fusão de outras cidades, ou até mesmo nações que mudem as suas fronteiras geográficas passando a abranger territórios até então pertencentes a outros países. Esta reorganização acaba por ser um processo em constante evolução, sendo também possível tornar obsoleto determinado WOEID quando se verificam situações de locais repetidos ou locais inválidos.

Ao nível da cobertura, o *Yahoo! Geoplanet* fornece informação para aproximadamente seis milhões de locais a nível global [Developer.yahoo.com, n.d.]. Essa cobertura varia consoante o país e inclui milhares de áreas administrativas distintas, com históricos incluídos, com milhões de zonas e códigos postais de aproximadamente 150 países.

### 2.4.1 Estrutura

O *Yahoo! Geoplanet* usa um modelo hierárquico para os locais fornecendo uma consistência a nível geográfico tanto vertical como horizontalmente. Este modelo garante que os locais em cada camada da hierarquia sobrepõem corretamente os locais correspondentes noutras camadas, fazendo com que as relações existentes ao nível geográfico sejam preservadas.

Esta hierarquia permite que se efetuem pesquisas num contexto geográfico a todos os locais representados por um WOEID. Na hierarquia é possível distinguir os seguintes locais administrativos: país, estado ou província e cidade. Cada local pertence a uma entidade superior (em termos geográficos) e pode eventualmente conter entidades inferiores penduradas em si.

O *Yahoo! Geoplanet* permite aos utilizadores a identificação de locais que tenham determinadas relações. Seguidamente são apresentadas as relações existentes nos locais:

- *Parent* (Pai) - O superior direto de um local. Por exemplo o estado da Califórnia com o (WOEID: 2347563) é “filho” do país Estados Unidos com o (WOEID: 23424977), ou seja, no sentido inverso da relação, os Estados Unidos são o “pai” da Califórnia. Nesta versão do *Yahoo! Geoplanet* os locais possuem apenas um local pai;

- *Children* (Filho) - O inferior direto de um local. Novamente o exemplo do estado da Califórnia com o (WOEID: 2347563) contém 58 zonas/municípios;
- *Neighbour* (Vizinho) - São locais adjacentes, ou seja, o estado do Nevada (WOEID: 2347587) é adjacente ao estado do Arizona (WOEID: 2347561). No entanto, um local pode não partilhar a mesma fronteira geográfica com um local vizinho apesar de estarem perto um do outro;
- *Sibling* (Irmão) - O estado da Califórnia (WOEID: 2347563) tem 49 estados denominados de irmãos;
- *Belong to* (Pertence a) - Um local filho pertence a um local pai;
- *Ancestors* (Antepassados) – São locais que são pais de outros locais que por sua vez também são pais, ou seja, formam uma cadeia de parentesco multinível.

## 2.4.2 Tipos de Locais

Os locais são categorizados de forma a permitir a sua correta identificação. Este tipo de locais (Tabela 7) tem códigos únicos que permitem uma melhor filtragem de resultados por parte de alguns recursos ou plataformas que usem os “locais” do *Yahoo! Geoplanet*.

Tabela 7 – Tipos de locais

<i>Tipo de Local</i>	<i>Código</i>	<i>Descrição</i>
<i>Continent</i>	29	<i>O GeoPlanet foi construído usando um modelo de sete continentes: Ásia (24865671), África (24865670), América do Norte (24865672), América do Sul (24865673), Antártida (28289421), Europa (24865675) e Pacífico que engloba (Austrália, Nova Zelândia e outras ilhas do Oceano Pacífico - 24865674).</i>
<i>Country</i>	12	<i>Países e territórios dependentes definidos pelo standard ISO 3166-1.</i>
<i>Admin</i>	8	<i>Zonas administrativas primárias de um país.</i>
<i>Admin2</i>	9	<i>Zonas administrativas secundárias de um país.</i>
<i>Admin3</i>	10	<i>Zonas administrativas terciárias de um país.</i>
<i>Town</i>	7	<i>Locais populosos dentro de um país podendo ser tanto cidades como vilas.</i>
<i>Suburb</i>	22	<i>Zonas dentro de uma cidade.</i>
<i>Postal Code</i>	11	<i>Códigos postais de um determinado país.</i>
<i>Supername</i>	19	<i>Local que se refere a uma região composta por vários países ou um país antigo que se dividiu em vários países. Alguns exemplos: Ibéria, Escandinávia, USSR ou Jugoslávia.</i>
<i>Colloquial</i>	24	<i>Local que se refere a uma zona que possui um determinado nome como por exemplo: Algarve, Riviera Francesa.</i>
<i>Time Zone</i>	31	<i>Local que se refere a uma área definida pelo standard de Olson frequentemente relacionado com os fusos horários. Alguns exemplos: América/Los Angeles, Ásia/Tóquio, Europa/Madrid.</i>

No que diz respeito à representação dos locais, foram definidos dois tipos distintos sendo eles, a representação curta composta por três elementos (*WOEID*, *placeTypeName*, *name*) e a representação longa que apresenta alguns elementos adicionais que permitem definir um determinado local com mais precisão. A Tabela 8 apresenta todos os códigos existentes tanto para continentes como para oceanos disponibilizados pelo serviço *Yahoo! Geoplanet*.

Tabela 8 – Códigos de continentes e oceanos

<i>Código</i>	<i>Descrição</i>
AF	África
AN	Antártida
AS	Ásia
AU	Austrália
EU	Europa
NA	América do Norte
SA	América do Sul
AR	Oceano Ártico
AT	Oceano Atlântico
IN	Oceano Índico
PA	Oceano Pacífico
SO	Oceano Antártico

### 2.4.3 Língua

O *Yahoo! Geoplanet* é compatível com UTF-8 (*8bit Unicode Transformation Format*) e suporta nomes de locais para usar em múltiplas línguas como, por exemplo, Inglês, Francês, Alemão, Italiano e Espanhol, assim como línguas com caracteres especiais como o Japonês, Chinês e Coreano. A Tabela 9 apresenta as línguas suportadas pelo *Yahoo! Geoplanet* e como se pode observar existem distinções entre línguas semelhantes mas que são faladas em diferentes partes do globo. Por exemplo, o inglês falado nos Estados Unidos e o inglês falado no Reino Unido apresentam códigos distintos.

Tabela 9 – Línguas suportadas pelo *Yahoo! Geoplanet*

<i>Código</i>	<i>Descrição</i>
en-US	US English
en-GB, en	UK English
en-CA	CA English
en-AU	AU English
en-ZA	ZA English
fr-FR, fr	FR French
fr-CA	CA French
it-IT, it	Italian
es-ES, es	Spanish
es-US	US Spanish
es-MX	MX Spanish
de-DE, de	DE German
de-AT	AT German
zh-TW	TW Chinese (traditional)
zh-HK	HK Chinese (traditional)
zh-MO	MO Chinese (traditional)
zh-CN	CN Chinese (simplified)
zh-SG	SG Chinese (simplified)
ja-JP, ja	JP Japanese
ko-KR	KR Korean
vi-VN	VN Vietnamese
el-GR	GR Greek
ru-RU	RU Russian

#### 2.4.4 Consistência Posicional

Os locais no *Yahoo! Geoplanet* são representados em coordenadas (latitude/longitude) que usam o *standard* WGS 84 (*World Geodetic System*) estabelecido em 1984 e revisto em 2004 difundido pelo IERS (*International Earth Rotation and Reference Systems Service*).

O WGS 84 é um sistema de referência fixado e centrado na Terra e ainda um sistema de dados geodésicos. Baseia-se num conjunto de constantes e parâmetros que descrevem o tamanho da Terra, a sua forma, gravidade e os seus campos geomagnéticos. É o *standard* usado como sistema de referência para o GPS (*Global Positioning System*) [UNOOSA, 2012]. Sendo um sistema convencional de referência terrestre, as observações efetuadas estão reduzidas a uma mesma época, eliminando-se o efeito do movimento do pólo nas coordenadas determinadas.

É um sistema tridimensional, comum e ortogonal. A origem do sistema de coordenadas WGS 84 também é usada como o centro geométrico do elipsóide WGS 84 e o eixo Z serve como eixo de rotação desse elipsóide de revolução.

Ao analisar a Figura 20, a origem surge como o centro de massa da terra (geocêntrico). O eixo Z encontra-se na direção do IRP (*IERS Reference Pole*). O eixo X representa a intersecção do meridiano de referência IRM (*IERS Reference Meridian*) com o plano que passa pela origem e é normal ao eixo Z. Por último o eixo Y define um sistema ortogonal com os outros dois (X e Z).

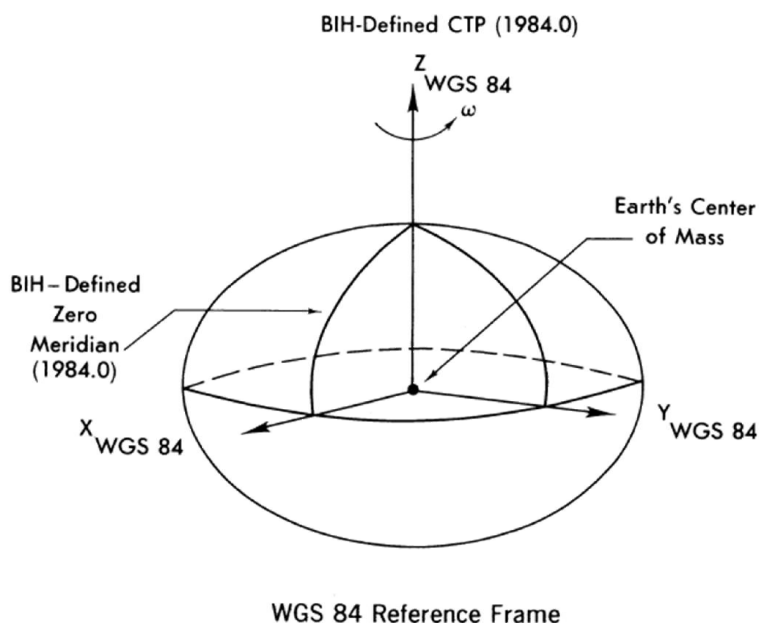


Figura 20 – Elipsóide WGS 84

Fonte:

[http://upload.wikimedia.org/wikipedia/commons/b/bd/WGS\\_84\\_reference\\_frame.png](http://upload.wikimedia.org/wikipedia/commons/b/bd/WGS_84_reference_frame.png)



## 3 Extração de Dados

Neste capítulo são apresentados os detalhes sobre o processo de extração de dados do Twitter através da sua API. São também consideradas outras fontes de dados com o intuito de complementar os dados extraídos do Twitter e assim poder alcançar determinadas conclusões.

São apresentadas as tecnologias que foram utilizadas neste processo, sendo também descritos os extratores do Twitter, das outras fontes de dados e a forma como a recolha foi automatizada.

### 3.1 *Tecnologias Utilizadas*

Neste trabalho optou-se por usar a linguagem de programação Python, por ter uma sintaxe bastante intuitiva e por possuir uma grande variedade de *packages* disponibilizados, para facilitar o acesso à Twitter API. A facilidade de manipulação de dados foi também um fator preponderante aquando da escolha da mesma.

Neste estudo os resultados das consultas estiveram sempre intimamente ligados com a recolha das tendências (*trending topics*) nos locais escolhidos previamente.

Para a automatização do processo de extração, foram definidas tarefas diárias no sistema operativo, que foram executadas várias vezes por dia durante precisamente oitenta dias. Os dados resultantes das consultas que foram feitas à API foram retornados no formato JSON, que posteriormente serviu para alimentar diversos *scripts* em SQL (*Structured Query Language*), que efetuaram a inserção das tendências numa base de dados denominada de *TwitterTrends*.

Foi também utilizada uma linguagem que é usada no âmbito da *Semantic Web*, a linguagem SPARQL (*SPARQL Protocol and RDF Query Language*) encapsulada em código Python, com o

objetivo de extrair dados da Internet. Esses dados extraídos referem-se à lista de países, cidades, desportistas e alguns eventos, para depois poderem ser comparados com algumas das tendências extraídas.

### 3.1.1 Python

Para poder escrever o código Python num ambiente controlado e amigável, procedeu-se à instalação do interpretador *IPython*. Este funciona como uma *shell* que permite a utilização de várias linguagens de programação estando, no entanto, otimizado para Python. Um dado positivo sobre o *IPython* é que permite manter um histórico da linha de comandos.

No processo de implementação do código e no decorrer das tarefas que surgiram, foram instalados vários *packages* no *IPython* (ver Tabela 10), que permitiram a comunicação com a Twitter API, a escrita/leitura de JSON, a interação com endereços URL e o encapsulamento de outras linguagens.

Tabela 10 – *Packages* instalados no *IPython*

<i>Package</i>	<i>Função</i>
<i>pip</i>	<i>Instalação e gestão de packages do Python</i>
<i>python-twitter</i>	<i>Comunicação com a Twitter API</i>
<i>simplejson</i>	<i>Leitura/escrita de JSON</i>
<i>httplib2</i>	<i>Comunicação com o protocolo http</i>
<i>urllib2</i>	<i>Leitura de URLs</i>
<i>SPARQLWrapper</i>	<i>Simula um ambiente para usar queries SPARQL</i>

### 3.1.2 SQL

Neste trabalho foi usado o *SQL Management Studio* que é uma aplicação usada na configuração, gestão e administração de bases de dados usando a linguagem SQL.

Para este estudo, foi criada uma base de dados denominada de *TwitterTrends*, onde inicialmente foram criadas apenas duas tabelas (*locations* e *trends*). A tabela *locations* tem dois campos, *woeid* e *name*, relativos aos códigos e respetivos nomes dos locais em análise. A tabela *trends* armazena toda a informação relativa às tendências recolhidas. Nela surgem os seguintes campos: *trendsID*, *name*, *as\_of*, *time\_of\_day* e *woeid*. Estes campos correspondem respetivamente a um código numérico, à tendência propriamente dita, à data e hora da tendência num formato (yyyy-mm-dd hh:mm:ss), à altura do dia da recolha (manhã, tarde ou noite) e ao *woeid* de um determinado local.

Posteriormente foram criadas outras tabelas com dados oriundos da *DBpedia* como indicado na Tabela 11. Na secção 3.3 são apresentados mais detalhes sobre o processo de criação das mesmas.



Tabela 11 – Tabelas criadas posteriormente na base de dados

Nome da tabela	Campos da tabela
countries	country, country1, country2
european_capital_cities	capital_cities, capital_cities_desc
pt_cities	cities
players	player_name, type_of_player
festivals	festival_name, festival_name_desc

### 3.1.3 Semantic Web e SPARQL

A *Semantic Web* pretende melhorar a partilha de conteúdos para além dos limites estabelecidos por *sites* e aplicações [Semanticweb.org, 2014]. A *Semantic Web* promove a organização da informação e um acesso seletivo à mesma, fornecendo meios para distribuição e troca de ontologias e de metadados. Por seu lado as ontologias são modelos de dados que representam um conjunto de conceitos dentro de um domínio e os seus relacionamentos. É baseada no formato RDF (*Resource Description Framework*) que é um modelo uniformizado construído para a troca de dados através da Internet usando URI (*Uniform Resource Identifier*). O URI é uma sequência de caracteres utilizado para identificar um recurso na Internet. Na Figura 21 estão visíveis as várias camadas da *Semantic Web*. Note-se que o SPARQL é a linguagem usada para consultas de dados no formato RDF no âmbito da *Semantic Web*.

A *DBpedia* é um projeto que se insere na iniciativa LOD (*Linking Open Data*) e disponibiliza dados provenientes da Wikipedia e que podem ser consultados com recurso a *queries* SPARQL. Neste trabalho essas *queries* foram embutidas em código Python.

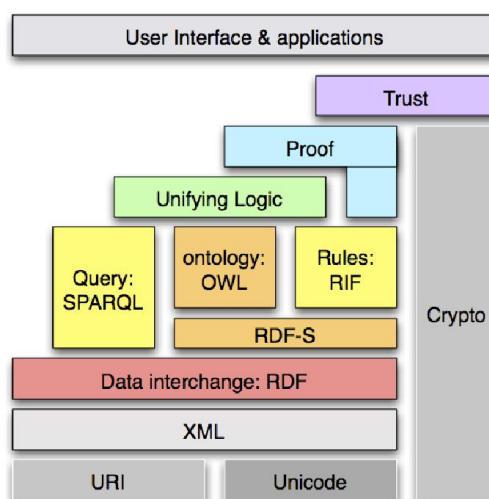


Figura 21 – As camadas da *Semantic Web*

Fonte: <http://www.w3.org/2006/Talks/1023-sb-W3CTechSemWeb/SemWebStack-tbl-2006a.png>

## **3.2 Extração das Tendências do Momento**

Neste trabalho optou-se pela escolha de seis países (locais) distintos, pertencentes aos continentes europeu e americano. Foram escolhidos dois países de língua portuguesa, dois países de língua espanhola e dois países de língua inglesa sendo esta escolha motivada por uma maior familiarização com a cultura e línguas desses países. Com esta decisão pretendeu-se observar semelhanças nas tendências de países com línguas idênticas mas de continentes distintos.

De referir também que a recolha das tendências efetuou-se da mesma forma mas à escala mundial. Esta opção surgiu da necessidade de analisar a importância que uma determinada tendência pode assumir em alguns países catapultando-a por vezes para um nível superior de destaque (tendência mundial).

As recolhas diárias das tendências no Twitter iniciaram-se no dia 3 de Maio de 2014 e estenderam-se até dia 21 de Julho de 2014 perfazendo um total de oitenta dias.

Numa primeira fase, que decorreu entre o dia 3 de Maio de 2014 e o dia 15 de Junho de 2014, extraíram-se as tendências apenas de Portugal, Espanha e Estados Unidos mas, face à necessidade, já referida, de comparar as tendências em diferentes países com a mesma língua, mas faladas em continentes diferentes, optou-se por alargar a recolha também a países como o Brasil, Argentina e o Reino Unido. Esse alargamento teve início no dia 16 de Junho e perdurou até ao dia 21 de Julho de 2014. A opção de considerar as tendências a nível mundial também decorreu neste último período.

Neste período temporal (oitenta dias), tiveram lugar acontecimentos bastante importantes essencialmente a nível desportivo mas também no que diz respeito a festivais de música, temas políticos, entre outros.

O mundial de futebol 2014 no Brasil superou durante mais de um mês qualquer tipo de eventos a nível mundial quer fossem desportivos ou de outra índole. Foi sem dúvida o assunto mais falado e debatido nos locais escolhidos para análise bem como em muitos outros locais.

Em Portugal, os festivais de verão Rock in Rio e NOS Alive tiveram também grande repercussão. Em Espanha por exemplo, a final da Liga dos Campeões jogada em Lisboa, entre o Real Madrid e Atlético de Madrid, foi um dos temas de interesse durante o mês de Maio, logo seguido da abdicação do Rei Juan Carlos e posterior coroação de Felipe VI durante o mês de Junho.

### **3.2.1 Tendências em Números**

Seguidamente surgem duas tabelas (Tabela 12 e 13) que refletem o número de tendências armazenadas durante o período de recolha de dados sendo apresentados os totais diários.

Tabela 12 – Número de tendências extraídas na 1ª fase (por dia)

País	10H	16H	22H	Total
PT	10	10	10	30
ES	10	10	10	30
USA	10	10	10	30
<b>Total</b>	<b>30</b>	<b>30</b>	<b>30</b>	<b>90</b>

As 90 tendências multiplicadas pelos 44 dias que durou a 1ª fase da recolha (3 de Maio a 15 de Junho) perfazem um total de 3.960 tendências.

Tabela 13 – Número de tendências extraídas na 2ª fase (por dia)

País	10H	16H	22H	Total
PT	10	10	10	30
ES	10	10	10	30
USA	10	10	10	30
BRA	10	10	10	30
ARG	10	10	10	30
UK	10	10	10	30
WORLD	10	10	10	30
<b>Total</b>	<b>70</b>	<b>70</b>	<b>70</b>	<b>210</b>

As 210 tendências multiplicadas pelos 36 dias que durou a 2ª fase da recolha (16 de Junho a 21 de Julho) perfazem um total de 7.560 tendências.

Estes dois totais parciais somados, formam um total final de 11.520 tendências, sendo esta a amostra usada para este estudo.

### 3.2.2 Tarefas Criadas no Sistema Operativo

Para a extração automática das tendências, existem três tarefas criadas no programador de tarefas do sistema operativo. Essas tarefas são executadas às 10h, 16h e por fim 22h de cada dia. Cada tarefa executa um *batch file* que contém a execução de sete *scripts* em Python, pertencentes a cada um aos locais em estudo (Portugal, Espanha, Estados Unidos, Brasil, Argentina, Reino Unido e Mundial). Seguidamente está um exemplo de um desses *batch files* que são executados através das tarefas criadas.

```
@echo off
python "C:\Users\Tiago Martins\Desktop\trend_extraction_PT.py" >
"C:\Users\Tiago Martins\Desktop\Trends\trend_PT_10h_XXJun14.json"
python "C:\Users\Tiago Martins\Desktop\trend_extraction_ES.py" >
"C:\Users\Tiago Martins\Desktop\Trends\trend_ES_10h_XXJun14.json"
python "C:\Users\Tiago Martins\Desktop\trend_extraction_USA.py" >
"C:\Users\Tiago Martins\Desktop\Trends\trend_USA_10h_XXJun14.json"
python "C:\Users\Tiago Martins\Desktop\trend_extraction_BRA.py" >
"C:\Users\Tiago Martins\Desktop\Trends\trend_BRA_10h_XXJun14.json"
python "C:\Users\Tiago Martins\Desktop\trend_extraction_ARG.py" >
"C:\Users\Tiago Martins\Desktop\Trends\trend_ARG_10h_XXJun14.json"
python "C:\Users\Tiago Martins\Desktop\trend_extraction_UK.py" >
"C:\Users\Tiago Martins\Desktop\Trends\trend_UK_10h_XXJun14.json"
```

```
python "C:\Users\Tiago Martins\Desktop\trend_extraction_WORLD.py" >
"C:\Users\Tiago Martins\Desktop\Trends\trend_WORLD_10h_XXJun14.json"
```

Código 6 – Script executado pela tarefa “Trends\_10H”

As restantes tarefas (16h e 22h) são semelhantes e os *batch files* usados apenas diferem no nome do ficheiro JSON a criar.

### 3.2.3 Extratores das Tendências

Os ficheiros em código Python, foram criados para ser possibilitada a comunicação com a Twitter API e dessa forma realizar a extração das tendências num determinado momento definido pelo utilizador, através da execução das tarefas do sistema operativo descritas anteriormente.

Estão presentes no código os seguintes elementos:

- Importação de dois *packages* do *IPython*;
- Utilização das quatro credenciais *OAuth* descritas no capítulo anterior para comunicação com a Twitter API;
- Estabelecimento da ligação com a Twitter API;
- Atribuição do WOEID que se pretende analisar;
- Extração das dez tendências (*trends.place*) do local definido no WOEID do ponto anterior;
- Escrita do resultado no formato JSON.

#### *trend\_extraction\_PT.py*

```
import twitter
import json
CONSUMER_KEY = 'xxxxxxxxxx'
CONSUMER_SECRET = 'xxxxxxxxxx'
OAUTH_TOKEN = 'xxxxxxxxxx'
OAUTH_TOKEN_SECRET = 'xxxxxxxxxx'
auth = twitter.oauth.OAuth(OAUTH_TOKEN, OAUTH_TOKEN_SECRET, CONSUMER_KEY,
CONSUMER_SECRET)
twitter_api = twitter.Twitter(auth=auth)
PT_WOE_ID = 23424925
pt_trends = twitter_api.trends.place(_id=PT_WOE_ID)
print json.dumps(pt_trends, indent=1)
```

Código 7 – Extração das tendências em Portugal

Para os restantes locais (Espanha, Estados Unidos, Brasil, Argentina, Reino Unido e Mundial) apenas é alterado o código WOEID e o respetivo nome da variável.

### 3.2.4 Sistema Implementado

O sistema apresentado na Figura 22 explica detalhadamente todo o processo implementado durante a execução deste trabalho e que permitiu a extração das tendências do Twitter.

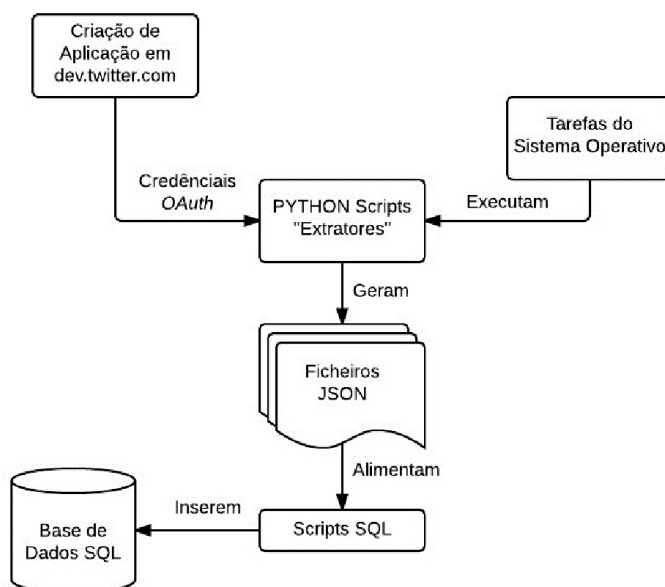


Figura 22 – Sistema de extração das tendências

Inicialmente procede-se à criação da aplicação em (<https://dev.twitter.com/apps>) para obter as credenciais *OAuth* e à criação das tarefas no sistema operativo e respetivos extratores escritos em Python. Esses extratores, quando executados, geram os ficheiros JSON com as tendências que depois servem para preencher os *scripts* SQL que inserem as tendências numa tabela da base de dados criada para esse efeito através de um simples *script* em SQL como exemplificado parcialmente no código demonstrado em seguida (Código 8).

```
INSERT INTO trends (trendsID, name, as_of, time_of_day, woeid)
VALUES
('11311', '#NiallAppreciationDay', '2014-07-21 10:00', 'Manhã', '23424925'),
('11312', 'Tiago Garrinhas', '2014-07-21 10:00', 'Manhã', '23424925'),
('11313', '#thevoiceportugal', '2014-07-21 10:00', 'Manhã', '23424925'),
('11314', 'Bianca', '2014-07-21 10:00', 'Manhã', '23424925'),
('11315', 'Alexandre', '2014-07-21 10:00', 'Manhã', '23424925'),
('11316', 'Rising Star', '2014-07-21 10:00', 'Manhã', '23424925'),
('11317', 'Algarve', '2014-07-21 10:00', 'Manhã', '23424925'),
('11318', 'The Voice', '2014-07-21 10:00', 'Manhã', '23424925'),
('11319', 'xDxD', '2014-07-21 10:00', 'Manhã', '23424925'),
('11320', '#CarregaBenfica', '2014-07-21 10:00', 'Manhã', '23424925'),...
```

Código 8 – *Script* de inserção das tendências na base de dados

### 3.3 Outras Fontes de Dados Consideradas

Neste trabalho, como já foi referido anteriormente, foram recolhidos dados do Twitter. No entanto, outras fontes de dados são também consideradas, de forma a complementar os dados existentes e também com o intuito de atingir determinadas conclusões. Por exemplo, se um termo aparece num determinado jornal, ou se um termo numa tendência corresponde a um local, a um evento ou a uma pessoa. Optou-se assim por efetuar também uma recolha automática de alguns jornais *online* diários durante um determinado período de tempo e ainda por proceder à recolha de dados provenientes da plataforma *DBpedia*.

#### 3.3.1 Jornais Diários

Foram escolhidos textos de alguns dos jornais diários dos países em estudo (ver Tabela 14). Estes jornais foram usados como forma de complementar a análise de dados proveniente do Twitter e verificar se algumas das tendências estariam relacionadas com tópicos noticiosos.

Tabela 14 – Jornais extraídos e respetivos países de origem

País de origem	Nome do jornal
<i>Portugal</i>	<i>Jornal de Notícias</i>
<i>Espanha</i>	<i>El Mundo</i>
<i>Estados Unidos</i>	<i>The New York Times</i>
<i>Brasil</i>	<i>Folha de S. Paulo</i>
<i>Argentina</i>	<i>La Nación</i>
<i>Reino Unido</i>	<i>The Guardian</i>

#### 3.3.2 DBpedia

A *DBpedia* é uma plataforma construída com o intuito de extrair conteúdos/informação estruturada existente na *Wikipedia*. A informação estruturada é posteriormente disponibilizada na Internet sendo possível efetuar consultas às propriedades e relações dos recursos existentes na *Wikipedia*, sendo também possível realizar consultas a outros conjuntos de dados existentes na Internet e ligá-los com dados existentes na *Wikipedia* [Bizer et al., 2009].

A versão inglesa tem uma base de conhecimento que comporta quatro milhões de itens, dos quais 3,22 milhões estão classificados em modelos de dados consistentes, sendo 832.000 pessoas, 639.000 locais e também 209.000 organizações [Wiki.dbpedia.org, 2014].

Neste trabalho a *DBpedia* foi utilizada para efetuar a extração de países, capitais europeias e ainda principais cidades portuguesas usando diversas *queries* SPARQL embutidas em código Python. Também foi possível extrair uma lista completa com desportistas de várias modalidades (futebol, basquetebol e ténis). Por último foram extraídos alguns eventos (festivais de música) de Portugal.

Torna-se assim importante referir que todo este processo de extração de dados provenientes da *DBpedia* revela-se bastante útil quando se pretende complementar os dados já existentes na base de dados sobre as tendências do Twitter. Neste estudo são apresentados apenas pequenos exemplos, que serviram para demonstrar certos aspetos relacionados com a utilização completamente escalável da *Semantic Web* permitindo aos utilizadores inúmeras possibilidades de extração dos mais variados dados existentes na Internet.

Na Figura 23, é exemplificado o sistema de extração de dados da *DBpedia* através do uso de *queries* SPARQL encapsuladas em código Python.

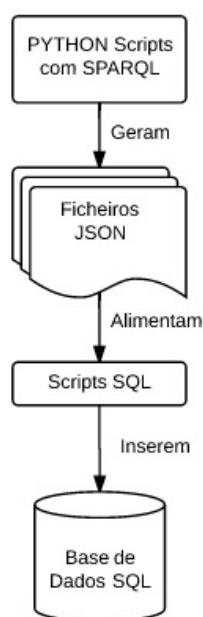


Figura 23 – Sistema de extração de dados com *queries* SPARQL

Por exemplo, para extrair a lista de países da *DBpedia*, o *script* em Python com SPARQL gera um ficheiro JSON com os países, que serve para preencher o *script* SQL que seguidamente os insere na tabela *countries*.

Esta lógica repete-se para as tabelas de capitais europeias (*european\_capital\_cities*), cidades portuguesas (*pt\_cities*), jogadores de futebol, basquetebol e ténis (*players*) e ainda para a tabela que contém os festivais de música (*festivals*) bastando para isso alterar a *query* SPARQL a usar (ver secção 3.3.4).

### 3.3.3 Extração de Dados dos Jornais Diários

Durante o mês de Julho adicionou-se um novo elemento ao estudo das tendências. Uma extração diária no fim de cada dia, dos jornais de maior expressão em cada um dos países escolhidos para a recolha de dados.

Para a extração automática dos jornais diários nos seis países de referência, foi criada uma tarefa no programador de tarefas do sistema operativo. Essa tarefa seria executada às 22h01m logo depois de ser executada a tarefa das 22h da extração das tendências.

```
@echo off
python "C:\Users\Tiago Martins\Desktop\Jornais\PT_Jornal.py"
python "C:\Users\Tiago Martins\Desktop\Jornais\ES_Jornal.py"
python "C:\Users\Tiago Martins\Desktop\Jornais\USA_Jornal.py"
python "C:\Users\Tiago Martins\Desktop\Jornais\BRA_Jornal.py"
python "C:\Users\Tiago Martins\Desktop\Jornais\ARG_Jornal.py"
python "C:\Users\Tiago Martins\Desktop\Jornais\UK_Jornal.py"
```

Código 9 – Script executado pela tarefa “jornais”

Estão presentes no código os seguintes elementos:

- Importação do *package* do *IPython*;
- Comunicação com a página *web* pretendida;
- Leitura do conteúdo;
- Escrita do conteúdo (página *web* principal).

```
import urllib2
page = urllib2.urlopen('http://www.jn.pt')
page_content = page.read()
with open('PT_jn.html', 'w') as fid:
    fid.write(page_content)
```

Código 10 – Extração diária do *Jornal de Notícias* (Portugal)

```
import urllib2
page = urllib2.urlopen('http://www.elmundo.es')
page_content = page.read()
with open('ES_elmundo.html', 'w') as fid:
    fid.write(page_content)
```

Código 11 – Extração diária do *El Mundo* (Espanha)

```
import urllib2
page = urllib2.urlopen('http://www.nytimes.com')
page_content = page.read()
with open('USA_nytimes.html', 'w') as fid:
    fid.write(page_content)
```

Código 12 – Extração diária do *The New York Times* (Estados Unidos)



```
import urllib2
page = urllib2.urlopen('http://www.folha.uol.com.br')
page_content = page.read()
with open('BRA_folhasaopaulo.html', 'w') as fid:
    fid.write(page_content)
```

Código 13 – Extração diária do *Folha de São Paulo* (Brasil)

```
import urllib2
page = urllib2.urlopen('http://www.lanacion.com.ar')
page_content = page.read()
with open('ARG_lanacion.html', 'w') as fid:
    fid.write(page_content)
```

Código 14 – Extração diária do *La Nación* (Argentina)

```
import urllib2
page = urllib2.urlopen('http://www.theguardian.com/uk')
page_content = page.read()
with open('UK_theguardian.html', 'w') as fid:
    fid.write(page_content)
```

Código 15 – Extração diária do *The Guardian* (Reino Unido)

### 3.3.4 Extração de Dados da DBpedia

No processo de extração de dados da *DBpedia* foram usadas algumas *queries* SPARQL embutidas em código Python, sendo o SPARQL a linguagem usada para efetuar consultas sobre um determinado domínio no âmbito da *Semantic Web*.

Estão presentes no código os seguintes elementos:

- Importação de dois *packages* do *IPython*;
- Encapsulamento da *query* SPARQL;
- Escrita do resultado no formato JSON.

```
import json
from SPARQLWrapper import SPARQLWrapper, JSON
sparql = SPARQLWrapper("http://dbpedia.org/sparql")
sparql.setQuery("""
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
select ?country where { ?country <http://purl.org/dc/terms/subject>
<http://dbpedia.org/resource/Category:Member_states_of_the_United_Nations>;
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://schema.org/Country> } order by ?country
""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
for result in results["results"]["bindings"]:
```

```
print result["country"]["value"]
```

Código 16 – Extração da lista de países membros das Nações Unidas da *DBpedia*

O resultado da execução do código 16 deu origem a uma lista de países que foram adicionados à base de dados através da execução de um *script* em SQL parcialmente apresentado no fim desta secção, que os inseriu na tabela *countries* da base de dados.

```
import json
from SPARQLWrapper import SPARQLWrapper, JSON
sparql = SPARQLWrapper("http://dbpedia.org/sparql")
sparql.setQuery("""PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT * WHERE {
?subject rdfs:type <http://dbpedia.org/class/yago/CapitalsInEurope>.
?subject rdfs:label ?label.
?subject rdfs:comment ?abstract.
FILTER (lang(?label) = "pt" && lang(?abstract) = "pt")
}""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
for result in results["results"]["bindings"]:
    print result["label"]["value"]
```

Código 17 – Extração da lista de capitais europeias da *DBpedia*

O código 17 deu origem a uma lista com as capitais europeias e respetivas descrições extraídas da *DBpedia*. Seguidamente foram adicionadas à base de dados através da execução de um *script* em SQL também apresentado parcialmente no fim desta secção e que inseriu os dados na tabela *european\_capital\_cities* da base de dados.

Também foi usada uma consulta que extraiu uma lista com as cidades portuguesas presentes na *DBpedia*, sendo posteriormente adicionadas à base de dados através da execução de um *script* em SQL apresentado parcialmente no fim desta secção, que inseriu os dados na tabela *pt\_cities* da base de dados.

Foi usada outra consulta, esta com a particularidade de gerar uma lista com jogadores de futebol presentes no mundial de futebol de 2010, uma vez que ainda não existiam dados referentes ao mundial de 2014.

A Tabela 15 apresenta algumas das classes utilizadas para a extração dos dados mencionados.

Tabela 15 – Classes que suportaram a extração de dados da *DBpedia*

Recurso da <i>DBpedia</i>	Objetivo
<a href="http://dbpedia.org/class/yago/CapitalsInEurope">http://dbpedia.org/class/yago/CapitalsInEurope</a>	Extração de dados sobre as capitais europeias
<a href="http://dbpedia.org/class/yago/CitiesInPortugal">http://dbpedia.org/class/yago/CitiesInPortugal</a>	Extração de dados sobre as cidades de Portugal
<a href="http://dbpedia.org/class/yago/2010FIFAWorldCupPlayers">http://dbpedia.org/class/yago/2010FIFAWorldCupPlayers</a>	Extração de dados sobre os jogadores do Mundial 2010

Apesar de ser já bastante completa, a lista de jogadores de futebol foi atualizada com alguns jogadores em falta através de uma simples *query* SPARQL usando a classe *SoccerPlayer* (código 18).

```

import json
from SPARQLWrapper import SPARQLWrapper, JSON
sparql = SPARQLWrapper("http://dbpedia.org/sparql")
sparql.setQuery("""PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT * WHERE {
?subject rdf:type <http://dbpedia.org/ontology/SoccerPlayer>.
?subject rdfs:label ?label.
?subject rdfs:comment ?abstract.
FILTER (lang(?label) = "pt" && lang(?abstract) = "pt")
}
""")
sparql.setReturnFormat(JSON)
results = sparql.query().convert()
for result in results["results"]["bindings"]:
    print result["label"]["value"]

```

Código 18 – Extração de dados da *DBpedia* (jogadores de futebol)

Também foram extraídos atletas de outros desportos como por exemplo o ténis e basquetebol. Nestes exemplos são usadas as classes *TennisPlayer* e *BasketballPlayer*, subclasses da classe *Athlete* (<http://dbpedia.org/ontology/Athlete>), que pode ser utilizada para retornar dados de todos os atletas presentes na *DBpedia*.

A Tabela 16 apresenta as classes usadas para extrair dados da *DBpedia* relativamente a desportistas.

Tabela 16 – Classes usadas para extrair dados sobre desportistas

Recurso da <i>DBpedia</i>	Objetivo
<a href="http://dbpedia.org/ontology/SoccerPlayer">http://dbpedia.org/ontology/SoccerPlayer</a>	Extração de dados sobre futebolistas
<a href="http://dbpedia.org/ontology/TennisPlayer">http://dbpedia.org/ontology/TennisPlayer</a>	Extração de dados sobre tenistas
<a href="http://dbpedia.org/ontology/BasketballPlayer">http://dbpedia.org/ontology/BasketballPlayer</a>	Extração de dados sobre basquetebolistas

Tendo em conta a quantidade de dados existente na *Wikipedia*, torna-se necessário definir bem o domínio dos dados que se pretendem extrair de forma a evitar a sua recolha excessiva e obtendo apenas o estritamente necessário.

Outra extração efetuada debruçou-se na recolha da lista de festivais de música de Portugal com utilização da ontologia YAGO. Esses recursos encontram-se descritos na Tabela 17.

Tabela 17 – Recursos usados para extrair dados sobre festivais portugueses da *DBpedia*

Recurso da <i>DBpedia</i>	Objetivo
<a href="http://dbpedia.org/class/yago/RockFestivalsInPortugal">http://dbpedia.org/class/yago/RockFestivalsInPortugal</a>	Extração de dados sobre festivais de rock
<a href="http://dbpedia.org/class/yago/MusicFestivalsInPortugal">http://dbpedia.org/class/yago/MusicFestivalsInPortugal</a>	Extração de dados sobre festivais de música genéricos

Seguidamente são parcialmente apresentados os *scripts* de inserção usados para inserir na base de dados *TwitterTrends*, os dados provenientes da *DBpedia* nas tabelas *countries*, *european\_capital\_cities*, *pt\_cities*, *players* e *festivals*.

Na tabela *countries* foram adicionados posteriormente mais dois campos (*country1* e *country2*) face à necessidade de ter alguns países apresentados noutras línguas (português e espanhol) respetivamente, para assim poder efetuar um cruzamento de dados mais completo.

```
INSERT INTO countries (country)
VALUES
('Afghanistan'),
('Albania'),
('Algeria'),
('Andorra'),
('Angola'),
('Antigua and Barbuda'),
('Argentina'),
('Armenia'),
('Australia'),
('Austria'),...
```

Código 19 – *Script* de inserção dos países na base de dados

```
INSERT INTO european_capital_cities (capital_cities, capital_cities_desc)
VALUES
('Amesterdão', 'Amesterdão é a capital dos Países Baixos.'),
('Ancara', 'Ancara é a capital da República da Turquia.'),
('Andorra-a-Velha', 'Andorra-a-Velha é a capital do co-principado de
Andorra.'),
('Atenas', 'Atenas é a capital da Grécia.'),
('Baku', 'Baku é a capital do Azerbaijão.'),
('Belfast', 'Belfast é a capital da Irlanda do Norte.'),
('Belgrado', 'Belgrado é a capital da Sérvia.'),
('Berlim', 'Berlim é a capital da Alemanha.'),
('Berna', 'Berna é a capital da Suíça.'),
('Bratislava', 'Bratislava é a capital da Eslováquia.'),...
```

Código 20 – *Script* de inserção das capitais europeias na base de dados

```
INSERT INTO pt_cities (cities) VALUES
('Abrantes'),
('Aguilva-Cacém'),
('Águeda'),
('Albufeira'),
('Alcobaça'),
('Almada'),
('Almeirim'),
('Alverca do Ribatejo'),
('Amadora'),
('Angra do Heroísmo'),...
```

Código 21 – *Script* de inserção das cidades portuguesas na base de dados

```
INSERT INTO players (player_name,type_of_player) VALUES ('A. C.
Green', 'Basquetebolista'),
('Aaron Brooks', 'Basquetebolista'),
('Aaron Clapham', 'Futebolista'),
('Aaron Lennon', 'Futebolista'),
('Aaron Mokoena', 'Futebolista'),
('Abdelkader Ghezzal', 'Futebolista'),
('Abdelkader Laïfaoui', 'Futebolista'),
('Abdul Rahim Ayew', 'Futebolista'),
('Abigail Spears', 'Tenista'),
('Abou Diaby', 'Futebolista'),...
```

Código 22 – *Script* de inserção de desportistas de várias modalidades na base de dados

```

INSERT INTO festivals (festival_name,festival_name_desc) VALUES
 ('Rock in Rio','Rock in Rio é um festival de música realizado pela
 primeira vez em 1985, sendo, desde sua criação, reconhecidamente, o maior
 festival musical do mundo latino. '),
 ('Super Bock Super Rock','O Super Bock Super Rock é um festival de música
 de Verão realizado anualmente no Meco, em Sesimbra. '),...

```

Código 23 – *Script* de inserção de eventos de música na base de dados

### 3.3.5 Consultas SQL com Dados Oriundos da DBpedia

Os dados extraídos da *DBpedia* depois de gravados nas respetivas tabelas da base de dados, podem ser alvo de consultas por forma a poder associar as tendências existentes na tabela *trends* com os dados existentes nas outras tabelas (*countries*, *european\_capital\_cities*, *pt\_cities*, *players* e *festivals*).

Seguidamente são apresentadas algumas consultas e os respetivos resultados dessas consultas na forma de tabelas com alguns dados. É apresentado detalhadamente um primeiro exemplo seguindo os restantes exemplos a mesma lógica.

Se o objetivo pretendido for a visualização das tendências relacionadas com países e que tem correspondência direta na tabela de países, pode ser efetuada a consulta indicada no código 24.

```

(select * from trends, countries, locations
 where trends.name = countries.country
 and locations.woeid = trends.woeid)

union all

(select * from trends, countries, locations
 where trends.name = countries.country1
 and locations.woeid = trends.woeid)

union all

(select * from trends, countries, locations
 where trends.name = countries.country2
 and locations.woeid = trends.woeid)
order by as_of

```

Código 24 – Consulta SQL com a combinação de tendências com países

O resultado parcial obtido após a execução da consulta é apresentado na Tabela 18. Os campos (*trendsID*, *name*, *as\_of*, *time\_of\_day*, *woeid*) pertencem à tabela *trends* e apresentam dados que já foram apresentados na secção 3.1.2. À direita do campo *woeid* surge o campo *name* que representa o local correspondente a esse código *woeid*. Por último os restantes campos (*country*, *country1* e *country2*) são dados extraídos da *DBpedia* em inglês, português e espanhol.

Tabela 18 – Combinação de dados das tabelas *trends* e *countries*

<i>trendsID</i>	<i>name</i>	<i>as_of</i>	<i>time_of_day</i>	<i>woeid</i>	<i>name</i>	<i>country</i>	<i>country1</i>	<i>country2</i>
19	Ucrania	2014-05-03 10:00:00.000	Manhã	23424950	Spain	Ukraine	Ucrânia	Ucrania
199	España	2014-05-05 10:00:00.000	Manhã	23424950	Spain	Spain	Espanha	España
339	Holanda	2014-05-06 22:00:00.000	Noite	23424925	Portugal	Kingdom of the Netherlands	Holanda	Holanda
3630	Brazil	2014-06-12 10:00:00.000	Manhã	23424977	United States	Brazil	Brasil	Brasil
4033	Alemanha	2014-06-16 16:00:00.000	Tarde	23424925	Portugal	Germany	Alemanha	Alemania

Seguidamente são apresentadas algumas consultas e as respetivas tabelas com os dados parcialmente obtidos.

```
select * from trends, european_capital_cities
where trends.name = capital_cities
order by as_of
```

Código 25 – Consulta SQL com a combinação de tendências com cidades capitais europeias

Tabela 19 – Combinação de dados das tabelas *trends* e *european\_capital\_cities*

<i>trendsID</i>	<i>name</i>	<i>as_of</i>	<i>time_of_day</i>	<i>woeid</i>	<i>capital_cities</i>	<i>capital_cities_desc</i>
40	Lisboa	2014-05-03 16:00:00.000	Tarde	23424925	Lisboa	Lisboa é a capital de Portugal.
1399	Roma	2014-05-18 16:00:00.000	Tarde	23424950	Roma	Roma é a capital da Itália.
4188	Madrid	2014-06-17 10:00:00.000	Manhã	23424950	Madrid	Madrid é a capital da Espanha.
8459	Londres	2014-07-07 16:00:00.000	Tarde	23424950	Londres	Londres é a capital da Inglaterra e do Reino Unido.

```
select * from trends, pt_cities
where trends.name = cities
order by as_of
```

Código 26 – Consulta SQL com a combinação de tendências com cidades portuguesas

Tabela 20 – Combinação de dados das tabelas *trends* e *pt\_cities*

<i>trendsID</i>	<i>name</i>	<i>as_of</i>	<i>time_of_day</i>	<i>woeid</i>	<i>cities</i>
40	Lisboa	2014-05-03 16:00:00.000	Tarde	23424925	Lisboa
2140	Valongo	2014-05-26 22:00:00.000	Noite	23424925	Valongo
5577	Porto	2014-06-23 22:00:00.000	Noite	23424925	Porto
5929	Seixal	2014-06-25 16:00:00.000	Tarde	23424925	Seixal

```
select * from trends, players, locations
where trends.name = player_name
and locations.woeid = trends.woeid
order by as_of
```

Código 27 – Consulta SQL com a combinação de tendências com desportistas

Tabela 21 – Combinação de dados das tabelas *trends* e *players*

<i>trendsID</i>	<i>name</i>	<i>as_of</i>	<i>time_of_day</i>	<i>woeid</i>	<i>name</i>	<i>player_name</i>	<i>type_of_player</i>
180	Tony Parker	2014-05-04 22:00:00.000	Noite	23424977	United States	Tony Parker	Basquetebolista
1895	David Luiz	2014-05-24 10:00:00.000	Manhã	23424925	Portugal	David Luiz	Futebolista
4068	Cristiano Ronaldo	2014-06-16 16:00:00.000	Tarde	23424768	Brasil	Cristiano Ronaldo	Futebolista
7208	Maria Sharapova	2014-07-01 16:00:00.000	Tarde	23424977	United States	Maria Sharapova	Tenista

```
select * from trends, festivals
where trends.name = festival_name
order by as_of
```

Código 28 – Consulta SQL com a combinação de tendências com festivais de música

Tabela 22 – Combinação de dados das tabelas *trends* e *festivals*

<i>trendsID</i>	<i>name</i>	<i>as_of</i>	<i>time_of_day</i>	<i>woeid</i>	<i>festival_name</i>	<i>festival_name_desc</i>
2619	Rock in Rio	2014-06-01 10:00:00.000	Manhã	23424925	Rock in Rio	Rock in Rio é um festival de música realizado pela primeira vez em 1985, sendo, desde sua criação, reconhecidamente, o maior festival musical do mundo latino.
3635	Optimus Alive	2014-06-12 16:00:00.000	Tarde	23424925	Optimus Alive	Optimus Alive é um festival de música anual realizado no Passeio Marítimo de Algés, em Oeiras, Portugal.
10617	Super Bock Super Rock	2014-07-17 22:00:00.000	Noite	23424925	Super Bock Super Rock	O Super Bock Super Rock é um festival de música de Verão realizado anualmente no Meco, em Sesimbra.

Estas tabelas apresentam alguns resultados que constituem apenas uma porção dos dados que eventualmente podem ser combinados servindo apenas para demonstrar as possibilidades de combinação de dados entre as diversas tabelas existentes.





## 4 Análise de Dados

Neste quarto capítulo, depois de apresentado o funcionamento, as características, a API do Twitter e de explorado o processo da extração de dados, apresenta-se a análise aos dados extraídos através da Twitter API. São analisadas as tendências tendo em consideração uma série de parâmetros definidos previamente e são também analisadas as tendências que de igual modo surgiram noutras fontes de dados.

As tendências extraídas e guardadas na base de dados são posteriormente trabalhadas e exploradas através do uso de tabelas dinâmicas (*pivot tables*) com ligação direta à base de dados permitindo o manuseamento de grandes volumes de dados e facilitando o processo de análise de dados.

### 4.1 Análise das Tendências

Como já foi referido anteriormente no terceiro capítulo, foram extraídas as tendências do Twitter durante um período de tempo compreendido entre os meses de Maio e Julho, mais propriamente entre os dias 3 de Maio e 21 de Julho de 2014. Durante esse período extraíram-se as tendências de seis países e ainda as tendências mundiais. Numa fase inicial o estudo recaiu sobre três locais sendo posteriormente alargados para sete os locais a considerar no processo de extração. Definiram-se três períodos diários (10h, 16h, 22h) em que uma tarefa do sistema operativo despoletava um *script* que extraía as tendências da Twitter API. Essa frequência de recolha das tendências foi escolhida por representar um compromisso entre os recursos disponíveis para o trabalho e assim poder ter mais do que uma recolha diária e poder verificar se ao longo do dia haveriam grandes alterações nas tendências.

A Figura 24 apresenta o total de tendências extraídas neste estudo e a sua distribuição quer por locais quer por período do dia. Devido ao facto de Portugal, Espanha e Estados Unidos possuírem um tempo de extração mais alargado, o número de tendências de cada um destes

locais (2.400) é superior ao de Brasil, Argentina, Reino Unido e Mundial (1.080). Esse aspecto é tido em conta ao longo dos cálculos efetuados na presente análise.

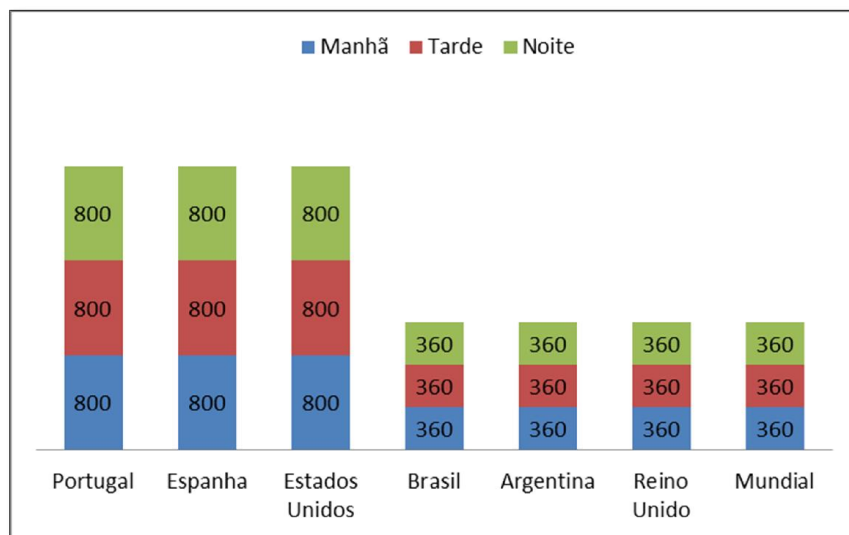


Figura 24 – Total de tendências extraídas

#### 4.1.1 Tendências Diferentes

A Figura 25 apresenta o total de tendências extraídas e o número de tendências distintas por local de extração, servindo essencialmente para concluir que as tendências podem repetir-se ao longo do tempo dependendo da força e importância dessas mesmas tendências. De referir um dado curioso que apresenta Portugal como o local que neste estudo verifica uma menor dinâmica na mudança de tendências (apenas 743 diferentes em 2.400 possíveis), o que leva a concluir que as tendências em Portugal tem um tempo de vida mais elevado do que nos outros locais em análise, possivelmente por se tratar de um país com menos utilizadores.

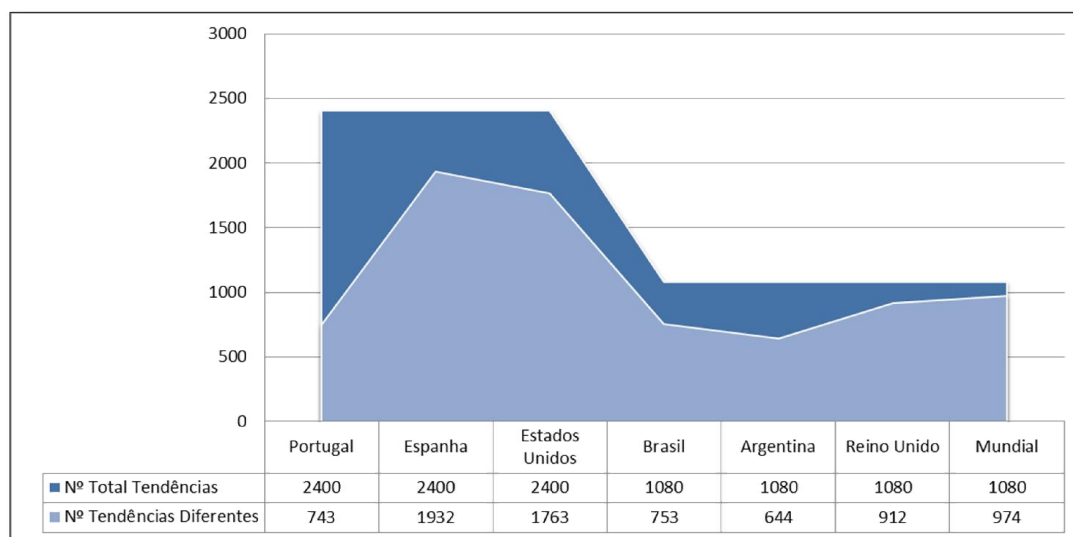


Figura 25 – Comparação de totais extraídos e totais de tendências diferentes por local

A Figura 25 apresenta o número de tendências diferentes por local, resultando a sua soma total em 7.721. Este número é obtido se a análise for feita individualmente por local, mas se for realizada uma análise global tendo em conta os sete locais indiferenciadamente, o número de tendências diferentes desce para 6.637 o que prova que essas tendências podem ser comuns a mais do que um local.

Por exemplo em 4 de Julho de 2014 a tendência “#BRAvsCOL” relativa ao jogo de futebol do mundial entre o Brasil e a Colômbia, foi tendência nos sete locais em análise. Neste caso se a análise for efetuada por local este resultado repete-se sete vezes mas se for tido em conta o global das tendências apenas é contabilizada uma única vez.

#### 4.1.2 Repetição de Tendências

A repetição de tendências é um facto incontornável pelo que surge a necessidade de analisar algumas dessas situações. Na secção anterior (4.1.1) fez-se referência à possibilidade de existirem tendências que por um determinado motivo, se repetem ao longo do tempo. Nos dados extraídos são inúmeros os casos em que esta situação se verifica, no entanto e face ao número elevado de repetições optou-se por apresentar na Tabela 23 apenas aquelas tendências que se repetiram em cinco ou mais ocasiões, sendo também apresentados alguns exemplos. Na secção de anexos apresentam-se na totalidade algumas tabelas com o número de tendências que surgiram em cinco ou mais ocasiões.

Tabela 23 – Número de tendências repetidas (cinco ou mais repetições)

<i>Local</i>	<i>Nº de Tendências</i>	<i>Exemplos</i>
<i>PT</i>	<i>97</i>	<i>"Lisboa", "Inês"</i>
<i>ES</i>	<i>16</i>	<i>"España", "Madrid"</i>
<i>USA</i>	<i>19</i>	<i>"Netflix", "Waffle House"</i>
<i>BRA</i>	<i>7</i>	<i>"Gabriel", "Lucas"</i>
<i>ARG</i>	<i>8</i>	<i>"Agus", "Disney"</i>
<i>UK</i>	<i>3</i>	<i>"Jeremy Kyle", "#tracerequest"</i>
<i>WORLD</i>	<i>2</i>	<i>"Happy Thanksgiving", "XXXXX"</i>

Face ao exposto no ponto 4.1.1, Portugal surge nesta tabela como o local que apresenta o maior número de tendências repetidas visto que já tinha sido demonstrado que também era o local que tinha um menor número de tendências diferentes.

#### 4.1.3 Número Médio de Caracteres

Outro dos pontos em análise é o número médio de caracteres das tendências extraídas (Figura 26), sendo importante referir que para este efeito os espaços também foram contabilizados como caracteres. Verifica-se que os termos em Portugal possuem uma menor dimensão face aos de outros locais (inferior a 11 caracteres por tendência), sendo os termos a nível mundial aqueles que têm em média maior dimensão motivados sobretudo por tendências de países que não fazem parte do estudo. Como exemplos de tendências com

elevado número de caracteres surgem “RengimSariLacivert HayatimFENERBAHÇE”, “Pemerintah Tiongkok Larang Warganya Berpuasa” ou ainda “KoyuncaKupaCikaranTakim HazirmisinBuSezon”.

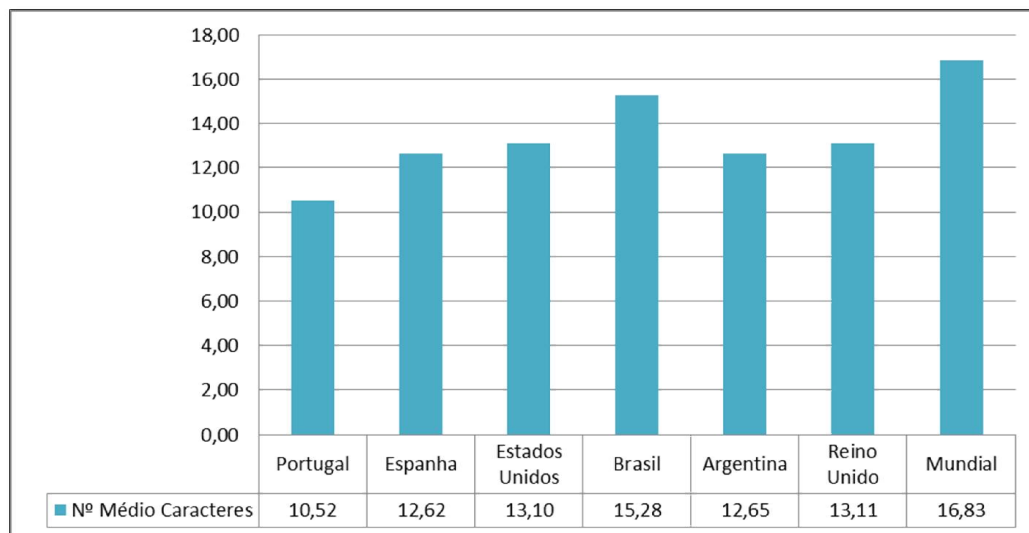


Figura 26 – Número médio de caracteres por tópico

#### 4.1.4 Número Médio de Palavras

Na Figura 27 surge um gráfico que apresenta o número médio de palavras presentes nas tendências. Neste caso são contabilizadas as palavras que se encontram separadas por espaços. Verifica-se novamente que os valores mais elevados registam-se nas tendências a nível mundial e que nesta análise a Argentina regista o menor número médio de palavras.

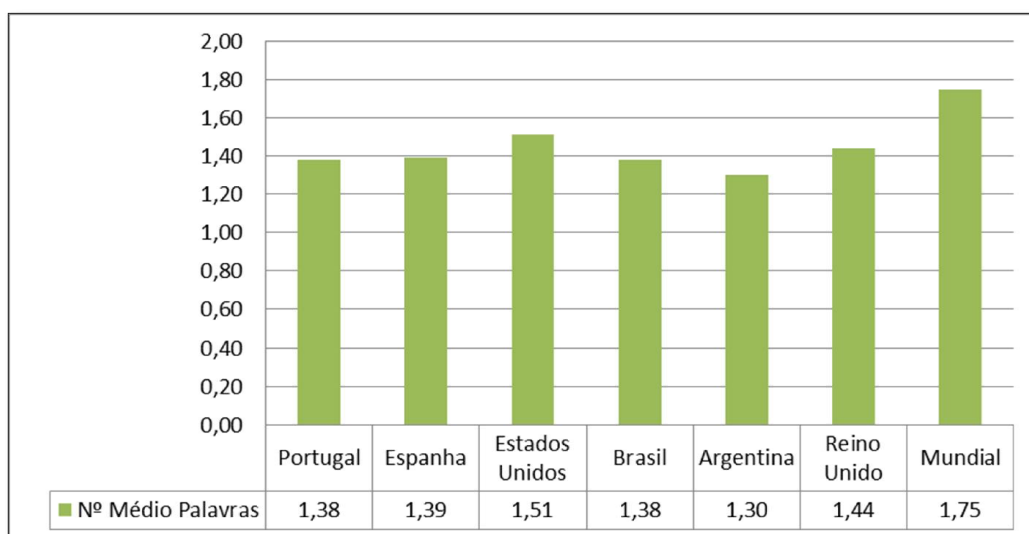


Figura 27 – Número médio de palavras por tópico

#### 4.1.5 Número de Palavras por Tópico

Por exemplo o termo “30 Things Only Antisocial People Understand” tendência registada a nível mundial em 24 de Junho de 2014 é uma das 34 tendências que apresenta seis palavras distintas. De realçar o facto de que a maioria, cerca de aproximadamente 69% das tendências são compostas apenas por uma palavra e que o número de tendências diminui à medida que aumenta o número de palavras que as compõem. Por exemplo, a tendência “Queens Of The Stone Age” apresenta cinco palavras distintas enquanto a tendência “#Aos25AnosQuero” conta apenas como uma palavra.

A Figura 28 apresenta a distribuição das tendências extraídas consoante o número de palavras que as compõem, mostrando que regra geral as tendências do Twitter são apenas compostas por uma palavra.

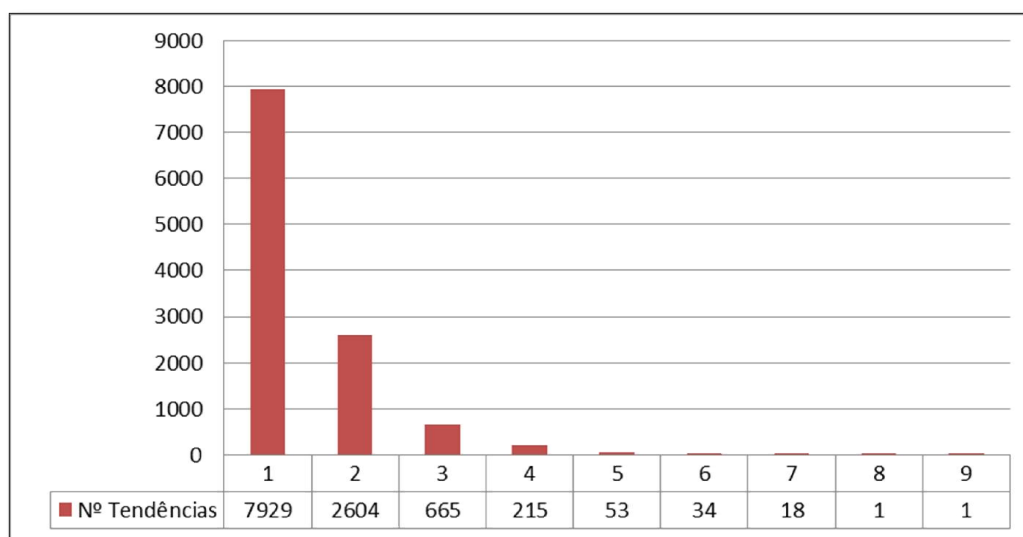


Figura 28 – Número de palavras por tópico

#### 4.1.6 Análise de Hashtags

Como já foi apresentado no segundo capítulo, as *hashtags* ou marcadores são expressões iniciadas pelo símbolo # e que servem essencialmente como forma de categorização de mensagens permitindo o seu agrupamento no Twitter.

Do total de tendências extraídas e analisadas, cerca de 44,37% (5.111 em 11.520) começam pelo símbolo #, sendo notório o grau de adoção das *hashtags* por parte dos utilizadores do Twitter. A Figura 29 apresenta o número de *hashtags* existentes face ao total de tendências extraído para cada um dos locais.

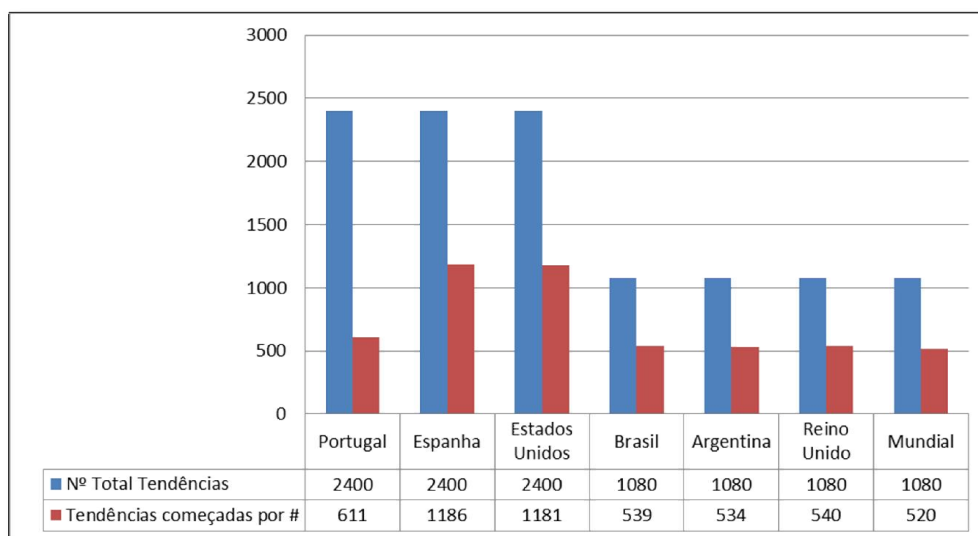


Figura 29 – Comparação de totais extraídos e totais de *hashtags*

Por seu lado, a Figura 30 apresenta um dado curioso relativamente às percentagens de uso de *hashtags* nos vários locais em estudo. Portugal surge com uma percentagem bastante inferior face a todos os outros locais. Isto poderá estar relacionado com alguma resistência ou pouca familiarização que ainda se verifica em Portugal no uso das *hashtags* por parte dos utilizadores do Twitter. Associado a este dado está o facto de o Twitter ainda não registar o mesmo grau de popularidade em Portugal do que nos outros locais em estudo como já demonstrado no início do segundo capítulo deste trabalho. Nos outros locais escolhidos para análise, as *hashtags* surgem como tendência em aproximadamente 50% dos casos.

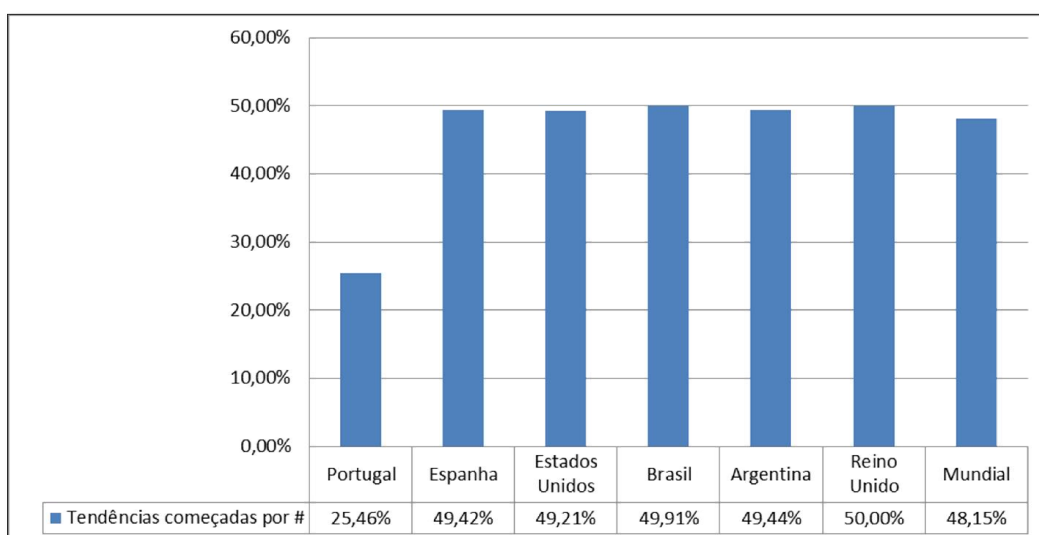


Figura 30 – Percentagem de *hashtags* face ao total de tendências

As tendências extraídas apresentam um dado importante relativamente ao número de palavras existentes em tópicos começados pelo símbolo #. Apenas uma palavra compõe este tipo de registos, ou seja, as *hashtags* extraídas são na verdade uma única palavra. Exemplos de *hashtags*: “#TwitterOff”, “#DescribeYourFriendshipInAPicture” ou “#Aos25AnosQuero”. O

mundial de futebol também se caracterizou por apresentar um conjunto de *hashtags* que simbolizavam os jogos que estavam a decorrer nesses dias como por exemplo: “#BRAvsCOL”, “#NEDvsARG”, “#GERvsARG”, “#BRAvsGER”, “#BRAvsNED”, “#NEDvsESP” ou “#USAvsPOR”.

Relativamente ao número médio de caracteres que compõem as *hashtags*, a Figura 31 demonstra um dado importante relativamente ao aumento da dimensão dos tópicos e que mostra que todos os locais em análise registam um aumento da sua dimensão face às tendências que não são *hashtags*. Verifica-se neste tipo de casos que os países de língua portuguesa (Portugal e Brasil) são os que registam um aumento mais significativo do número de caracteres usados nos tópicos.

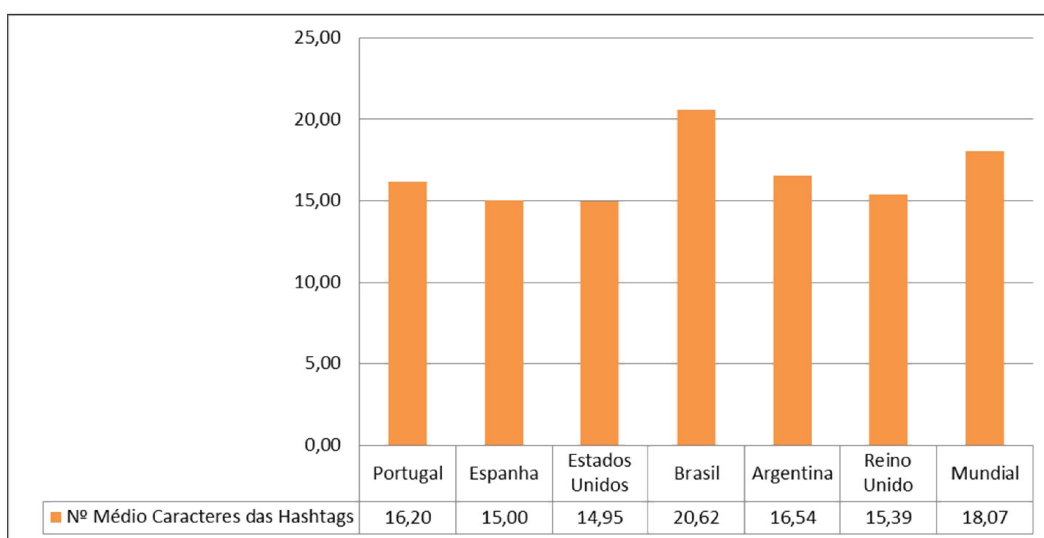


Figura 31 – Número médio de caracteres por *hashtag*

Ainda no contexto das *hashtags* são apresentados alguns números relativamente à repetição deste tipo de tendências. Nos dados extraídos são vários os casos em que se regista a repetição de determinadas *hashtags*. Assim na Tabela 24 surgem as *hashtags* que se repetiram em três ou mais ocasiões, sendo também apresentados alguns exemplos. Tal como na análise das repetições de tendências da secção 4.1.2, optou-se por restringir os dados por forma a filtrar apenas as *hashtags* mais dominantes e continuando Portugal a ser o local com maior número de repetições. Na secção de anexos apresentam-se na totalidade algumas tabelas com o número de *hashtags* que surgiram em três ou mais ocasiões.

Tabela 24 – Número de *hashtags* repetidas (três ou mais repetições)

País	Nº de Hashtags	Exemplos
PT	80	"#TwitterOff", "#RIR_SIC"
ES	27	"#FelizLunes", "#TourTVE"
USA	41	"#cantsleep", "#teamnosleep"
BRA	30	"#FelizDiaDoAmigo", "#MadrugaSporTV"
ARG	64	"#QuieroUnaSelfieCon", "#EsDeBuenAmigo"
UK	6	"#tracerequest", "#Wimbledon2014"
WORLD	2	"#AkhimyaMilihJokowi", "#BOYKOTVarSesVerTürkiye"

#### 4.1.7 Periodicidade de Tópicos

Algumas tendências têm a capacidade de se repetirem ao longo do tempo, quer seja no decorrer de um evento com algum impacto ou por mera vontade de utilizadores do Twitter em fazer de um determinado termo uma tendência do momento.

Seguidamente é apresentada a Tabela 25 que contém as tendências que mais se repetiram no período de recolha de dados.

Tabela 25 – Tendências repetidas (quinze ou mais repetições)

<i>Tendência</i>	<i>Repetições</i>	<i>Local da Tendência</i>	<i>Primeira Utilização</i>	<i>Última Utilização</i>
<i>Inês</i>	72	<i>PT</i>	<i>03-05-2014</i>	<i>06-07-2014</i>
<i>Lisboa</i>	70	<i>PT, ES</i>	<i>03-05-2014</i>	<i>21-07-2014</i>
<i>Netflix</i>	63	<i>USA</i>	<i>03-05-2014</i>	<i>21-07-2014</i>
<i>Mariana</i>	56	<i>PT</i>	<i>03-05-2014</i>	<i>21-07-2014</i>
<i>Raquel</i>	55	<i>PT</i>	<i>03-05-2014</i>	<i>19-07-2014</i>
<i>#TwitterOff</i>	53	<i>PT</i>	<i>03-05-2014</i>	<i>17-07-2014</i>
<i>Portimão</i>	43	<i>PT</i>	<i>07-06-2014</i>	<i>21-07-2014</i>
<i>Waffle House</i>	42	<i>USA</i>	<i>03-05-2014</i>	<i>20-07-2014</i>
<i>Algarve</i>	39	<i>PT</i>	<i>20-06-2014</i>	<i>21-07-2014</i>
<i>FaceTime</i>	38	<i>USA</i>	<i>06-05-2014</i>	<i>21-07-2014</i>
<i>Lucas</i>	36	<i>BRA, ARG</i>	<i>16-06-2014</i>	<i>19-07-2014</i>
<i>Rita</i>	36	<i>PT</i>	<i>03-05-2014</i>	<i>08-07-2014</i>
<i>Jeremy Kyle</i>	35	<i>UK, WORLD</i>	<i>16-06-2014</i>	<i>18-07-2014</i>
<i>Disney</i>	34	<i>PT, BRA, ARG</i>	<i>20-06-2014</i>	<i>19-07-2014</i>
<i>Joana</i>	32	<i>PT</i>	<i>03-05-2014</i>	<i>20-06-2014</i>
<i>Cali</i>	29	<i>USA</i>	<i>03-05-2014</i>	<i>21-07-2014</i>
<i>Madrid</i>	28	<i>ES</i>	<i>17-06-2014</i>	<i>21-07-2014</i>
<i>IHOP</i>	27	<i>USA</i>	<i>04-05-2014</i>	<i>08-07-2014</i>
<i>XDXD</i>	26	<i>PT</i>	<i>07-05-2014</i>	<i>21-07-2014</i>
<i>España</i>	25	<i>ES, ARG</i>	<i>05-05-2014</i>	<i>21-07-2014</i>
<i>#tracerequest</i>	24	<i>UK</i>	<i>16-06-2014</i>	<i>21-07-2014</i>
<i>Portugal</i>	24	<i>PT, ES, ARG</i>	<i>07-05-2014</i>	<i>17-07-2014</i>
<i>#cantsleep</i>	23	<i>USA</i>	<i>05-05-2014</i>	<i>15-07-2014</i>
<i>Sesimbra</i>	23	<i>PT</i>	<i>10-05-2014</i>	<i>16-07-2014</i>
<i>Agus</i>	21	<i>ARG</i>	<i>18-06-2014</i>	<i>21-07-2014</i>
<i>Nico</i>	21	<i>ARG, UK</i>	<i>17-06-2014</i>	<i>21-07-2014</i>
<i>Rock In Rio</i>	21	<i>PT</i>	<i>08-05-2014</i>	<i>01-06-2014</i>
<i>Matemática</i>	19	<i>PT</i>	<i>05-05-2014</i>	<i>23-06-2014</i>
<i>NyQuil</i>	19	<i>USA</i>	<i>05-05-2014</i>	<i>03-07-2014</i>
<i>#MTVHottest</i>	18	<i>PT, ES, USA, BRA, ARG, UK</i>	<i>19-07-2014</i>	<i>21-07-2014</i>
<i>Starbucks</i>	18	<i>USA</i>	<i>12-05-2014</i>	<i>21-07-2014</i>
<i>Sumol</i>	18	<i>PT</i>	<i>02-06-2014</i>	<i>29-06-2014</i>
<i>Sofia</i>	17	<i>PT, BRA</i>	<i>06-05-2014</i>	<i>22-06-2014</i>
<i>#RIR_SIC</i>	16	<i>PT</i>	<i>25-05-2014</i>	<i>02-06-2014</i>
<i>London</i>	16	<i>USA, UK</i>	<i>22-05-2014</i>	<i>21-07-2014</i>
<i>Argentina</i>	15	<i>PT, ES, USA, BRA, ARG</i>	<i>15-06-2014</i>	<i>14-07-2014</i>
<i>Tommy Ramone</i>	15	<i>PT, ES, USA, BRA, ARG, UK, WORLD</i>	<i>12-07-2014</i>	<i>12-07-2014</i>

Ao analisar a tabela anterior, a tendência “Tommy Ramone” referente ao falecimento de um dos fundadores da banda de música *Ramones*, sobressai pelo facto de ser uma tendência que



surgiu apenas durante um dia (12/07/2014) tendo no entanto registado quinze repetições e tendo sido tendência em todos os locais escolhidos neste estudo sendo inclusivamente tendência a nível mundial.

Por exemplo, o termo “#TwitterOff” usado em Portugal repete-se 53 vezes apresentando assim uma periodicidade elevada, distribuída entre dia 3 de Maio e 17 de Julho e que é usado quando os utilizadores pretendem comunicar aos seus contactos que vão sair das suas contas do Twitter. O termo “#cantsleep” surge 23 vezes nos Estados Unidos e aparece relacionado com utilizadores do Twitter que se encontram acordados de madrugada sem conseguir dormir visto que o significado desta *hashtag* e a hora a que se referem assim o indica. O termo “Netflix” apresenta também uma periodicidade elevada surgindo 63 vezes apenas nos Estados Unidos. Este termo refere-se ao serviço de televisão por Internet e é tendência nos Estados Unidos sempre em períodos noturnos.

O festival Rock in Rio, um evento musical com grande expressão em Portugal, teve lugar nos dias 25, 29, 30, 31 de Maio e 1 de Junho. No entanto, o evento começou a ser tendência em Portugal algum tempo antes sendo isso demonstrado na Tabela 25 e na Figura 32.

A primeira referência feita ao festival captada na extração surge no dia 8 de Maio com termo “Rock In Rio”, repetindo-se novamente dia 19 de Maio. Este termo com o aproximar do primeiro dia de festival surge diversas vezes, sendo que no dia 25 de Maio, primeiro dia do festival, surge uma nova tendência “#RIR\_SIC”, sugerida pelo canal de televisão Sic Radical que transmitiu o evento. No dia 26 surge mais um tópico associado ao festival, “#RIR2014”, registando-se um período de dois dias de inatividade (27 e 28 de Maio). No dia 29 de Maio, segundo dia do festival, as tendências “Rock In Rio” e “#RIR\_SIC” voltam a emergir mantendo-se diariamente até ao último dia do evento (1 de Junho). No dia 2 de Junho surge a tendência “Rir 2014” e repete-se a “#RIR\_SIC”, sendo este o último dia em que são captadas as tendências com referência direta ao evento Rock in Rio.

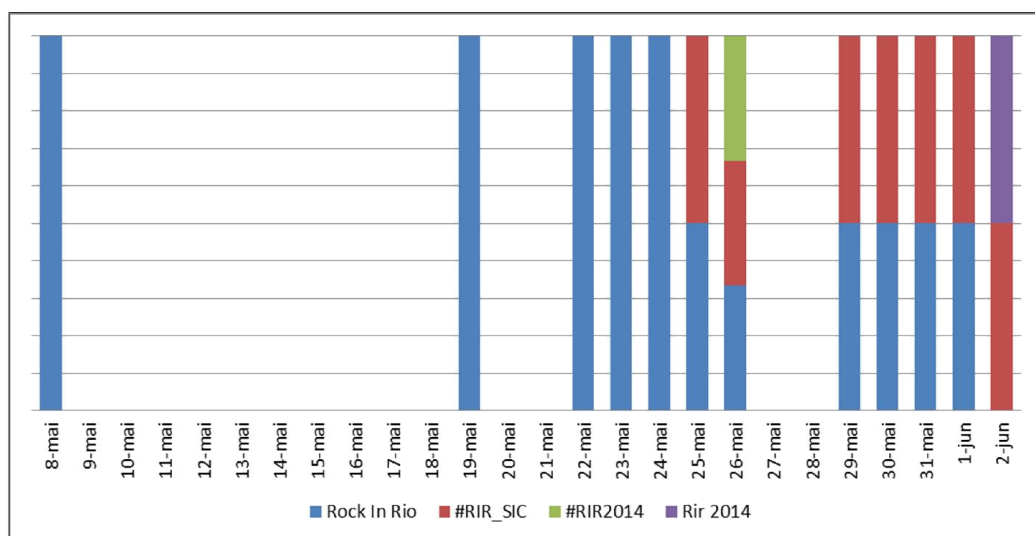


Figura 32 – Tendências do festival Rock in Rio

Outro dos fenómenos em análise está relacionado com os artistas que atuaram no festival e que foram também tendências em Portugal estando presentes na Figura 33.

Na maioria dos casos, os termos referentes a estes artistas tornam-se tendência no dia do concerto e no dia seguinte. No entanto, o termo “*Rolling Stones*” alusivo à banda de música oriunda do Reino Unido foi tendência durante quatro dias (dois dias anteriores, no dia do concerto e no dia seguinte), o que pode ser um indicador da popularidade que a banda possui em Portugal.

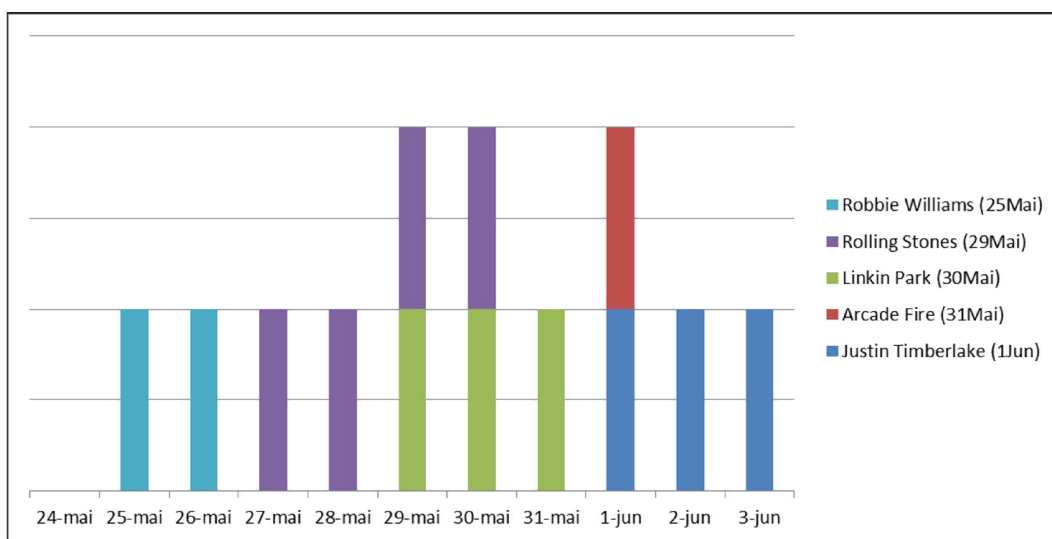


Figura 33 – Tendências de artistas presentes no festival Rock in Rio

## 4.2 Análise das Tendências - Outras Fontes de Dados

Como já foi referido anteriormente, foi feita uma extração diária das tendências do Twitter em vários países/locais distintos, durante um período de tempo que decorreu nos meses de Maio, Junho e parte de Julho.

Juntamente com a extração das tendências efetuou-se também uma extração diária de alguns dos jornais diários dos países em estudo. Essa recolha de dados teve uma duração mais curta (19 dias), mas suficiente para concluir que por vezes algumas das tendências do Twitter podem eventualmente ser mencionadas nos jornais diários, dizendo respeito a notícias da atualidade quer sejam de índole desportiva, política, musical, entre outras.

Para tornar a análise mais completa adicionou-se também uma terceira fonte, com os dados provenientes da *DBpedia* onde se extraíram dados sobre países, capitais europeias, cidades portuguesas, desportistas (futebolistas, tenistas e basquetebolistas) e ainda alguns festivais de música portugueses.

Seguidamente são apresentadas diversas tabelas e gráficos que apresentam algumas das tendências que também surgiram nos jornais escolhidos neste estudo. A comparação

efetuada apresenta para um determinado termo, o local onde esse termo foi tendência, a descrição oriunda da *DBpedia* e as repetições que esse termo apresenta nos jornais escolhidos sendo demonstrados apenas alguns exemplos.

Na Figura 34 as tendências que se destacam abordam o tema do mundial de futebol do Brasil com o termo “*Copa*”, sendo importante referir que este termo surge nos jornais do Brasil, Espanha e Argentina mas apenas é tendência do Twitter no Brasil. O termo “*Madrid*” apresenta muitas referências no jornal espanhol *El Mundo*, sendo também referido em menor número no diário português *Jornal de Notícias*. O termo “*Lisboa*” surge apenas no jornal português e como tendência em Portugal.

Tabela 26 – Comparação de dados do Twitter, *DBpedia* e jornais online (03/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da <i>DBpedia</i>	Repetições nos jornais dos países
<i>Copa</i>	Brasil	-----	ES – 1; BRA – 129; ARG – 4;
<i>Madrid</i>	Espanha	<i>Madrid é a capital da Espanha</i>	PT – 6; ES – 56;
<i>Lisboa</i>	Portugal	<i>Lisboa é a capital de Portugal</i>	PT – 17;

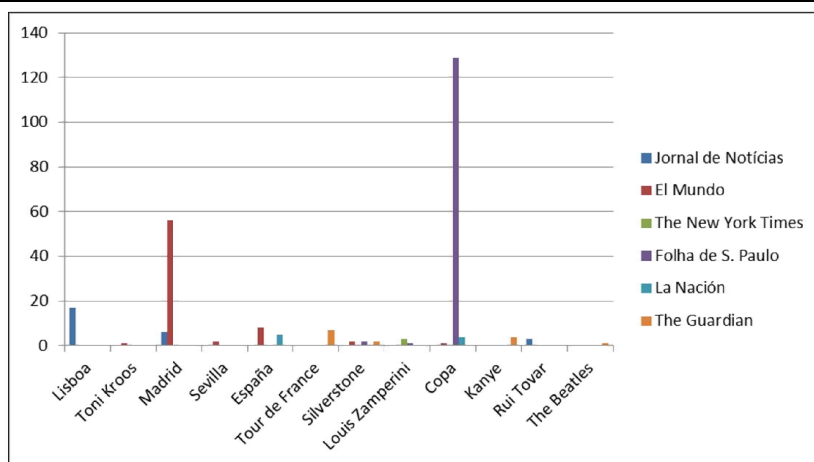


Figura 34 – Tendências nos jornais em 03/07/2014

No dia 4 de Julho (Tabela 27 e Figura 35), surge novamente o termo “*Madrid*”, acompanhado de perto pelo termo “*Lisboa*”. Por último, o termo “*Colombia*” foi tendência em vários locais e surgiu na maioria dos jornais extraídos motivado sobretudo pelo jogo da seleção de futebol da Colômbia a decorrer nesse dia.

Tabela 27 – Comparação de dados do Twitter, *DBpedia* e jornais online (04/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da <i>DBpedia</i>	Repetições nos jornais dos países
<i>Madrid</i>	Espanha	<i>Madrid é a capital da Espanha</i>	PT – 2; ES – 43
<i>Lisboa</i>	Portugal	<i>Lisboa é a capital de Portugal</i>	PT – 26;
<i>Colombia</i>	Estados Unidos, Argentina	<i>País</i>	ES – 7; USA – 8; BRA – 11; ARG – 15; UK – 12;

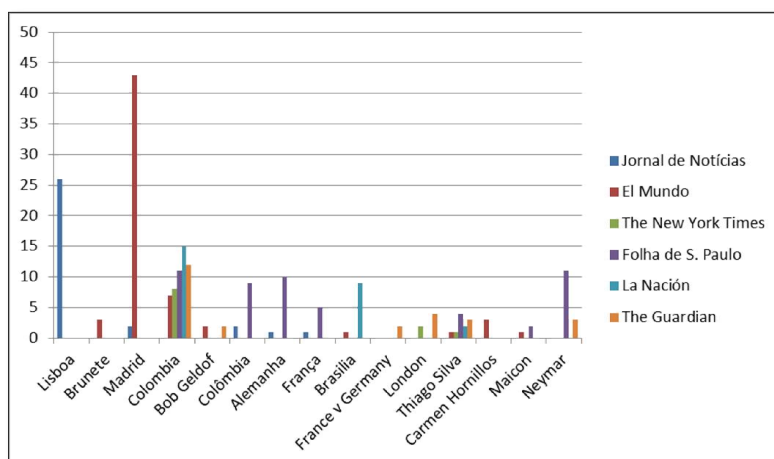


Figura 35 – Tendências nos jornais em 04/07/2014

No dia 5 de Julho (Tabela 28 e Figura 36), surgem os termos “Argentina”, “Neymar” e “Holanda” todos eles relacionados com o mundial de futebol.

Tabela 28 – Comparação de dados do Twitter, DBpedia e jornais online (05/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Argentina	Brasil	País	PT – 3; ES – 8; USA – 7; BRA – 19; ARG – 119; UK – 10;
Neymar	Portugal, Argentina	Futebolista	PT – 3; ES – 7; USA – 2; BRA – 47; ARG – 26; UK – 18;
Holanda	Portugal	País	ES – 9; BRA – 8; ARG – 16;

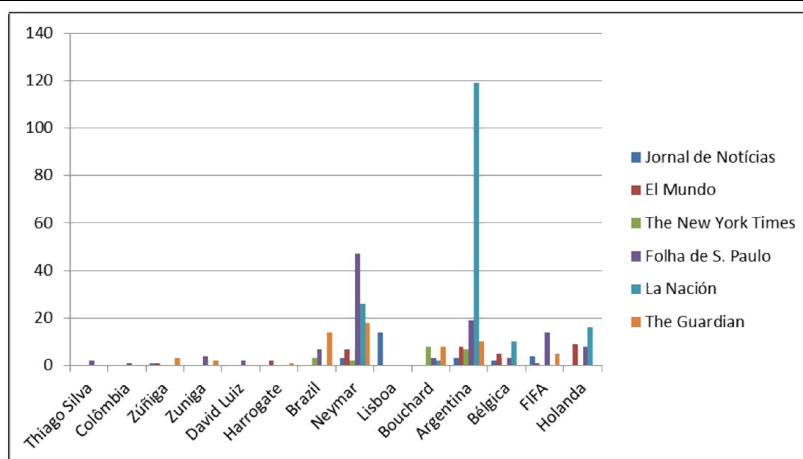


Figura 36 – Tendências nos jornais em 05/07/2014

No dia 6 de Julho (Tabela 29 e Figura 37), surgem os termos “Novak” e “Federer”, referindo-se aos tenistas que disputavam nesse dia a final do torneio de Wimbledon 2014.

Tabela 29 – Comparação de dados do Twitter, DBpedia e jornais online (06/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Novak	Estados Unidos, Reino Unido	Tenista	PT – 1; USA – 8; BRA – 3; ARG – 9; UK – 12;
Federer	Portugal	Tenista	ES – 1; USA – 9; BRA – 10; ARG – 12; UK – 10;
Lisboa	Portugal	Lisboa é a capital de Portugal	PT – 23;

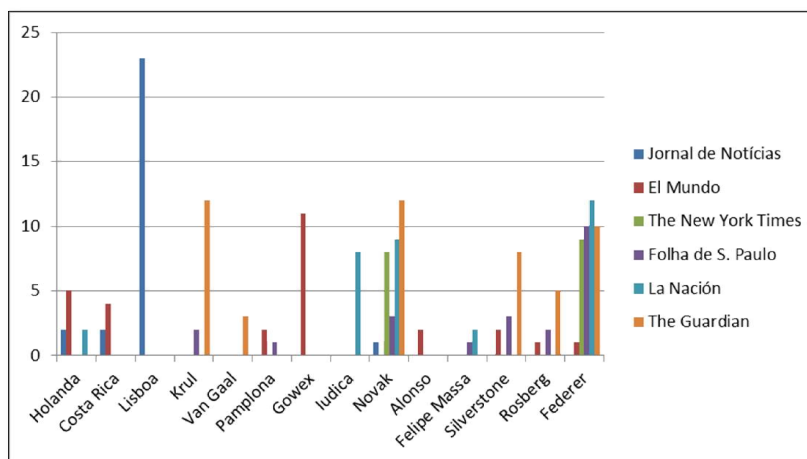


Figura 37 – Tendências nos jornais em 06/07/2014

No dia 7 de Julho (Tabela 30 e Figura 38), como principais termos surgem apenas nomes de cidades como por exemplo “Madrid”, “Barcelona” ou “Lisboa”.

Tabela 30 – Comparação de dados do Twitter, DBpedia e jornais online (07/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Madrid	Espanha	Madrid é a capital da Espanha	PT – 4; ES – 36; BRA – 3; UK - 4
Barcelona	Espanha	-----	PT – 2; ES – 12; ARG – 2; UK – 6;
Lisboa	Portugal	Lisboa é a capital de Portugal	PT – 18;

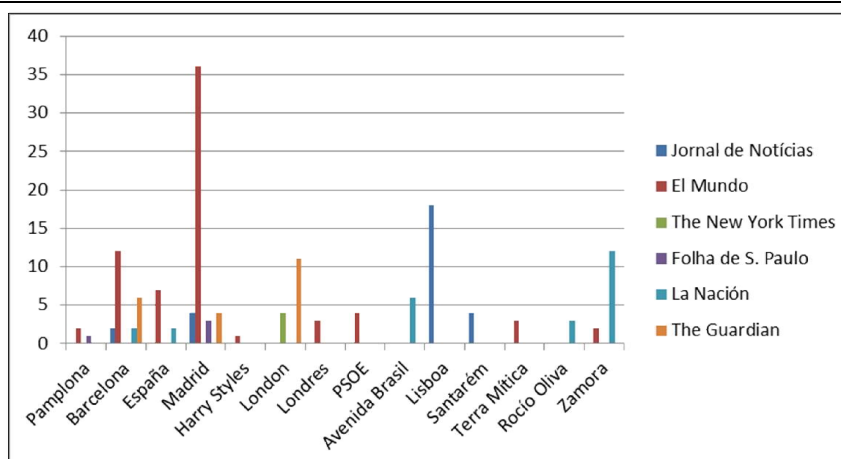


Figura 38 – Tendências nos jornais em 07/07/2014

No dia 8 de Julho (Tabela 31 e Figura 39), surge novamente o termo “Madrid”. Associados ao mundial de futebol surgem os termos “Alemanha” e “Klose” numa clara alusão ao jogo de futebol da seleção alemã a decorrer nesse dia e ao jogador alemão.

Tabela 31 – Comparação de dados do Twitter, DBpedia e jornais online (08/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Madrid	Espanha	Madrid é a capital da Espanha	PT – 2; ES – 33; USA – 1;
Alemanha	Brasil	País	PT – 2; BRA – 22;
Klose	Portugal	Futebolista	ES – 1; BRA – 6; ARG – 15; UK – 2;

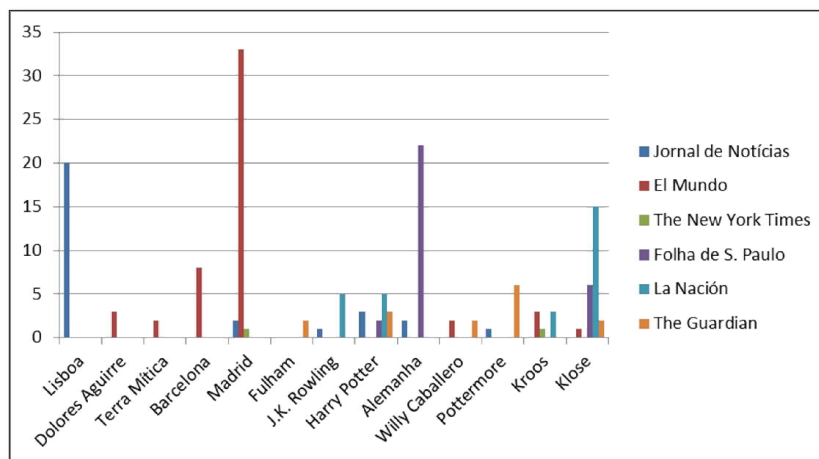


Figura 39 – Tendências nos jornais em 08/07/2014

No dia 9 de Julho (Tabela 32 e Figura 40), surgem os termos “Argentina” e “Holanda” estando associados ao jogo de futebol entre as duas seleções. O termo “Israel” emerge relacionado com o conflito armado na faixa de Gaza.

Tabela 32 – Comparação de dados do Twitter, DBpedia e jornais online (09/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Argentina	Portugal	País	ES – 10; USA – 11; BRA – 17; ARG – 43; UK – 26;
Holanda	Portugal	País	ES – 10; BRA – 5; ARG – 28;
Israel	Espanha	País	ES – 6; USA – 24; BRA – 2; ARG – 25; UK – 19;

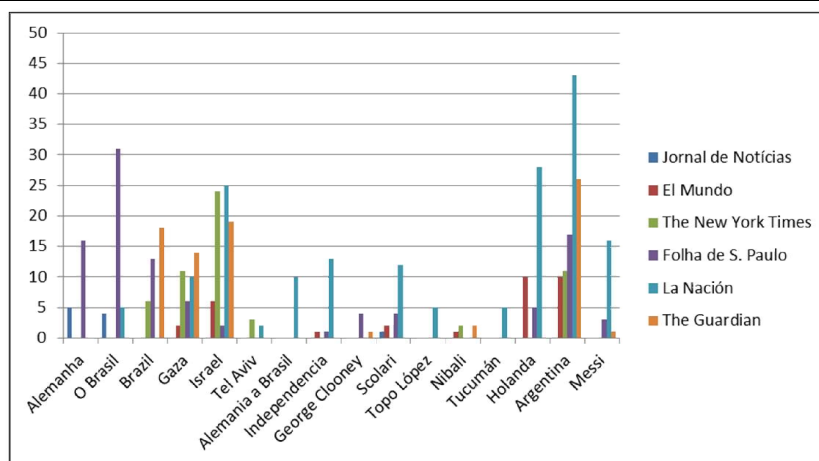


Figura 40 – Tendências nos jornais em 09/07/2014

No dia 10 de Julho (Tabela 33 e Figura 41), surgem os termos “Argentina” associado à seleção de futebol e também os termos “Israel” e “Gaza” motivado pelo conflito na faixa de Gaza.

Tabela 33 – Comparação de dados do Twitter, DBpedia e jornais online (10/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Argentina	Portugal, Espanha	País	PT – 2; ES – 8; USA – 6; BRA – 14; ARG – 81; UK – 11;
Israel	Brasil	País	PT – 2; ES – 6; USA – 10; BRA – 4; ARG – 14; UK – 20;
Gaza	Espanha	-----	ES – 2; USA – 12; BRA – 8; ARG – 10; UK – 22;

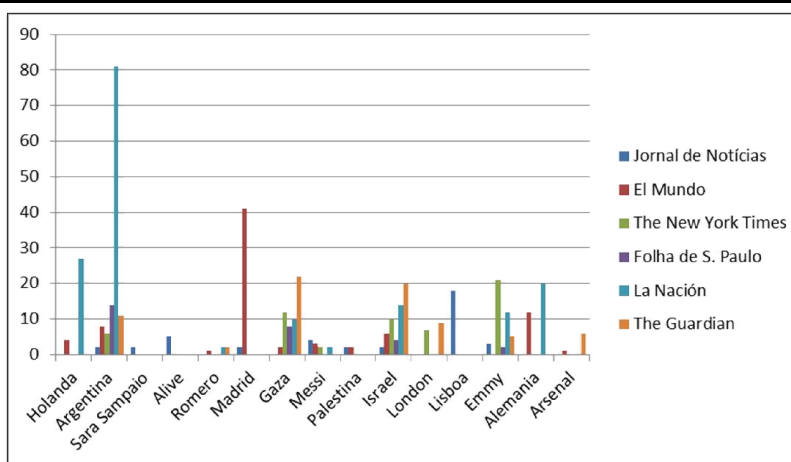


Figura 41 – Tendências nos jornais em 10/07/2014

No dia 11 de Julho (Tabela 34 e Figura 42), surgem os termos “Madrid” e “Barcelona” referências a cidades espanholas e o termo “LeBron” associado ao jogador de basquetebol americano LeBron James.

Tabela 34 – Comparação de dados do Twitter, DBpedia e jornais online (11/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Madrid	Espanha	Madrid é a capital da Espanha	PT – 3; ES – 35;
Barcelona	Brasil, Reino Unido	-----	ES – 16; USA – 4; BRA – 2; UK – 9;
LeBron	Estados Unidos	Basquetebolista	ES – 4; USA – 14; BRA – 2; UK – 5;

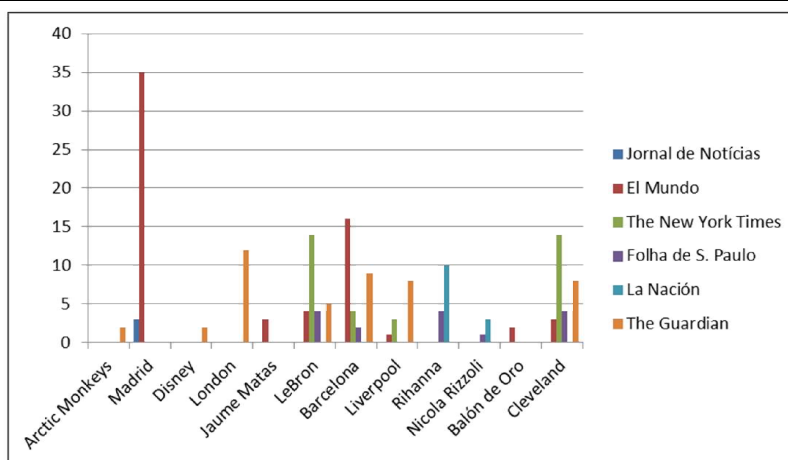


Figura 42 – Tendências nos jornais em 11/07/2014

No dia 12 de Julho (Tabela 35 e Figura 43), emergem os termos “*Holanda*” relacionado com o jogo da seleção holandesa e “*Madrid*” numa referência à capital espanhola. O termo “*Putin*” refere-se ao presidente Russo e à visita que realizava à Argentina nesse dia, sendo importante referir que este termo também era tendência na Argentina.

Tabela 35 – Comparação de dados do Twitter, *DBpedia* e jornais online (12/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da <i>DBpedia</i>	Repetições nos jornais dos países
<i>Holanda</i>	Brasil	País	PT – 3; ES – 10; BRA – 12; ARG – 27;
<i>Madrid</i>	Espanha	Madrid é a capital da Espanha	PT – 2; ES – 46;
<i>Putin</i>	Argentina	-----	ES – 5; BRA – 2; ARG – 23;

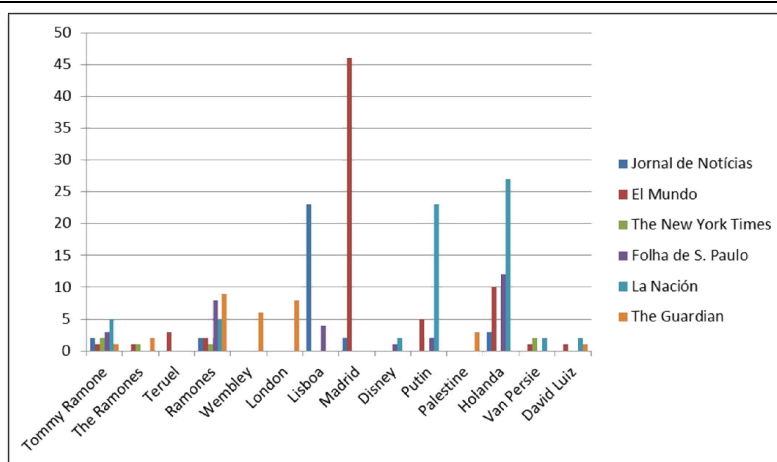


Figura 43 – Tendências nos jornais em 12/07/2014

No dia 13 de Julho (Tabela 36 e Figura 44), destacaram-se vários termos sendo quase todos relacionados com a final do mundial de futebol no Brasil. Os termos “*Maracanã*” e “*Maracanã*” referem-se ao estádio onde se realizou a final. O termo com mais referências é “*Alemania*” numa clara alusão dos jornais de língua espanhola (*El Mundo* e *La Nación*) à seleção alemã campeã do mundo seguido do termo “*Germany*” que foi referido pelos jornais de língua inglesa (*The New York Times* e *The Guardian*).

Tabela 36 – Comparação de dados do Twitter, *DBpedia* e jornais online (13/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da <i>DBpedia</i>	Repetições nos jornais dos países
<i>Alemania</i>	Argentina	País	ES – 22; ARG – 38;
<i>Germany</i>	Reino Unido	País	USA – 20; UK – 27;
<i>Lisboa</i>	Portugal	Lisboa é a capital de Portugal	PT – 33; BRA – 2;



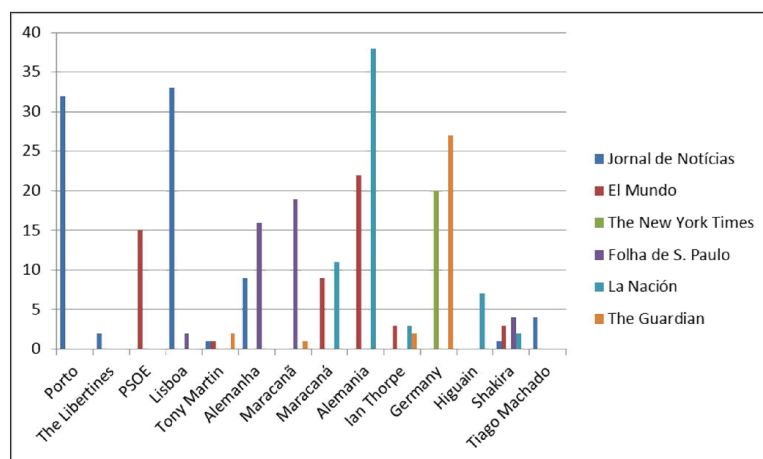


Figura 44 – Tendências nos jornais em 13/07/2014

No dia 14 de Julho (Tabela 37 e Figura 45), surgem os termos “Argentina”, “Obelisco” e “Messi” todos eles relacionados com a seleção Argentina e com o jogador Lionel Messi.

Tabela 37 – Comparação de dados do Twitter, DBpedia e jornais online (14/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Argentina	Argentina	Pais	PT – 2; ES – 4; USA – 4; BRA – 12; ARG – 20; UK – 13;
Obelisco	Argentina	-----	ARG – 50;
Messi	Espanha, Argentina	Futebolista	PT – 3; ES – 4; BRA – 9; ARG – 16; UK – 5;

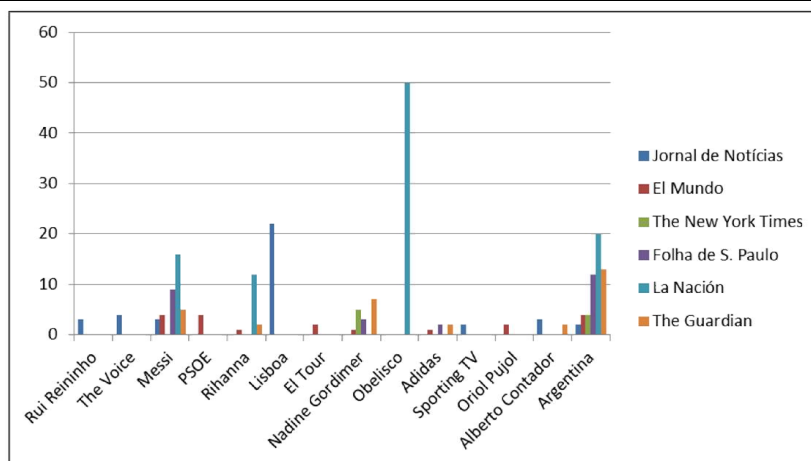


Figura 45 – Tendências nos jornais em 14/07/2014

No dia 15 de Julho (Tabela 38 e Figura 46), o termo “BRICS” é o único que surge como novidade face às tendências de dias anteriores, numa alusão aos países economicamente emergentes dos quais o Brasil faz parte. De realçar o facto do termo em questão ser tendência precisamente no Brasil.

O termo “BRICS” repete-se também no dia 16 de Julho (Tabela 39 e Figura 47).

Tabela 38 – Comparação de dados do Twitter, DBpedia e jornais online (15/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Madrid	Espanha	Madrid é a capital da Espanha	PT – 2; ES – 28;
BRICS	Brasil	-----	BRA – 14; ARG – 14;
Lisboa	Portugal	Lisboa é a capital de Portugal	PT – 24;

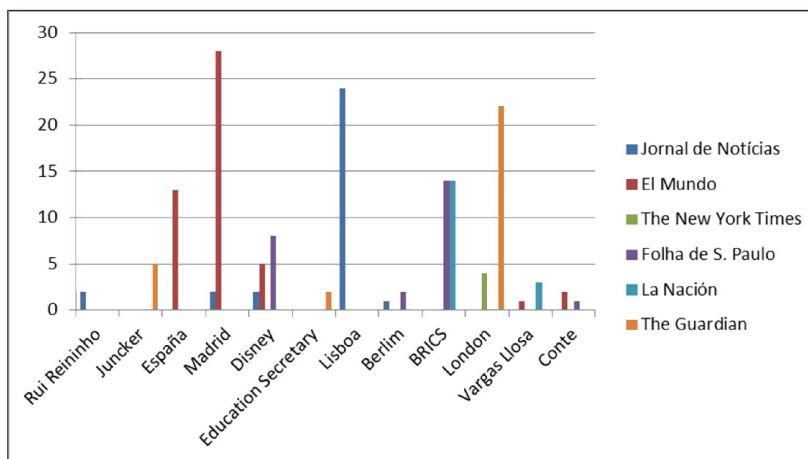


Figura 46 – Tendências nos jornais em 15/07/2014

Tabela 39 – Comparação de dados do Twitter, DBpedia e jornais online (16/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
BRICS	Brasil	-----	BRA – 8; ARG – 27; UK – 2;
Pergolini	Argentina	-----	ARG – 28;
Riquelme	Argentina	-----	ARG – 26;

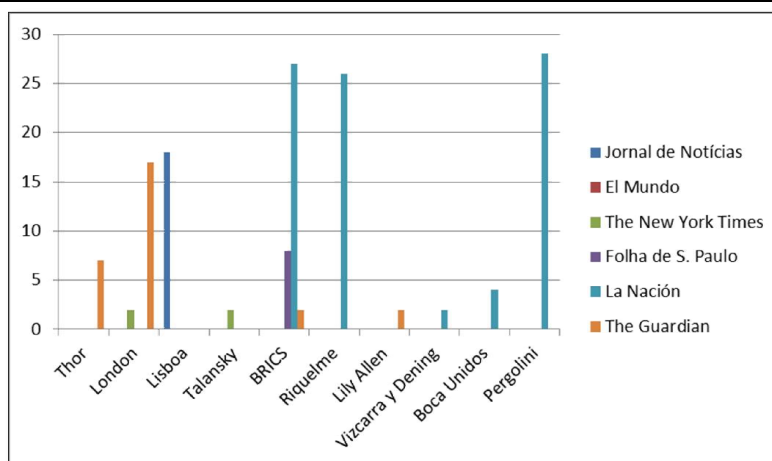


Figura 47 – Tendências nos jornais em 16/07/2014

No dia 17 de Julho (Tabela 40 e Figura 48), surge novamente o termo “Madrid”, acompanhado de perto pelos termos “Ucrania”, “Ukraine” e “Malaysia Airlines” que se referem à queda de um avião da companhia aérea da Malásia na Ucrânia.

Tabela 40 – Comparação de dados do Twitter, *DBpedia* e jornais online (17/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da <i>DBpedia</i>	Repetições nos jornais dos países
Ucrania	Espanha, Argentina	País	ES – 5; BRA – 13; ARG – 37;
Madrid	Espanha	Madrid é a capital da Espanha	PT – 2; ES – 48;
Malaysia Airlines	Argentina	-----	ES – 2; USA – 15; BRA – 1; ARG – 20; UK – 9;

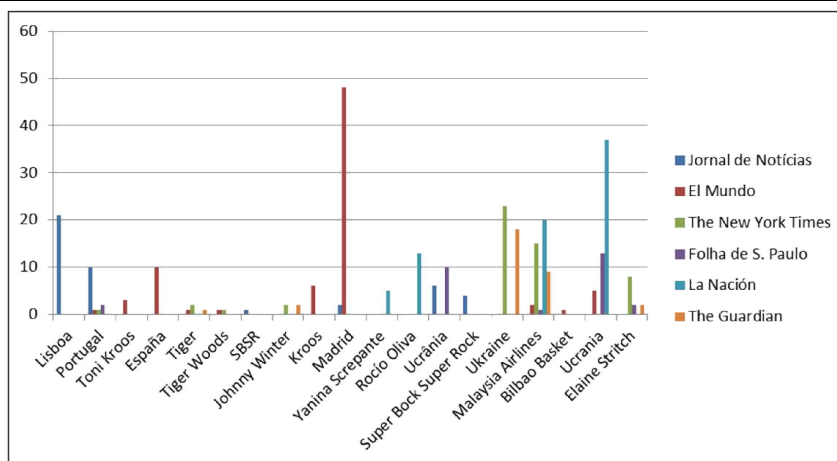


Figura 48 – Tendências nos jornais em 17/07/2014

No dia 18 de Julho (Tabela 41 e Figura 49), os termos sobre o desastre de avião na Ucrânia surgem novamente através de “*Malaysia Airlines*”, “*Ucrania*”, “*Ukraine*” e “*Ucrânia*”. Os termos relativos à Ucrânia aparecem escritos consoante a língua dos jornais em questão.

Tabela 41 – Comparação de dados do Twitter, *DBpedia* e jornais online (18/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da <i>DBpedia</i>	Repetições nos jornais dos países
Malaysia Airlines	Espanha, Estados Unidos	-----	PT – 2; ES – 2; USA – 28; BRA – 2; ARG – 20; UK – 11;
Ukraine	Estados Unidos, Reino Unido	País	USA – 36; UK – 26;
Ucrania	Espanha, Argentina	País	ES – 6; BRA – 8; ARG – 19;

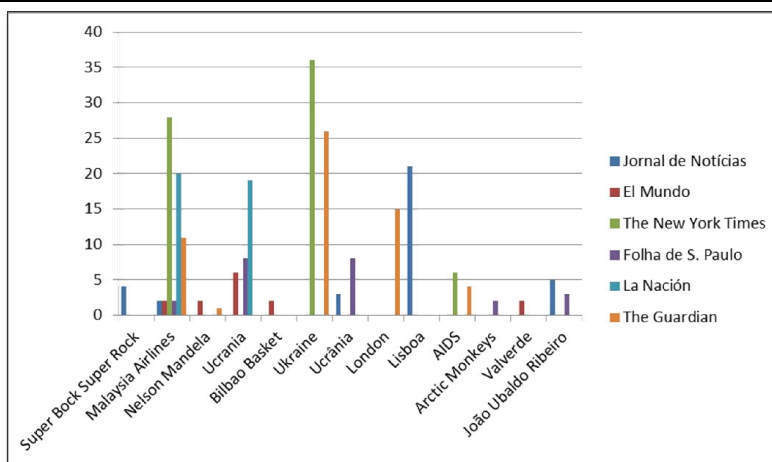


Figura 49 – Tendências nos jornais em 18/07/2014

Nos restantes dias de extração (19, 20 e 21 de Julho) não surgiram termos suscetíveis de comentários já que não diferiam muito dos apresentados anteriormente.

Tabela 42 – Comparação de dados do Twitter, DBpedia e jornais online (19/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Madrid	Espanha	Madrid é a capital da Espanha	PT – 3; ES – 27; BRA – 4;
China	Argentina	País	ES – 1; ARG – 19; UK – 4;
Lisboa	Portugal	Lisboa é a capital de Portugal	PT – 14;

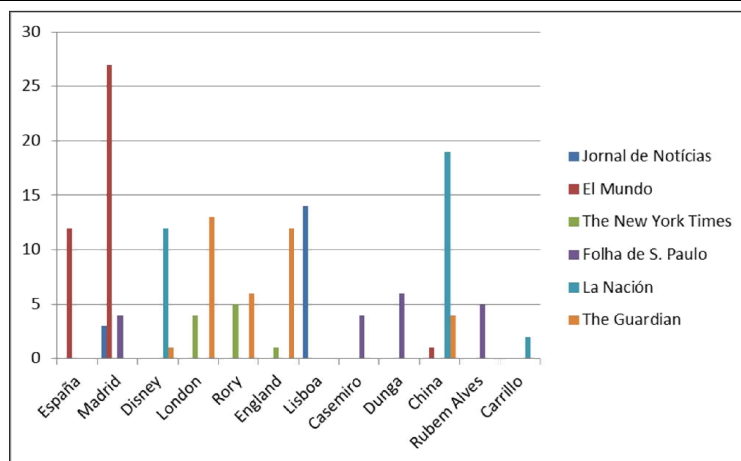


Figura 50 – Tendências nos jornais em 19/07/2014

Tabela 43 – Comparação de dados do Twitter, DBpedia e jornais online (20/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da DBpedia	Repetições nos jornais dos países
Madrid	Espanha	Madrid é a capital da Espanha	PT – 2; ES – 37; BRA – 2;
Rory	Reino Unido	-----	ES – 1; USA – 8; ARG – 12; UK – 10;
Rosberg	Argentina	-----	PT – 1; ES – 3; BRA – 4; ARG – 12; UK – 5;

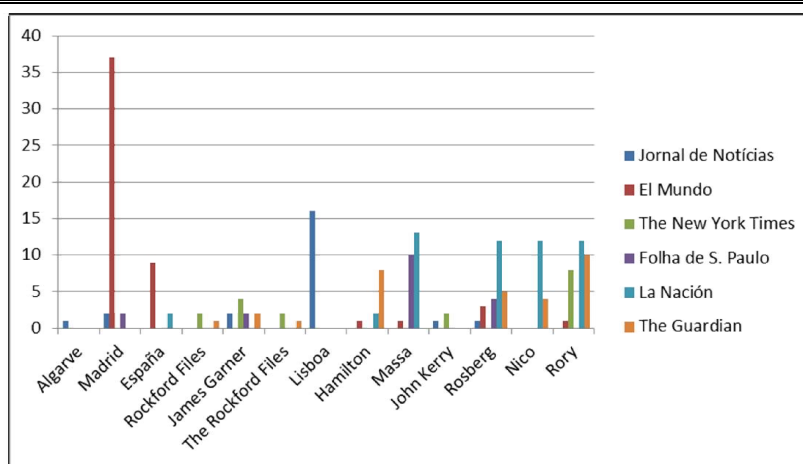


Figura 51 – Tendências nos jornais em 20/07/2014

Tabela 44 – Comparação de dados do Twitter, *DBpedia* e jornais *online* (21/07/2014)

Tendência	Local da Tendência	Descrição do termo extraído da <i>DBpedia</i>	Repetições nos jornais dos países
Madrid	Espanha	Madrid é a capital da Espanha	PT – 2; ES – 37;
Sarmiento	Argentina	-----	ARG – 28;
London	Reino Unido	-----	USA – 4; UK – 19;

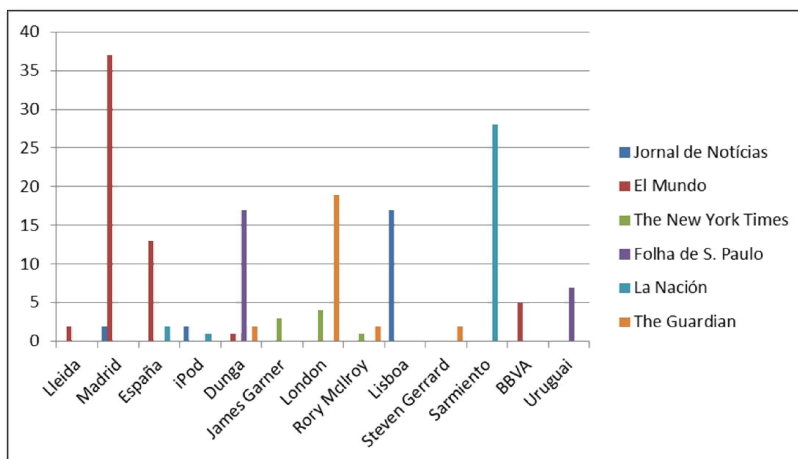


Figura 52 – Tendências nos jornais em 21/07/2014

Apesar do período de recolha de dados dos jornais *online*, ter sido bastante reduzido face ao período de extração das tendências do Twitter, foi possível comprovar que por vezes essas mesmas tendências podem surgir noutras fontes de dados. As comparações efetuadas basearam-se apenas em aspetos meramente sintáticos sem abordar questões de semântica.



## 5 Conclusões e Trabalho Futuro

Neste capítulo são apresentadas algumas conclusões através da análise de resultados mostrando resumidamente aquilo que foi desenvolvido ao longo do trabalho.

São também apresentadas algumas limitações encontradas no trabalho, sendo por último apresentadas possíveis melhorias a implementar no mesmo.

### 5.1 *Análise de Resultados*

Depois de efetuado um estudo alargado do Twitter e do seu interface de programação, procedeu-se à elaboração de um sistema que permitisse a extração e o armazenamento das tendências do Twitter para posterior análise desses dados.

O código desenvolvido e apresentado ao longo do trabalho contribuiu para a construção de diversos extratores, cujo objetivo foi a extração das tendências do Twitter durante um período de tempo alargado, o que permitiu a recolha suficiente de tendências para construir uma amostra de dados de tamanho considerável, tornando o processo de análise de dados mais consistente e completo no que respeita a obtenção de determinadas conclusões.

Os extratores das tendências foram utilizados mediante a execução de tarefas definidas no sistema operativo e as tendências foram posteriormente gravadas com sucesso na base de dados SQL criada no âmbito do trabalho.

Ainda na extração de dados, foram consideradas outras fontes de dados, nomeadamente jornais *online* e ainda dados provenientes da *DBpedia* sendo estes últimos extraídos com recurso a técnicas e metodologias pertencentes à *Semantic Web*.

Os dados extraídos foram seguidamente explorados e trabalhados com recurso a tabelas dinâmicas (*pivot tables*), que estando diretamente ligadas à base de dados permitiram um manuseamento mais fácil dos dados e a construção de gráficos e tabelas numa forma mais eficaz e intuitiva.

Os objetivos definidos inicialmente foram atingidos, tendo sido analisadas as tendências em parâmetros tão distintos como por exemplo, o número de tendências diferentes e repetidas registadas nos diversos locais em estudo, o número médio de caracteres e palavras por tendência e ainda o número de palavras existentes nos tópicos. Foi também analisada a utilização das *hashtags* em diversos aspetos como dimensão das mesmas ao nível de caracteres, repetições e ainda comparações entre totais de *hashtags* e totais de tendências extraídas. Quanto às tendências repetidas verificou-se que existem tópicos que são comuns a mais do que um local e que existem inúmeras tendências que se repetem ao longo do tempo sendo no entanto a maioria das tendências esporádicas sem repetições.

Foi analisada a periodicidade dos tópicos que foram tendências, nomeadamente aqueles que mais vezes surgiram durante o período de recolha. Alguns estavam relacionados com eventos a decorrer, como por exemplo, o festival Rock in Rio em que se conseguiu demonstrar que as tendências sobre o evento começaram a surgir alguns dias antes do mesmo ter início prolongando-se até um dia depois do fim do festival.

Por último analisaram-se algumas tendências que em determinados dias também surgiram em diversos jornais *online* permitindo assim atingir diversas conclusões sobre a capacidade de certas tendências se referirem a assuntos que também são notícia em um ou mais locais.

## 5.2 Limitações

Ao ser desenvolvido este trabalho foram surgindo algumas dificuldades no processo de extração das tendências onde foram encontradas algumas limitações relacionadas com as características do próprio interface de programação do Twitter, nomeadamente nas chamadas que se podiam fazer ao mesmo. Esses pedidos estavam limitados a um número reduzido e a uma janela temporal bastante curta com apenas 15 minutos. No entanto, após alguma familiarização com o código a usar para cada funcionalidade, o número de pedidos efetuados à Twitter API decresceu e os limites disponibilizados pela plataforma raramente foram ultrapassados.

A própria familiarização com a linguagem Python foi também uma adversidade que inicialmente dificultou o processo evolutivo do trabalho mas que foi ultrapassada através do recurso a diversas fontes *online*, incluindo alguma bibliografia sobre o tema.

Outra limitação deste estudo surgiu relacionada com o facto de não ser possível determinar quais as tendências mais fortes de um determinado dia. Por outras palavras, as tendências extraídas em períodos fixos não são necessariamente as maiores tendências do dia em questão, mas sim as tendências registadas à hora da extração especificada pelo utilizador.



Os próprios jornais *online*, ao serem recolhidos a uma hora em específico também estão limitados, visto que nas edições *online* as notícias podem mudar diversas vezes ao dia ao contrário dos jornais em formato papel que habitualmente registam apenas uma tiragem diária.

O processo de descoberta de conteúdo estruturado na *Semantic Web* também não se verificou muito intuitivo, pelo que houve alguma dificuldade aquando da construção das *queries* SPARQL que pudessem satisfazer as necessidades de recolha de dados em diversos domínios.

A própria comparação entre as tendências e os dados provenientes da *DBpedia* não é completamente precisa, já que por vezes os termos não aparecem escritos da mesma forma em relação ao disponibilizado na *DBpedia*. Ao considerar apenas aspetos sintáticos, por exemplo, a tendência "*Rock In Rio*" foi facilmente comparável com o termo presente nos jornais mas a tendência "*RIR*" já não foi encontrada nos jornais dizendo no entanto respeito ao mesmo evento.

### **5.3 Possíveis Melhorias**

Existem algumas melhorias que se poderiam introduzir neste trabalho, nomeadamente no que diz respeito ao processo de extração de tendências propriamente dito, a amostra extraída poderia ser maior através da recolha de dados com uma frequência superior à utilizada neste trabalho. Desta forma e com uma base de dados de tamanho consideravelmente superior, seriam atingidas outras conclusões ou poderiam ser consolidadas as conclusões já atingidas. No entanto, tal alteração exigiria outros recursos para este trabalho.

No futuro também se podiam analisar *tweets* completos relacionados com as tendências atuais. Assim, poderiam ser aplicadas técnicas de *Text Mining* e análise de sentimentos (*sentiment analysis*) para se verificar a forma como os utilizadores estão comentar esses tópicos.

Podiam ser também analisados utilizadores que utilizam os tópicos mais comentados nos seus *tweets*, consoante diversos aspetos como, por exemplo, o número de seguidores de um utilizador, a faixa etária dos utilizadores, a sua atividade na rede, a sua localização geográfica e as suas preferências.

Os dados extraídos da *DBpedia* podiam ter uma maior amplitude no que diz respeito à temática dos mesmos, abordando outras áreas de interesse como música, política, outros desportos e até programas de televisão, permitindo um cruzamento mais completo entre os dados extraídos do Twitter e os dados da *DBpedia*.

Por fim, o trabalho poderia ser estendido para outras redes sociais ou plataformas, pois o que as pessoas comentam, compram, consultam ou pesquisam podem ser considerados como sendo dados valiosos e de elevado interesse em muitas áreas de investigação.



# Referências

- [Alexa.com, 2014] Alexa.com, (2014). twitter.com Site Overview. [online] Disponível em: <http://www.alexa.com/siteinfo/twitter.com> [Acedido em 27 Jul. 2014].
- [Aristóteles, 350AC] Aristóteles. (350AC). Política, Livro 1, Capítulo II.
- [Asur and Huberman, 2010] Asur, S. and Huberman, B. (2010). Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on (Vol. 1, pp. 492-499).
- [Bizer et al., 2009] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. and Hellmann, S. (2009). DBpedia-A crystallization point for the Web of Data. Web Semantics: science, services and agents on the world wide web, 7(3), pp.154-165.
- [Bollen, Mao and Zeng, 2011] Bollen, J., Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), pp.1-8.
- [Chang, 2010] Chang, H. (2010). A new perspective on Twitter hashtag use: diffusion of innovation theory. Proceedings of the American Society for Information Science and Technology, 47(1), pp.1-4.
- [Cheong and Lee, 2009] Cheong, M., & Lee, V. (2009). Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. In Proceedings of the 2nd ACM workshop on Social web search and mining (pp. 1-8). ACM.
- [Costa, 2012] Costa, J. (2012). Current status: API v1.1 [online] Twitter Blogs. Disponível em: <https://blog.twitter.com/2012/current-status-api-v1.1> [Acedido em 13 Mar. 2014].
- [Dev.twitter.com, 2013] Dev.twitter.com, (2013). GET trends / place. [online] Disponível em: <https://dev.twitter.com/docs/api/1.1/get/trends/place> [Acedido em 24 Mai. 2014].
- [Dev.twitter.com, 2014a] Dev.twitter.com, (2014). API Overview. [online] Disponível em: <https://dev.twitter.com/overview/api> [Acedido em 11 Set. 2014].
- [Dev.twitter.com, 2014b] Dev.twitter.com, (2014). The Streaming APIs. [online] Disponível em: <https://dev.twitter.com/docs/api/streaming> [Acedido em 24 Mai. 2014].
- [Dev.twitter.com, 2014c] Dev.twitter.com, (2014). API Rate Limits. [online] Disponível em: <https://dev.twitter.com/docs/rate-limiting/1.1> [Acedido em 13 Mar. 2014].
- [Developer.yahoo.com, n.d.] Developer.yahoo.com, (n.d.). Yahoo! GeoPlanet. [online] Disponível em: <http://developer.yahoo.com/geo/geoplanet> [Acedido em 4 Set. 2013].
- [Dijck, 2011] Dijck, J. V. (2011). Tracing Twitter: The rise of a microblogging platform. International Journal of Media & Cultural Politics, 7(3), 333-348.
- [Fleischman, 2014] Fleischman, M. (2014). The reach and impact of Oscars 2014 Tweets. [online] Twitter Blogs. Disponível em: <https://blog.twitter.com/2014/the-reach-and-impact-of-oscars-2014-tweets> [Acedido em 1 Jun. 2014].
- [Griffiths, 2013] Griffiths, S. (2013). Twitter can predict where FLU outbreaks will occur and could help doctors prepare for busy times. [online] Mail Online. Disponível em: <http://www.dailymail.co.uk/sciencetech/article-2511960/Twitter-predict-FLU-outbreaks-occur-help-doctors-prepare-busy-times.html> [Acedido em 29 Dez. 2013].

- [Guskin and Hitlin, 2012] Guskin, E., & Hitlin, P. (2012). Hurricane sandy and twitter. [online] Pew Research Journalism Project. Disponível em: <http://www.journalism.org/2012/11/06/hurricane-sandy-and-twitter/> [Acedido em 29 Dez. 2013].
- [Hockenson, 2014] Hockenson, L. (2014). eMarketer: Nearly half of Twitter's users will be in Asia by 2018. [online] Gigaom.com. Disponível em: <http://gigaom.com/2014/05/27/emarketer-nearly-half-of-twiters-users-will-be-in-asia-by-2018/> [Acedido em 2 Ago. 2014].
- [Kassim, 2012] Kassim, S. (2012). Twitter Revolution: How the Arab Spring Was Helped By Social Media. [online] Mic. Disponível em: <http://mic.com/articles/10642/twitter-revolution-how-the-arab-spring-was-helped-by-social-media> [Acedido em 28 Dez. 2013].
- [Kumar, Morstatter and Liu, 2014] Kumar, S., Morstatter, F., & Liu, H. (2014). Twitter Data Analytics. Springer New York
- [Lambrecht, Tucker and Wiertz, 2014] Lambrecht, A., Tucker, C., & Wiertz, C. (2014). Should You Target Early Trend Propagators? Evidence from Twitter. pp.1-37.
- [Manfreda, 2013] Manfreda, P. (2013). What Is the Arab Spring?. [online] About. Disponível em: <http://middleeast.about.com/od/humanrightsdemocracy/a/Definition-Of-The-Arab-Spring.htm> [Acedido em 28 Dez. 2013].
- [Oed.com, 2014] Oed.com, (2014). hash, n.3 : Oxford English Dictionary. [online] Disponível em: <http://www.oed.com/view/Entry/389023#eid301493073> [Acedido em 14 Set. 2014].
- [OxfordWords blog, 2014] OxfordWords blog, (2014). From flexitarian to evil genius: new words in the OED. [online] Disponível em: <http://blog.oxforddictionaries.com/2014/06/oed-update-june-2014/> [Acedido em 14 Set. 2014].
- [Rios, 2014] Rios, M. (2014). Penalty kicks, as seen through Twitter data. [online] Twitter Blogs. Disponível em: <https://blog.twitter.com/2014/penalty-kicks-as-seen-through-twitter-data> [Acedido em 14 Jul. 2014].
- [Rogers, 2014a] Rogers, S. (2014). The Twitter #WorldCup group stage recap. [online] Twitter Blogs. Disponível em: <https://blog.twitter.com/2014/the-twitter-worldcup-group-stage-recap> [Acedido em 14 Jul. 2014].
- [Rogers, 2014b] Rogers, S. (2014). Insights into the #WorldCup conversation on Twitter. [online] Twitter Blogs. Disponível em: <https://blog.twitter.com/2014/insights-into-the-worldcup-conversation-on-twitter> [Acedido em 15 Jul. 2014].
- [Russell, 2013] Russell, M. (2013). Mining the social web. 2nd ed. Sebastopol, CA: O'Reilly, pp.5-18.
- [Semanticweb.org, 2014] Semanticweb.org, (2014). semanticweb.org. [online] Disponível em: [http://semanticweb.org/wiki/Main\\_Page](http://semanticweb.org/wiki/Main_Page) [Acedido em 28 Jul. 2014].
- [Singletary, 2013] Singletary, T. (2013). API v1 Retirement is Complete - Use API v1.1. [online] Twitter Blogs. Disponível em: <https://blog.twitter.com/2013/api-v1-is-retired> [Acedido em 13 Mar. 2014].
- [Stampler, 2013] Stampler, L. (2013). Twitter Was Almost Called "Friendstalker" And Other Fun Facts About The Tech Giant. [online] TIME.com. Disponível em: <http://newsfeed.time.com/2013/10/09/twitter-was-almost-called-friendstalker-and-other-fun-facts-about-the-tech-giant/> [Acedido em 3 Jun. 2014].

- [Statisticbrain.com, 2014] Statisticbrain.com, (2014). Twitter Statistics. [online]  
Disponível em: <http://www.statisticbrain.com/twitter-statistics/>  
[Acedido em 5 Mar. 2014].
- [Support.twitter.com, 2014a] Support.twitter.com, (2014). Twitter Help Center - Using hashtags on Twitter. [online]  
Disponível em: <https://support.twitter.com/articles/255508-o-que-sao-os-marcadores-simbolos-de> [Acedido em 10 Mai. 2014].
- [Support.twitter.com, 2014b] Support.twitter.com, (2014). Twitter Help Center - FAQs about verified accounts. [online]  
Disponível em: <https://support.twitter.com/groups/50-welcome-to-twitter/topics/203-faqs/articles/289914-perguntas-frequentes-sobre-contas-verificadas> [Acedido em 29 Mai. 2014].
- [Support.twitter.com, 2014c] Support.twitter.com, (2014). Twitter Help Center - Twitter API limits. [online]  
Disponível em: <https://support.twitter.com/articles/160385> [Acedido em 24 Mai. 2014].
- [Support.twitter.com, 2014d] Support.twitter.com, (2014). Twitter Help Center - Twitter limits (API, updates, and following). [online]  
Disponível em: <https://support.twitter.com/articles/289867-sobre-os-limites-do-twitter-atualizacao-api-md-e-seguindo> [Acedido em 28 Dez. 2013].
- [UNOOSA, 2012] United Nations Office for Outer Space Affairs, (2012). World Geodetic System 1984. [online]  
Disponível em:  
[http://www.oosa.unvienna.org/pdf/icg/2012/template/WGS\\_84.pdf](http://www.oosa.unvienna.org/pdf/icg/2012/template/WGS_84.pdf)  
[Acedido em 15 Mai. 2014].
- [Wang, Kisling and Lam, 2013] Wang, L., Kisling, W. & Lam, E. (2013). Fake Post Erasing \$136 Billion Shows Markets Need Humans. [online] Bloomberg.  
Disponível em: <http://www.bloomberg.com/news/2013-04-23/fake-report-erasing-136-billion-shows-market-s-fragility.html> [Acedido em 1 Jun. 2014].
- [Wiki.dbpedia.org, 2014] Wiki.dbpedia.org, (2014). The DBpedia Knowledge Base. [online]  
Disponível em: <http://wiki.dbpedia.org/About> [Acedido em 31 Jul. 2014].



# Anexo A

## Tendências Repetidas

Nas seguintes tabelas são apresentadas as tendências repetidas.

Tabela A1 – Tendências repetidas em Portugal (cinco ou mais repetições)

<i>Tendência</i>	<i>Nº de Repetições</i>	<i>Primeira Utilização</i>	<i>Última Utilização</i>
#AEscolaEraBoaSe	7	25-05-2014	28-05-2014
#CarregaBenfica	13	07-05-2014	21-07-2014
#CasaisPerfeitos	6	14-05-2014	15-05-2014
#ccastingwtf	13	24-06-2014	21-07-2014
#CoisasQueOsMeusPaisDizemTodosOsDias	5	18-05-2014	19-05-2014
#DemasiadoLeiteNosCereais	5	29-05-2014	31-05-2014
#GostavaDeTirarUmaSelfieCom	7	16-05-2014	18-05-2014
#HonestHour	8	04-07-2014	06-07-2014
#LesteIstoACantar	6	21-05-2014	23-05-2014
#MelhorDoQueFalecer	7	26-05-2014	28-05-2014
#MetexxxxNumProverbio	5	21-05-2014	22-05-2014
#MTVHottest	6	19-07-2014	21-07-2014
#NasAulasEuCostumo	7	15-05-2014	17-05-2014
#ParabensBipolarCiumenta	5	13-05-2014	14-05-2014
#PortugallsReadyForWWAT	5	08-05-2014	12-07-2014
#PorUmaVidaEmQue	5	11-05-2014	13-05-2014
#QuemMeDera	5	01-07-2014	03-07-2014
#RIPoreo	5	15-05-2014	16-05-2014
#RIR_SIC	16	25-05-2014	02-06-2014
#sbsr_radical	7	17-07-2014	20-07-2014
#SeMeConhecesSabes	7	19-05-2014	21-05-2014
#SeOMundoFosseJusto	6	08-07-2014	09-07-2014
#thevoiceportugal	14	08-06-2014	21-07-2014
#TodosOsDiasEuDigo	10	06-05-2014	09-05-2014
#TwitterOff	53	03-05-2014	17-07-2014
#WMAOneDirection	5	20-05-2014	21-05-2014
Alemanha	6	16-06-2014	09-07-2014
Algarve	39	20-06-2014	21-07-2014
Argentina	5	16-06-2014	10-07-2014
Benfica	5	15-05-2014	19-05-2014
Bruna Mateus	5	03-07-2014	04-07-2014
Chile	6	14-06-2014	28-06-2014
Costa Rica	6	14-06-2014	06-07-2014
David Luiz	7	24-05-2014	05-07-2014
Di Maria	6	24-05-2014	02-07-2014
Diogo	6	02-07-2014	06-07-2014
Filipa	5	04-06-2014	08-06-2014
Holanda	10	06-05-2014	10-07-2014
Inês	72	03-05-2014	06-07-2014
Inglaterra	5	15-06-2014	20-06-2014

<i>Inglês</i>	8	20-05-2014	29-05-2014
<i>Joana</i>	32	03-05-2014	20-06-2014
<i>Justin Timberlake</i>	5	01-06-2014	03-06-2014
<i>Linkin Park</i>	6	29-05-2014	31-05-2014
<i>Lisboa</i>	68	03-05-2014	21-07-2014
<i>Mac Miller</i>	6	01-06-2014	03-06-2014
<i>Mariana</i>	56	03-05-2014	21-07-2014
<i>Marta</i>	13	13-05-2014	26-06-2014
<i>Matemática</i>	19	05-05-2014	23-06-2014
<i>Miguel Veloso</i>	5	19-05-2014	26-06-2014
<i>Passeio Marítimo de Algés</i>	6	11-07-2014	13-07-2014
<i>Paulo Bento</i>	8	13-05-2014	27-06-2014
<i>Pedro Teixeira</i>	8	04-05-2014	07-07-2014
<i>Phineas e Ferb</i>	6	03-05-2014	24-05-2014
<i>Portimão</i>	43	07-06-2014	21-07-2014
<i>Porto</i>	5	23-06-2014	13-07-2014
<i>Portugal</i>	20	07-05-2014	17-07-2014
<i>Português</i>	9	06-05-2014	18-06-2014
<i>Psicopata Anónimo</i>	10	10-05-2014	10-06-2014
<i>Raquel</i>	55	03-05-2014	19-07-2014
<i>Renata</i>	5	12-05-2014	30-06-2014
<i>Rising Star</i>	13	04-05-2014	21-07-2014
<i>Rita</i>	36	03-05-2014	08-07-2014
<i>Rock In Rio</i>	21	08-05-2014	01-06-2014
<i>Rolling Stones</i>	8	27-05-2014	30-05-2014
<i>Rui Reininho</i>	5	30-06-2014	15-07-2014
<i>Rute</i>	6	22-05-2014	20-06-2014
<i>S.João</i>	14	19-06-2014	25-06-2014
<i>S.Pedro</i>	10	27-06-2014	30-06-2014
<i>Santarém</i>	11	29-05-2014	07-07-2014
<i>Sara Sampaio</i>	11	16-05-2014	10-07-2014
<i>Sebastianas</i>	5	10-07-2014	13-07-2014
<i>Seixal</i>	14	25-06-2014	03-07-2014
<i>Ser Benfiquista</i>	5	13-05-2014	15-05-2014
<i>Sesimbra</i>	23	10-05-2014	16-07-2014
<i>Setembro</i>	7	08-05-2014	16-07-2014
<i>Sexta-feira 13</i>	5	13-06-2014	14-06-2014
<i>Sofia</i>	16	06-05-2014	22-06-2014
<i>Somos Portugal</i>	10	11-05-2014	21-06-2014
<i>Sporting</i>	9	03-05-2014	20-07-2014
<i>Steve Aoki</i>	6	30-05-2014	14-07-2014
<i>Sumol</i>	18	02-06-2014	29-06-2014
<i>Sunset</i>	5	10-07-2014	13-07-2014
<i>Teen Wolf</i>	5	20-05-2014	24-06-2014
<i>Tenho 2</i>	7	08-05-2014	05-06-2014
<i>Tenho 3</i>	14	11-05-2014	06-06-2014
<i>The Originals</i>	5	09-05-2014	17-05-2014
<i>The Vampire Diaries</i>	6	03-05-2014	17-05-2014
<i>The Voice</i>	11	04-05-2014	21-07-2014
<i>Tiago</i>	7	22-05-2014	26-06-2014
<i>Tiago Bandeira</i>	10	04-05-2014	02-07-2014
<i>Tires</i>	10	07-06-2014	10-06-2014



Tony Carreira	9	14-05-2014	22-06-2014
Tunepics	6	14-07-2014	16-07-2014
Uruguai	5	14-06-2014	24-06-2014
Vilamoura	14	28-06-2014	16-07-2014
XDXD	26	07-05-2014	21-07-2014

Tabela A2 – Tendências repetidas em Espanha (cinco ou mais repetições)

<i>Tendência</i>	<i>Nº de Repetições</i>	<i>Primeira Utilização</i>	<i>Última Utilização</i>
#BuenosDías	5	12-05-2014	21-07-2014
#FelizJueves	6	22-05-2014	17-07-2014
#FelizLunes	8	05-05-2014	07-07-2014
#sorteodominguero	5	01-06-2014	13-07-2014
#TourTVE	6	05-07-2014	20-07-2014
Barcelona	5	16-05-2014	08-07-2014
España	23	05-05-2014	21-07-2014
Felipe VI	6	02-06-2014	18-06-2014
Kiko Rivera	5	14-07-2014	17-07-2014
Luis Enrique	5	05-05-2014	21-05-2014
Madrid	28	17-06-2014	21-07-2014
Noel	5	08-05-2014	27-06-2014
PSOE	6	05-05-2014	14-07-2014
San Juan	5	20-06-2014	23-06-2014
Sevilla	5	14-05-2014	03-07-2014
TFIOS	5	06-06-2014	04-07-2014

Tabela A3 – Tendências repetidas nos Estados Unidos (cinco ou mais repetições)

<i>Tendência</i>	<i>Nº de Repetições</i>	<i>Primeira Utilização</i>	<i>Última Utilização</i>
#ASCAfterShow	5	14-05-2014	25-06-2014
#ASCQuiz	6	14-05-2014	16-07-2014
#cantsleep	23	05-05-2014	15-07-2014
#FinallyA5sosAlbum	5	13-05-2014	26-06-2014
#insomnia	7	09-05-2014	09-07-2014
#oomf	9	03-05-2014	14-07-2014
#SamuraiJack	6	04-05-2014	06-07-2014
#teamnosleep	11	03-05-2014	24-06-2014
Cali	29	03-05-2014	21-07-2014
FaceTime	38	06-05-2014	21-07-2014
George Lopez	9	11-06-2014	08-07-2014
Godzilla	7	16-05-2014	19-05-2014
IHOP	27	04-05-2014	08-07-2014
Jacob Whitesides	5	20-06-2014	16-07-2014
Netflix	63	03-05-2014	21-07-2014
NyQuil	19	05-05-2014	03-07-2014
Starbucks	18	12-05-2014	21-07-2014
Taco Bell	11	08-05-2014	09-07-2014
Waffle House	42	03-05-2014	20-07-2014

Tabela A4 – Tendências repetidas no Brasil (cinco ou mais repetições)

<i>Tendência</i>	<i>Nº de Repetições</i>	<i>Primeira Utilização</i>	<i>Última Utilização</i>
<i>Carol</i>	8	19-06-2014	21-07-2014
<i>Disney</i>	6	23-06-2014	19-07-2014
<i>Gabriel</i>	11	19-06-2014	18-07-2014
<i>Jorge e Mateus</i>	8	19-06-2014	21-07-2014
<i>Lucas</i>	24	16-06-2014	19-07-2014
<i>Matheus</i>	6	20-06-2014	18-07-2014
<i>Paula Fernandes</i>	5	18-06-2014	29-06-2014

Tabela A5 – Tendências repetidas na Argentina (cinco ou mais repetições)

<i>Tendência</i>	<i>Nº de Repetições</i>	<i>Primeira Utilização</i>	<i>Última Utilização</i>
<i>Agus</i>	21	18-06-2014	21-07-2014
<i>Bariloche</i>	14	18-06-2014	19-07-2014
<i>Disney</i>	27	24-06-2014	19-07-2014
<i>Felices 15</i>	13	17-06-2014	19-07-2014
<i>Las 4</i>	7	22-06-2014	14-07-2014
<i>Las 5</i>	9	21-06-2014	21-07-2014
<i>Lucas</i>	12	18-06-2014	18-07-2014
<i>Nico</i>	20	17-06-2014	21-07-2014

Tabela A6 – Tendências repetidas no Reino Unido (cinco ou mais repetições)

<i>Tendência</i>	<i>Nº de Repetições</i>	<i>Primeira Utilização</i>	<i>Última Utilização</i>
<i>#tracerequest</i>	24	16-06-2014	21-07-2014
<i>Jeremy Kyle</i>	31	17-06-2014	18-07-2014
<i>London</i>	15	22-06-2014	21-07-2014

Tabela A7 – Tendências repetidas Mundial (cinco ou mais repetições)

<i>Tendência</i>	<i>Nº de Repetições</i>	<i>Primeira Utilização</i>	<i>Última Utilização</i>
<i>Happy Thanksgiving</i>	7	23-06-2014	13-07-2014
<i>XXXXXX</i>	20	16-06-2014	20-07-2014

# Anexo B

## Hashtags Repetidas

Nas seguintes tabelas são apresentadas as *hashtags* repetidas.

Tabela B1 – *Hashtags* repetidas em Portugal (três ou mais repetições)

<i>Hashtags</i>	<i>Nº de Repetições</i>	<i>Primeira Utilização</i>	<i>Última Utilização</i>
#1YearSince1DInPortugal	4	26-05-2014	27-05-2014
#5sosRockOutWithYourSocksOutTour	3	07-07-2014	08-07-2014
#5sosupindisstream	3	24-06-2014	01-07-2014
#AEscolaEraBoaSe	7	25-05-2014	28-05-2014
#AMelhorMúsicaDoMomentoÉ	3	24-05-2014	25-05-2014
#Aos25AnosQuero	4	03-05-2014	04-05-2014
#BandasQueValemAPenaOuvir	3	04-05-2014	05-05-2014
#BRAvsCOL	3	04-07-2014	05-07-2014
#BrazilvsGermany	3	08-07-2014	09-07-2014
#CamsBookClub	3	09-07-2014	16-07-2014
#CarregaBenfica	13	07-05-2014	21-07-2014
#CasaisPerfeitos	6	14-05-2014	15-05-2014
#cccastingwtf	13	24-06-2014	21-07-2014
#CoisasQueFaltamNoMeuQuarto	3	09-05-2014	10-05-2014
#CoisasQueOsMeusPaisDizemTodosOsDias	5	18-05-2014	19-05-2014
#DemasiadoLeiteNosCereais	5	29-05-2014	31-05-2014
#DONTSTOPLYRICVIDEO	3	05-05-2014	06-05-2014
#EsteVerãoQuero	3	23-05-2014	24-05-2014
#EuNasAulas	3	09-05-2014	10-05-2014
#EurovisionSongContest2014	3	06-05-2014	07-05-2014
#FinallyA5sosAlbum	3	13-05-2014	14-05-2014
#GostavaDeTirarUmaSelfieCom	7	16-05-2014	18-05-2014
#GostoMuitoDoVerãoPorque	3	13-05-2014	14-05-2014
#HappyBirthdayAshtonIrwin	3	07-07-2014	07-07-2014
#HappyBirthdayTheo	3	16-07-2014	17-07-2014
#heatJustinBieber	3	16-05-2014	17-05-2014
#HiCam	3	07-06-2014	07-06-2014
#HonestHour	8	04-07-2014	06-07-2014
#IMNEVERLEAVINGTHEONEDIRECTIONFAMILY	4	19-05-2014	20-05-2014
#justintimberlake	3	02-06-2014	03-06-2014
#LesteIstoACantar	6	21-05-2014	23-05-2014
#LiamAppreciationDay	3	20-07-2014	20-07-2014
#LouisAppreciationDay	3	18-07-2014	19-07-2014
#MelhorDoQueFalecer	7	26-05-2014	28-05-2014
#MelhoresMensagensParaReceber	3	24-05-2014	24-05-2014
#MeteAPalavraXxxxNoTituloDeUmLivro	3	19-06-2014	20-06-2014
#MeteXxxxNumProverbio	5	21-05-2014	22-05-2014
#MTV5SOS	3	03-07-2014	04-07-2014
#MTVHottest	6	19-07-2014	21-07-2014
#NasAulasEuCostumo	7	15-05-2014	17-05-2014

#NestasFériasEuVou	3	06-06-2014	07-06-2014
#NiallAppreciationDay	3	21-07-2014	21-07-2014
#NoConcertoDosOneDirectionEuVou	3	09-07-2014	09-07-2014
#NuncaÉBomSinalQuando	4	24-06-2014	25-06-2014
#OqueMaisCustaNaVidaDeUmAdolescenteÉ	3	03-05-2014	04-05-2014
#OsMeusPlanosParaOverãoSão	4	06-05-2014	07-05-2014
#ParabensBipolarCiumenta	5	13-05-2014	14-05-2014
#PAYZERISBACK	3	27-05-2014	28-05-2014
#PerguntasEstupidas	3	21-06-2014	22-06-2014
#PortugallsReadyForWWAT	5	08-05-2014	12-07-2014
#PortugalNeeds5SOS	3	11-06-2014	12-06-2014
#PortugalWantsOnTheRoadAgainTour	3	19-05-2014	20-05-2014
#PorUmaVidaEmQue	5	11-05-2014	13-05-2014
#PrecisoDeAlguémQue	3	12-05-2014	13-05-2014
#PretoQueÉPreto	3	16-07-2014	16-07-2014
#QuandoEraCriançaGostava	3	06-05-2014	07-05-2014
#QuemMeDera	5	01-07-2014	03-07-2014
#RazõesParaGostarDasFérias	4	12-05-2014	13-05-2014
#RIPOréo	5	15-05-2014	16-05-2014
#RIR_SIC	16	25-05-2014	02-06-2014
#risingstartvi	3	05-05-2014	09-06-2014
#sbsr_radical	7	17-07-2014	20-07-2014
#SeMeConhecesSabes	7	19-05-2014	21-05-2014
#SeOMundoFosseJusto	6	08-07-2014	09-07-2014
#SeTensMelhorAmiga	4	27-06-2014	28-06-2014
#songsfromthecouch	3	02-06-2014	03-06-2014
#TeamRecuperaçãoDoTwitter	3	15-06-2014	16-06-2014
#thevoice	4	05-05-2014	22-06-2014
#thevoiceportugal	14	08-06-2014	21-07-2014
#TodosOsDiasEuDigo	10	06-05-2014	09-05-2014
#TorresmoChato	3	27-05-2014	27-05-2014
#turnodanoite	4	19-06-2014	20-06-2014
#TwitterOff	53	03-05-2014	17-07-2014
#UmaLataDeCocaCola	4	06-07-2014	07-07-2014
#UmDosMeusViciosÉ	4	03-05-2014	04-05-2014
#WelcomeToPortugal1D	3	12-07-2014	12-07-2014
#WMAOneDirection	5	20-05-2014	21-05-2014
#WorldCup2014	4	12-06-2014	13-06-2014
#WWAQUESTIONS	3	01-07-2014	02-07-2014
#ZaynAppreciationDay	4	19-07-2014	20-07-2014

Tabela B2 – Hashtags repetidas em Espanha (três ou mais repetições)

Hashtags	Nº de Repetições	Primeira Utilização	Última Utilização
#BuenosDías	5	12-05-2014	21-07-2014
#CosasQueAmo	4	04-05-2014	05-05-2014
#CosasQueAmoMuchoComo	3	30-05-2014	31-05-2014
#ElReyAbdica	3	02-06-2014	02-06-2014
#FelizDomingo	4	15-06-2014	20-07-2014
#FelizJueves	6	22-05-2014	17-07-2014
#FelizLunes	8	05-05-2014	07-07-2014

#FelizMartes	3	20-05-2014	15-07-2014
#HalaMadridAPorLADecima	3	24-05-2014	24-05-2014
#HomoCañetus	3	16-05-2014	17-05-2014
#HoyNosAmordazan	3	11-07-2014	11-07-2014
#InventaTuMultas	4	10-06-2014	11-06-2014
#JunioConJDe	3	01-06-2014	02-06-2014
#larojasipuede	4	14-06-2014	18-06-2014
#MTVHottest	4	19-07-2014	21-07-2014
#party	4	01-06-2014	20-07-2014
#Podemos25M	3	25-05-2014	26-05-2014
#QuieroApretarUnBotonY	3	28-05-2014	29-05-2014
#QuieroUnaSelfieCon	4	16-06-2014	17-06-2014
#SiElFinDelMundoFueraHoy	4	23-05-2014	24-05-2014
#sorteodominguero	5	01-06-2014	13-07-2014
#TATGranada14	3	05-06-2014	06-06-2014
#TodosSomosCarlos	3	15-07-2014	16-07-2014
#TourTVE	6	05-07-2014	20-07-2014
#TuiteaParaEvitarElTalego	3	14-05-2014	15-05-2014
#VamosRafa	3	09-05-2014	08-06-2014
#YoPidoReferendum	3	07-06-2014	08-06-2014

Tabela B3 – Hashtags repetidas nos Estados Unidos (três ou mais repetições)

Hashtags	Nº de Repetições	Primeira Utilização	Última Utilização
#50FactsAboutMe	3	01-06-2014	02-06-2014
#ASCAfterShow	5	14-05-2014	25-06-2014
#ASCQuiz	6	14-05-2014	16-07-2014
#BadPrequels	3	07-05-2014	07-05-2014
#BattleOfTheFandoms	3	01-06-2014	02-06-2014
#BeforeMeloDecides	3	08-07-2014	08-07-2014
#cantsleep	23	05-05-2014	15-07-2014
#DeadWhiteandBlue	4	04-07-2014	06-07-2014
#FinallyA5sosAlbum	5	13-05-2014	26-06-2014
#fullmoon	3	13-06-2014	14-06-2014
#Happy21stBirthdayAriana	3	26-06-2014	26-06-2014
#HappyBirthdayJazmynBieber	3	30-05-2014	30-05-2014
#HappyBirthdayLanaDelRey	3	21-06-2014	21-06-2014
#happymonday	3	02-06-2014	21-07-2014
#heatJustinBieber	4	16-05-2014	17-05-2014
#HonestyHour	3	19-05-2014	01-07-2014
#insomnia	7	09-05-2014	09-07-2014
#jacobyoudi	3	06-06-2014	06-06-2014
#JayZSolange	3	12-05-2014	13-05-2014
#Letsblowuptwitterwithfetus1D	3	27-06-2014	27-06-2014
#MakeASongDepressing	3	23-05-2014	23-05-2014
#MayaAngelou	3	28-05-2014	29-05-2014
#MTVHottest	3	19-07-2014	20-07-2014
#nationalbestfriendday	3	08-06-2014	09-06-2014
#oomf	9	03-05-2014	14-07-2014
#ReplaceMovieTitlesWithThot	4	20-05-2014	21-05-2014
#RuinARapTrack	3	06-05-2014	06-05-2014

#SamuraiJack	6	04-05-2014	06-07-2014
#teamnosleep	11	03-05-2014	24-06-2014
#TeenWolfMarathon	3	23-06-2014	23-06-2014
#tfiostix	3	06-06-2014	06-06-2014
#ThingsGirlsSayDuringSex	3	26-05-2014	26-05-2014
#throwbackthursday	3	22-05-2014	03-07-2014
#UndateableBecause	3	21-05-2014	21-05-2014
#WeBeliveInYouKris	3	15-05-2014	15-05-2014
#WeLoveYouTaylor	3	23-06-2014	23-06-2014
#WhatJayZSaidToSolange	3	12-05-2014	13-05-2014
#WMAJustinBieber	3	20-05-2014	21-05-2014
#WMAOneDirection	3	20-05-2014	21-05-2014
#YouAintBlackIf	3	18-06-2014	18-06-2014
#YouWereCuteUntil	3	27-05-2014	27-05-2014

Tabela B4 – Hashtags repetidas no Brasil (três ou mais repetições)

Hashtags	Nº de Repetições	Primeira Utilização	Última Utilização
#2YearsOfBelieveAlbum	3	19-06-2014	19-06-2014
#5YearsOfOneTime	3	07-07-2014	07-07-2014
#askharmonizer	3	10-07-2014	17-07-2014
#Cite30CoisasQueVcAma	3	06-07-2014	06-07-2014
#encontro	3	25-06-2014	17-07-2014
#EuDesejoMeuXxxxSePara	3	14-07-2014	14-07-2014
#FelizDiaDoAmigo	4	20-07-2014	21-07-2014
#FifaOCamisa18daColombiaMereceSerPunido	3	05-07-2014	05-07-2014
#HappyBDayLaurenFromBrazil	3	27-06-2014	27-06-2014
#HappyBirthdayArianaFromBrazil	3	26-06-2014	26-06-2014
#HappyBirthdayDinahJaneFromBrazil	3	22-06-2014	22-06-2014
#HappyBirthdayLanaDelRey	3	21-06-2014	21-06-2014
#HappyBirthdayTheoHoranFromBrazil	3	16-07-2014	16-07-2014
#HarryDontLickAnything	3	07-07-2014	07-07-2014
#jackandjackdoingitright	3	01-07-2014	01-07-2014
#JogoDoSDVComJubs	3	06-07-2014	06-07-2014
#JustinDaUmVestidoParaTatiNeves	3	25-06-2014	25-06-2014
#lesbicafacts	3	02-07-2014	02-07-2014
#MadrugaSporTV	4	22-06-2014	01-07-2014
#NeymarHomenageiaoMCDaleste	4	27-06-2014	28-06-2014
#QueroSeguidorescomValentinoDasTag	3	03-07-2014	03-07-2014
#QuintaDoSegueSDVcomValentinoDasTag	3	17-07-2014	17-07-2014
#ReallyDontCareVideoTOMORROW	3	25-06-2014	25-06-2014
#SDVComLnstagrada	4	14-07-2014	15-07-2014
#SDVComOCafajeste	3	01-07-2014	02-07-2014
#SegueDeVoltaComODrogaSam	4	16-07-2014	17-07-2014
#SextaDeSeguidoresComValentino	3	20-06-2014	20-06-2014
#TeenAwardsBrasil2014	3	18-06-2014	18-06-2014
#ToAquiNoTwitterPorque	3	02-07-2014	02-07-2014
#Watch5HBOSSonVEVO	3	08-07-2014	09-07-2014

Tabela B5 – Hashtags repetidas na Argentina (três ou mais repetições)

Hashtags	Nº de Repetições	Primeira Utilização	Última Utilização
#20FrasasDeCallejeros	4	24-06-2014	19-07-2014
#AlGenioDeLaLamparaLePido	3	04-07-2014	04-07-2014
#AlViernesNoLePuedeFaltar	3	18-07-2014	18-07-2014
#AmigoEsAquel	4	20-07-2014	21-07-2014
#AMiLadoSiempreQuiero	3	03-07-2014	03-07-2014
#AMiMeGustaDormir	4	17-07-2014	18-07-2014
#AngelaTorresDedicateALavarBaños	3	17-07-2014	17-07-2014
#AprendiQueAveces	3	19-06-2014	19-06-2014
#ArgentinaCampeonBrasil2014	3	13-07-2014	13-07-2014
#ArgentinaHoyGanamosCarajo	3	05-07-2014	05-07-2014
#BrasilDecimeQueSe7	3	09-07-2014	09-07-2014
#CruzaLosDedos	3	02-07-2014	12-07-2014
#CuandoTermineElMundialYo	3	12-07-2014	12-07-2014
#DejoTodoPor	3	26-06-2014	26-06-2014
#DejoTodoSi	3	16-07-2014	16-07-2014
#DigoBoludecesCuando	3	28-06-2014	28-06-2014
#EnEsteMomentoQuiero	3	20-06-2014	21-06-2014
#EnVacacionesVoyA	3	04-07-2014	04-07-2014
#EsDeBuenAmigo	4	19-07-2014	20-07-2014
#esperandolafinal	3	12-07-2014	13-07-2014
#EstaCopaEsDeArgentina	3	13-07-2014	13-07-2014
#FandomParaLaSeleccion	4	15-07-2014	16-07-2014
#FelizCumplePatoFontanet	3	16-06-2014	16-06-2014
#FelizDiaDeLaBandera	3	20-06-2014	20-06-2014
#FelizDiaDeLaIndependencia	3	09-07-2014	09-07-2014
#FelizDiaDelAmigo	3	20-07-2014	20-07-2014
#FuerzaTiti	3	02-07-2014	03-07-2014
#HoyJugamosTodos	3	09-07-2014	09-07-2014
#JulioConJDe	4	01-07-2014	02-07-2014
#LoQueMasOdioDeLunes	3	07-07-2014	07-07-2014
#maschefacts	3	10-07-2014	10-07-2014
#MeCambiaElHumorCuando	3	27-06-2014	27-06-2014
#MeDaAsco	3	03-07-2014	04-07-2014
#MeDanGanasDe	4	17-06-2014	18-06-2014
#MeDejoDelImportar	4	08-07-2014	09-07-2014
#MeHaceFalta	3	23-06-2014	23-06-2014
#MePongoCortante	3	02-07-2014	02-07-2014
#MeReContraEnamoraQue	3	23-06-2014	23-06-2014
#MeReContraPuedeQue	4	27-06-2014	28-06-2014
#MiEspecialidad	3	11-07-2014	12-07-2014
#MiPromesaPorArgentina	3	10-07-2014	10-07-2014
#MovimientoParaQueLasLeonasJueguenEnTanga	3	30-06-2014	30-06-2014
#NadaMasFeoQue	4	22-06-2014	23-06-2014
#NoMeFumo	3	17-06-2014	18-06-2014
#NuncaPeroNunca	3	21-07-2014	21-07-2014

#PiroposParaLavezzi	4	29-06-2014	30-06-2014
#QuieroUnaSelfieCon	4	16-06-2014	17-06-2014
#ReDaPara	4	26-06-2014	27-06-2014
#RihannaRomeroNoTeDioBola	3	14-07-2014	14-07-2014
#RohayhuParaguay	3	15-07-2014	15-07-2014
#SeriaFelizCon	4	02-07-2014	03-07-2014
#SiArgentinaGana	3	07-07-2014	08-07-2014
#SiMeDasBolaYo	3	18-06-2014	18-06-2014
#SiPudieraMeCompraria	4	20-06-2014	21-06-2014
#SiTuvieraUnHadaMadrina	4	06-07-2014	07-07-2014
#TeChapoSi	4	17-07-2014	18-07-2014
#TipicasExcusasMias	3	18-06-2014	18-06-2014
#TiremosPiroposALoArgentino	3	22-06-2014	22-06-2014
#TraumasDeLaInfancia	4	03-07-2014	04-07-2014
#TwOff	3	27-06-2014	11-07-2014
#UnaMinitahQuiere	3	23-06-2014	23-06-2014
#UnOdioEternoA	4	07-07-2014	08-07-2014
#UnVerdaderoAmigoEs	4	20-07-2014	21-07-2014
#YaQuiero	3	28-06-2014	29-06-2014

Tabela B6 – Hashtags repetidas no Reino Unido (três ou mais repetições)

Hashtags	Nº de Repetições	Primeira Utilização	Última Utilização
#5sosRockOutWithOurSocksOutTour	3	11-07-2014	11-07-2014
#BrazilvsGermany	3	08-07-2014	09-07-2014
#MTVHottest	3	19-07-2014	20-07-2014
#RemoveALetterRuinABand	4	19-07-2014	20-07-2014
#tracerequest	24	16-06-2014	21-07-2014
#Wimbledon2014	4	23-06-2014	02-07-2014

Tabela B7 – Hashtags repetidas Mundial (três ou mais repetições)

Hashtags	Nº de Repetições	Primeira Utilização	Última Utilização
#AkhirnyaMilihJokowi	3	03-07-2014	04-07-2014
#BOYKOTVarSesVerTürkiye	3	21-07-2014	21-07-2014