

Miguel Santiago Lopes

Towards the characterization and prediction of
trends and collective dynamics in Instagram

Orientador: Manuel Marques Pita

Universidade Lusófona de Humanidade e Tecnologias

Escola de Comunicação, Arquitectura, Artes e Tecnologias da Informação

Lisboa, 10 de Abril de 2014

Agradecimentos

Este espaço é dedicado a todos aqueles que deram a sua contribuição para que este trabalho fosse realizado. A todos eles deixo aqui o meu agradecimento sincero.

Em primeiro lugar agradeço ao Professor Doutor Manuel Marques Pita a forma como motivou e orientou o meu trabalho. A receptividade sempre presente para discutir temas e para se adaptar aos vários aspectos que iam aparecendo ao decorrer da construção do projecto. Estou grato pela liberdade de acção que me permitiu, que foi decisiva para que este trabalho contribuísse para o meu desenvolvimento pessoal.

Em segundo lugar, gostaria de agradecer à Dra. Maria do Rosário Duarte, pela ajuda e disponibilidade que demonstrou aquando da construção deste relatório.

Gostaria ainda de agradecer a todos os professores da Licenciatura de Engenharia Informática por me terem ajudado a crescer e por sempre terem a paciência de falar e tirar quaisquer dúvidas que fossem aparecendo.

Agradecimento especial ao Professor Pedro Freire, por me ter ajudado quando mais precisava, ao Professor Doutor José Rogado pela motivação e constante preocupação pelo estado do projecto, à Professora Alexandra Campos por sempre estar disponível para falar e tirar qualquer dúvida, ao Professor Mário Cardoso por ter apostado em mim e me ter deixado à vontade para falar e pensar e por fim ao Professor Doutor Sérgio Guerreiro por sempre estar disponível.

Aos meus pais por sempre me apoiarem nos meus projectos, mesmo que por vezes não concordem ou não compreendam, mas sempre ao meu lado.

Por fim, o mais importante, à minha namorada Isabel Carvalho, que está permanentemente comigo, caminha comigo e sempre me motiva, apoia e ajuda, sem ela este trabalho não tinha sido possível e sem ela não seria a pessoa que sou.

A todos um muito obrigado.



Índice

| | |
|---|----|
| Agradecimentos..... | 2 |
| Índice..... | 3 |
| Índice de Figuras | 4 |
| Índice de Tabelas..... | 4 |
| I. Resumo..... | 5 |
| II. Abstract | 6 |
| III. Introdução | 7 |
| IV. Enquadramento Teórico | 10 |
| 1. Internet | 10 |
| 1.1. Políticas sobre a Internet | 12 |
| 2. Redes Sociais | 13 |
| 2.1. Instagram..... | 14 |
| 2.2. API do Instangram | 15 |
| 3. PHP | 15 |
| V. Métodos..... | 16 |
| 1. Tipo de estudo | 16 |
| 2. Procedimentos | 16 |
| VI. Resultados | 20 |
| 1. API access | 20 |
| 2. Base de Dados | 22 |
| 3. Mapa..... | 24 |
| VIII. Conclusões e trabalho futuro..... | 27 |
| IX. Bibliografia | 28 |
| X. Anexos..... | 30 |
| 1. Manual Técnico da Aplicação..... | 31 |
| XI. Glossário | 35 |
| XII. Índice Remissivo | 36 |

Índice de Figuras

| | |
|---|----|
| Figura 1 - Acesso à Internet no Mundo..... | 7 |
| Figura 2 - Crescimento de utilizadores da Internet entre 1995-2008 | 10 |
| Figura 3 - Estatísticas da utilização da Internet entre 2000-2012 | 10 |
| Figura 4 - Distribuição de pessoas que utilizam a Internet por país em 2009 | 11 |
| Figura 5 - Graus de censura da Internet no Mundo..... | 12 |
| Figura 6 - API do Instagram..... | 15 |
| Figura 7 - Esquema inicial da base de dados..... | 17 |
| Figura 8 - Esquema de relação entre entidades da base de dados | 18 |
| Figura 9 - Alteração de vários posts na base de dados..... | 20 |
| Figura 10 - Imagens obtidas da API do Instagram | 21 |
| Figura 11 - Alteração de valores em um post da página popular | 23 |
| Figura 12 - Mapa com localização dos posts da página popular..... | 24 |
| Figura 13 - Número de posts na página popular por país/hora..... | 26 |

Índice de Tabelas

| | |
|--|----|
| Tabela 1 - Campos da base de dados..... | 22 |
| Tabela 2 - Números reais da base de dados..... | 23 |
| Tabela 3 - Quantis de Gostos, Comentários e Seguidores..... | 25 |
| Tabela 4 - 10 países mais significativos na página popular | 25 |

I. Resumo

Nos últimos três anos o Instagram cresceu de uma pequena rede de pessoas que partilham fotos na área geográfica de Stanford na Califórnia (EUA), para uma das redes sociais mais activa no mundo, com mais de 200 milhões de utilizadores activos. De forma inédita, a comunicação visual é cada vez mais a linguagem escolhida para divulgar informações sobre estilos de vida, marcas, alimentação, eventos sócio-políticos e outras. A rede de interações num determinado momento, numa rede social como o Instagram tem uma complexidade aparentemente insuperável. Neste trabalho vamos concentrar-nos na implementação dos mecanismos computacionais que nos permitem tirar "snapshots" desta dinâmica, e caracterizar as suas propriedades básicas. Este projecto concentra-se no (1) acesso; (2) na estruturação e (3) no armazenamento de dados dinâmicos do Instagram. Vamos ilustrar o uso de tais dados, concentrando-nos na recolha de informações sobre a página 'explorar': uma parte da aplicação do Instagram que apresenta conteúdo "popular".

Análises de pequenas amostras de dados permitiram-nos supor que, após o post entrar na página popular, a condição socioeconómica pode ser correlacionada em alguns casos. No entanto, parece ser mais extensamente determinado, pela proporção de pessoas de um determinado país que utiliza o aplicativo, onde os EUA e a Tailândia apresentam a taxa mais alta. Entre outras observações, iremos também analisar "valores atípicos" - eventos especiais que acontecem na página popular, que podem ser de interesse para os antropólogos ou especialistas em marketing, por exemplo.

Palavras-Chave: Instagram, engagement, redes sociais

II. Abstract

Over the last three years Instagram has grown from a small network of people sharing photos around the Stanford area of California (USA) to the most active social network in the world today, with more than 200M active users. In an unprecedented manner, visual communication is becoming more and more the chosen language to disseminate information about lifestyles, branding, food, socio-political events and so on. The network of interactions at any given moment in a network such as Instagram has a seemingly insurmountable complexity. In this work we focus on implementing the computational mechanisms that enable us to take 'snapshots' of these dynamics, and characterize its basic properties. This work concentrates on (1) accessing; (2) structuring and (3) storing instagram dynamic data. We illustrate the uses of such data by focusing on gathering information about the 'explore' page: a portion of the instagram app that features "popular" content. Analysis of small data samples allowed us to hypothesize that post entering the popular page may be correlated to the socio-economic status of posters in some cases. Yet it seems to be most largely determined by the proportion of people in a given country using the app, where the USA and Thailand fare highest. Amongst other observations, we will analyse 'outliers' - special events happening in the popular page that may be of interest to e.g. anthropologists or marketing experts.

Key words: Instagram, engagement, social networking

III. Introdução

Com o aumento da população mundial entre 2005 e 2013 de 6,5 mil milhões para 7,1 mil milhões, a percentagem de pessoas que utilizam a internet aumentou 23%. A variação de utilização nos países em desenvolvimento, no mesmo período de tempo, é de 23%, sendo que nos países desenvolvidos o crescimento é de 26% (Figura 1). É de notar que os dados de 2013 são uma estimativa aproximada.[1]

| | 2005 | 2010 | 2013 ^a |
|---------------------------------|-------------|-------------|-------------------|
| World population ^[3] | 6.5 billion | 6.9 billion | 7.1 billion |
| Not using the Internet | 84% | 70% | 61% |
| Using the Internet | 16% | 30% | 39% |
| Users in the developing world | 8% | 21% | 31% |
| Users in the developed world | 51% | 67% | 77% |

^a Estimate.

Figura 1 - Acesso à Internet no Mundo

Normalmente a investigação nesta área tem-se concentrado na estrutura de ligações em redes, embora isso seja útil para identificar as suas características, muitas vezes não consegue explicar a sua dinâmica e como controlá-la. Só esta estrutura, tem pouco valor de previsão para o efeito de interações específicas, que apesar de terem um baixo índice de conectividade, são dinâmicas e contextualmente uma força motriz numa rede.[2]

Uma interacção no Instagram contém normalmente (1) uma foto, (2) uma descrição; (3) hashtags; (4) menções a outras pessoas e (5) comentários, tudo em uma sequência através de uma linha de tempo. Com esta informação, é possível construir uma rede da dinâmica de um qualquer utilizador, ou algum post específico, para determinar quais os aspectos chave para o sucesso ou fracasso da informação que foi partilhada em termos de número de gostos e comentários.

O Instagram é uma rede social que está a atrair muita atenção, devido aos níveis de *engagement* proporcionado pelo número de utilizadores activos em qualquer período de tempo.[3]

Esta rede social foi fundada em Outubro de 2010, sendo que dois meses depois apresentava no seu espólio dois milhões de utilizadores activos. Oito meses depois deste acontecimento cerca de cento e cinquenta milhões de fotos tinham sido partilhadas. Em

Setembro de 2013 o número de membros activos chegou aos cento e cinquenta milhões, fazendo com que esta rede seja mais procurada do que por exemplo o Google +. Porém o efeito que estas redes têm nos indivíduos e na sociedade permanece desconhecido.[3], [4]

Após a descrição do contexto deste projecto, surge o problema a ser estudado. A problemática que advém do que foi pesquisado é a tentativa de percepção do comportamento de um post/utilizador após entrada na página popular do Instagram e como os seus dados (que lhes estão associados) se alteram por terem chegado a esta página de popularidade. Será interessante também perceber através de geolocalização onde se encontram estes posts colocados no mundo.

Em termos do trabalho a ser realizado vão-se recolher dados dinâmicos das redes sociais através da API fornecida pelo Instagram. E, de seguida implementar algoritmos para recolha automática de dados, com foco na busca de informações úteis sobre a associação entre de um post e sua dinâmica.

Este projecto proposto pelo orientador, suscitou desde logo o nosso interesse, na medida em que existem poucos estudos científicos sobre esta matéria e de tornar possível a contribuição para a percepção de quais são as condições ideais para obter o maior número de seguidores, gostos e comentário de maneira a singrar na rede social objecto escolhida.

Tendo em conta todo o contexto apresentado acima os objectivos deste projecto são entender a localização dos postes do Instagram num período de tempo. Sendo possível especificar este objectivo através da análise da API de desenvolvimento do Instagram, da recolha dados fornecidos pela API, da criação de uma base de dados com os dados fornecidos pela API e do desenvolvimento de um mapa de geolocalização sobre a informação colocada na base de dados.

Constituíram-se como limitações no curso deste projecto a limitada experiência com a linguagem de programação PHP (Hypertext Preprocessor originalmente Personal Home Page), as falhas da API do Instagram, problemas no Apache /Wamp com a introdução de excepções devido aos pedidos HTTPS e o conhecimento limitado de queries MYSQL.

O presente projecto divide-se em introdução, enquadramento teórico, métodos, resultados e conclusão e trabalho futuro. Na introdução, tendo em conta o contexto em que o projecto se insere serão abordados alguns temas genéricos, a identificação do problema, qual a motivação que levou ao desenvolvimento deste projecto e por fim, a divisão que foi realizada para apresentação do trabalho.

No enquadramento teórico será feita uma pequena abordagem à Internet, onde irão ser abordados assuntos como o que é, os utilizadores da Internet, a censura da mesma, a diferença de utilização entre países desenvolvidos e em vias de desenvolvimento. Será também abordado o tema das redes sociais analisando-se entre outras as questões da liberdade de expressão, o potencial destas redes e as vantagens e desvantagens da sua utilização. Em relação à rede social escolhida para análise, o Instagram será analisado, tendo em conta o modo de inserção dos tópicos para a organização da informação, a informação geral do Instagram, a API disponibilizado pela rede social e a linguagem PHP.

Nos métodos iremos identificar as ferramentas utilizadas neste projecto, caracterizar o tipo de estudo e explicar os procedimentos desenvolvidos.

Nos resultados serão apresentados todas as informações obtidas na API access através da base de dados com recursos a mapas.

Na conclusão e trabalho futuro, apresentaremos as conclusões após análise da informação recolhida nos resultados e identificaremos vias possíveis de desenvolvimento do projecto após o término do presente estudo.

IV. Enquadramento Teórico

1. Internet

A internet é nas suas origens uma rede de computadores que estão ligados entre si, com um número de regras muito reduzido. Não existe uma entidade que seja reconhecida como sua proprietária.[5]

Tendo em conta a evolução da internet no Mundo, verificamos que desde 1995 até ao ano de 2008 (Figura 2), existiu um crescimento exponencial da utilização da internet sendo que 1500 milhões de pessoas utilizariam a internet em 2008.[6]

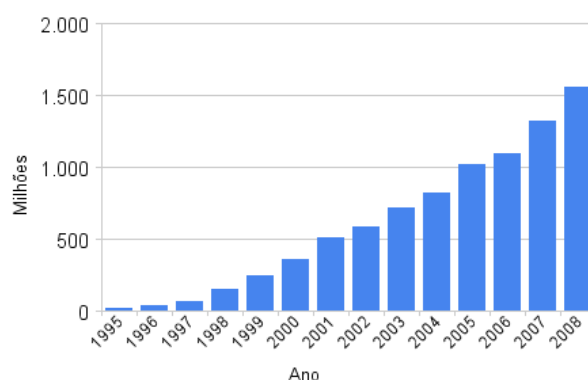


Figura 2 - Crescimento de utilizadores da Internet entre 1995-2008

Os dados recolhidos em Junho de 2012, segundo o site Internet World Stats[7], proporcionam-nos uma visão sobre a utilização da internet no mundo, podendo esta divisão ser feita também por continentes, no período entre 2000 e 2012. A estimativa da população total mundial em 2012 era de 7.017.846.922 de pessoas. Em Dezembro de 2000 o número total de utilizadores da internet era de 360.985.492, sendo que os últimos resultados obtidos em 2012 eram de 2.405.518.376. Assim, analisando o período de 12 anos acima referido, o aumento de utilizadores da internet cresceu para 566,4% (Figura 3).

| WORLD INTERNET USAGE AND POPULATION STATISTICS June 30, 2012 | | | | | | |
|---|----------------------------|---------------------------------|-------------------------------|-------------------------------|---------------------|---------------------|
| World Regions | Population (2012 Est.) | Internet Users Dec. 31, 2000 | Internet Users Latest Data | Penetration (% Population) | Growth 2000-2012 | Users % of Table |
| Africa | 1,073,380,925 | 4,514,400 | 167,335,676 | 15.6 % | 3,606.7 % | 7.0 % |
| Asia | 3,922,066,987 | 114,304,000 | 1,076,681,059 | 27.5 % | 841.9 % | 44.8 % |
| Europe | 820,918,446 | 105,096,093 | 518,512,109 | 63.2 % | 393.4 % | 21.5 % |
| Middle East | 223,608,203 | 3,284,800 | 90,000,455 | 40.2 % | 2,639.9 % | 3.7 % |
| North America | 348,280,154 | 108,096,800 | 273,785,413 | 78.6 % | 153.3 % | 11.4 % |
| Latin America / Caribbean | 593,688,638 | 18,068,919 | 254,915,745 | 42.9 % | 1,310.8 % | 10.6 % |
| Oceania / Australia | 35,903,569 | 7,620,480 | 24,287,919 | 67.6 % | 218.7 % | 1.0 % |
| WORLD TOTAL | 7,017,846,922 | 360,985,492 | 2,405,518,376 | 34.3 % | 566.4 % | 100.0 % |

Figura 3 - Estatísticas da utilização da Internet entre 2000-2012

Quando analisamos a distribuição do número de pessoas que utilizam a internet em 2009 por país, observamos que, segundo a figura abaixo (Figura 4), numa escala de 0 a

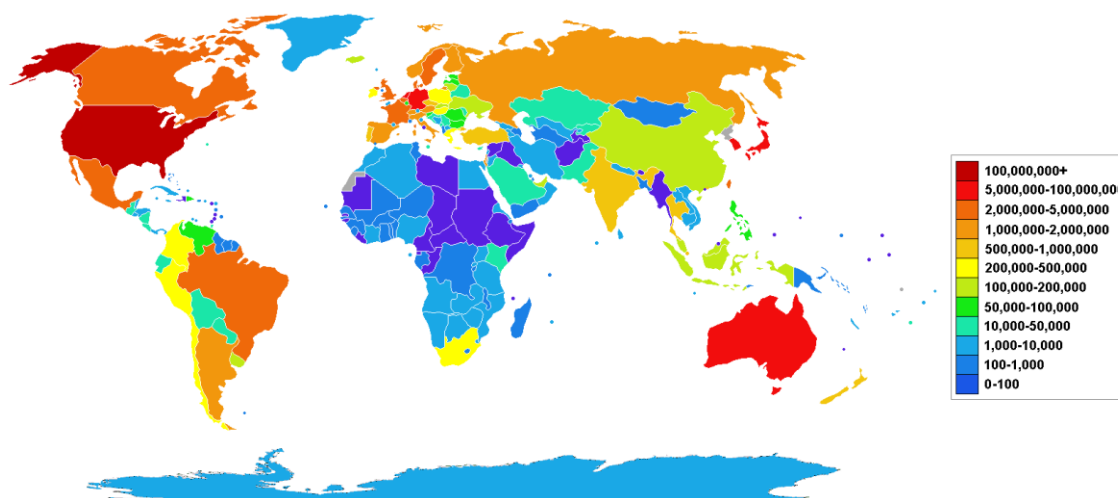


Figura 4 - Distribuição de pessoas que utilizam a Internet por país em 2009

mais de 100.000.000 utilizadores, nos países africanos, exceptuando a África do Sul, países do médio oriente (exemplos como o Iémen, o Iraque, o Afeganistão e o Irão), países a sul da China (Vietnam, Camboja, Laos) e países como o Suriname, Nicarágua, Guatemala, Cuba, Albânia e Bielorrússia, a utilização da internet é limitada ou quase inexistente. Paralelamente entre os maiores utilizadores de internet encontramos países como os Estados Unidos da América, a Alemanha, o Japão e a Austrália, sendo estes países bastante desenvolvidos e que a maioria da sua população recorrer à internet.[8]

1.1. Políticas sobre a Internet

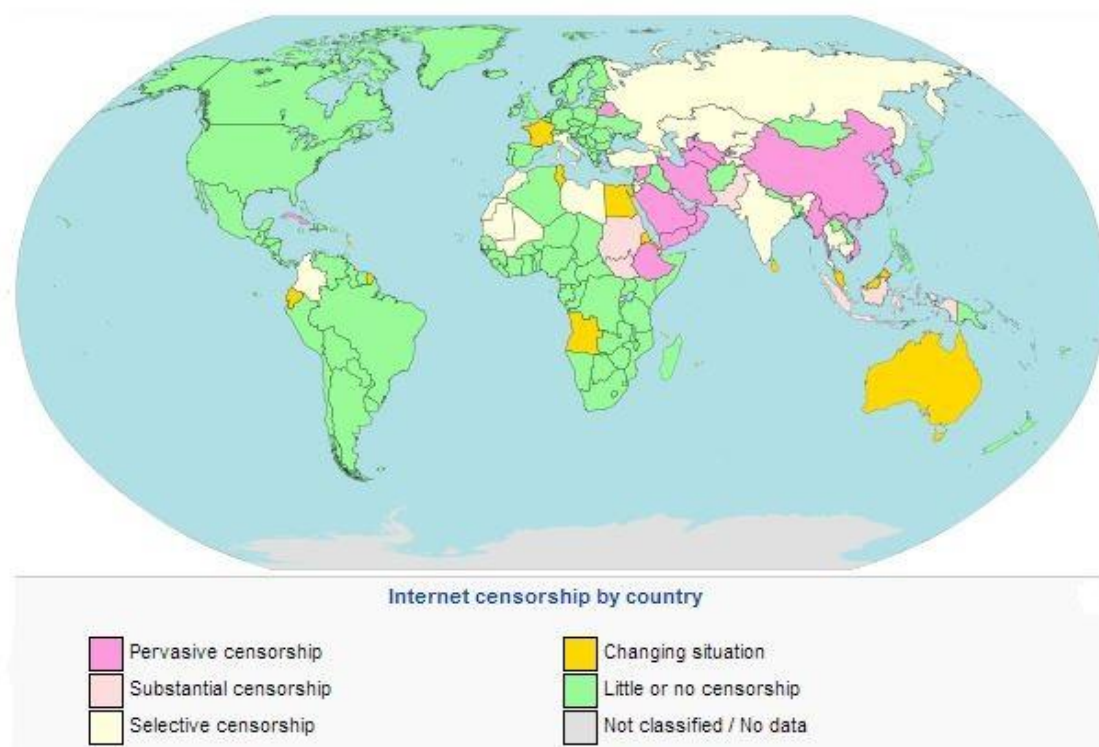


Figura 5 - Graus de censura da Internet no Mundo

As políticas de cada país sobre a Internet variam em relação à censura. Na figura acima (Figura 5), verificamos que em certos países/zonas existe uma maior limitação de acesso à Internet, impossibilitando a partilha e a troca de informação entre pessoas. Na maioria dos países existe uma ligeira censura ou mesmo nenhuma. [9]

Um dos exemplos que se pode retirar do mapa acima é o da China, local onde a Internet é censurada de 3 formas distintas. Primeiramente o governo chinês bloqueia o acesso a determinados *sites* como por exemplo o Twitter. Existe também uma censura das ferramentas a que muitos utilizadores de todo o mundo têm acesso, sendo criada uma ferramenta alternativa ao Twitter (o Weibo) onde certos termos não podem ser utilizados, correndo o risco de a conta ser fechada. Por fim, o encorajamento da visão nacionalista domina as conversas na Internet, tornando difícil a existência de outro tipo de opiniões. [9], [10]

2. Redes Sociais

As redes sociais surgiram como um meio de ligar pessoas através da Internet. Estas redes sociais funcionam tendo em base perfis dos utilizadores, que englobam todo o tipo de informação como por exemplo: o que a pessoa gosta e não gosta, as suas informações pessoais e tudo o que queira partilhar mais. O objectivo principal das redes sociais é aproximar através da Internet, pessoas que tenham vários aspectos em comum ou que se assemelhem de alguma forma. É possível através destas redes partilhar experiências, acontecimentos de vida, lugares onde se esteve e até mesmo conhecimentos sobre determinadas temáticas. [11], [12]

Apesar desta realidade ser comum na maioria dos países, existem outros em que a liberdade de expressão de cada um é oprimida por razões políticas. Foi aliás para combater esta falta de liberdade de expressão, que, a partir do dia 15 de Maio de 2011, em muitas destas redes sociais foram convocadas manifestações, chegando mesmo a transformarem-se em revoluções, como por exemplo a Primavera Árabe que teve começo na Tunísia.[13], [14]

As redes sociais apresentam vantagens e desvantagens na sua utilização. Entre as vantagens podem-se contar aspectos como a comunicação instantânea entre pessoas, a partilha de qualquer tipo de informação que ocorra no lado oposto do mundo onde a pessoa se encontra, sem nenhum custo associado e com bastante rapidez, a discussão de ideias entre grupos de pessoas com interesses comuns e até a troca de informação sobre empregos disponíveis, entre outras. [15], [16]

As desvantagens apresentadas podem ir desde a utilização de dados pessoais por criminosos com fins maliciosos (mesmo sendo o próprio utilizador a fornecer esses dados na rede e qualquer pessoa pode visualiza-los.), à incapacidade das redes sociais oferecerem protecção total, mesmo que possuam níveis de privacidade elevados, não garantindo a segurança total dos dados dos seus utilizadores, à dependência/vício, ao dificultar a comunicação entre as pessoas, não permitindo estabelecer relações interpessoais o que num futuro poderá levar à incapacidade de expressarem sentimentos.[15], [16]

A integração das sociedades à volta do mundo está a atravessar uma mudança significativa devido à crescente interacção massiva da população, que acontece no dia-a-dia, através das redes sociais. De todas as redes sociais conhecidas como por exemplo, Facebook, Twitter, Pinterest e Google +, entre outras, o Instagram possibilita uma análise da informação através da sua interacção que permite partilhar e consumir informação de diversas fontes (principalmente texto e imagens).[3]

2.1. Instagram

O Instagram é uma rede social que permite a partilha de fotos através de aplicações que têm Android e iOS. É um aplicativo gratuito através do qual é possível tirar fotografias com o telemóvel, contendo efeitos que se podem aplicar às imagens e posteriormente partilhar com a rede de amigos que tem nessa mesma rede social. É possível publicar estas mesmas fotografias em redes sociais como o Facebook ou o Twitter. Os utilizadores do Instagram podem gostar e comentar as fotografias e utilizar os hashtags para encontrar imagens relacionadas com a mesma temática, mesmo que as fotos dessa temática não sejam de pessoas que fazem parte da sua rede de amigos. [4]

É possível definir a privacidade em cada uma destas redes sociais. São exemplo disso o Facebook que: *“permite que outras pessoas encontrem o seu perfil, procurando pelo seu nome ou endereço de email, mas você pode proteger as informações particulares do seu perfil de qualquer um que você não tenha aprovado especificamente.”* [17]

Tendo em conta a definição de privacidade do Instagram, este controlo de definições de privacidade não é tão restrito como o Facebook. É possível qualquer pessoa aceder a informação disponibilizada pelos utilizadores desta rede social sem grandes restrições. Com o aparecimento das redes sociais, utilizadas em todo o mundo, os utilizadores podem partilhar qualquer tipo de informação sob o chapéu do anonimato, pseudónimos ou mesmo adoptando outras identidades. [17]

2.2. API do Instagram

O acesso à API do Instagram é público e fácil, como pode ser verificado na figura acima (Figura 6).

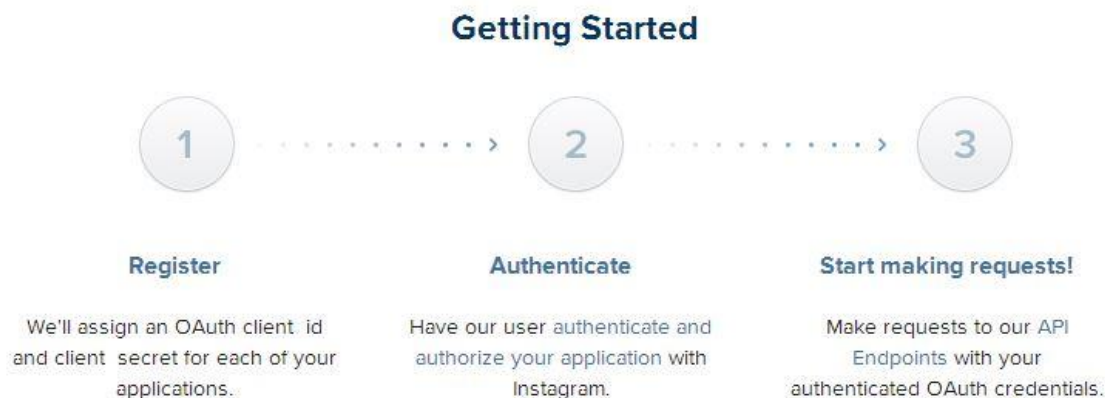


Figura 6 - API do Instagram

A API do Instagram é segura porque obriga o developer a um registo, utilizando um protocolo de autenticação bastante forte mas, a partir desta acção é bastante fácil aceder a todo o tipo de informação que o Instagram disponibiliza (com excepção a alguns casos específicos) é pública.[18], [19]

3. PHP

A linguagem PHP é uma linguagem de programação bastante utilizada para quem desenvolve programas para web. É uma linguagem bastante útil, uma vez que tem por base a programação orientada a objectos que facilita a sua integração com variáveis dinâmicas. As características mais significativas desta linguagem são: a velocidade e a robustez, o facto de ser open-source e server-side. Trata-se ainda de uma linguagem que torna possível a ligação entre a web e bases de dados já existentes num dado computador. [20], [21]

V. Métodos

Para a realização do presente projeto recorreremos ao programa Eclipse com a ferramenta PTI – PHP Tool Integration, ao Notepad ++ e a um aglomerado de softwares denominado Wamp.

1. Tipo de estudo

Existem duas formas de abordagem deste estudo. A primeira é uma forma em que identificamos características estáticas da rede social (que não são passíveis de alterar com o tempo). A segunda é de forma dinâmica que utiliza métodos de teoria da informação e modelos da rede para identificar a formação e propagação de informação.

A primeira abordagem foi utilizada para identificar características de interações no Instagram. Para isto, iremos recolher um número significativo de posts, utilizadores e comentários durante um determinado período de tempo e posteriormente analisámos o conteúdo obtido.

2. Procedimentos

O processo iniciou-se com o acesso à API do Instagram. Seguidamente foi realizado o acesso ao site <http://instagram.com/developer/> para, mediante registo ser possível actuar como um developer e poder realizar chamadas à API do Instagram. Após esta etapa foi gerado um Access Token através do site <http://jelled.com/instagram/access-token> de forma a haver um reconhecimento seguro por parte da API do Instagram, dado que todas as chamadas à API têm que ser realizadas por HTTPS. Com a consulta do site <https://apigee.com/console/instagram> para nos inteirarmos dos pedidos que a API do Instagram aceita, tentámos perceber quais os métodos relevantes para o projecto e consequentemente qual o tipo de dados que retorna.

Criou-se com o auxílio da ferramenta Eclipse um workspace para se desenvolver o ficheiro PHP que faz a ligação entre a API do Instagram e a Base de Dados criada posteriormente.

De seguida acedeu-se à API do Instagram realizando pedidos através de HTTPS, característica que obrigou a uma configuração mais avançada do servidor Apache e consequentemente a transformação do pacote de *softwares* Wamp.

Tendo este passo finalizado, foram feitos pedidos através do PHP à API do Instagram, que devolveu objectos Json em formato string. Após esta recolha, existiu a necessidade das string serem alteradas, para posteriormente serem transformadas em objectos passíveis de serem acedidos e mapeados através do PHP.

Depois da transformação descrita acima, ocorreu uma análise e um mapeamento do objecto em maior detalhe, para que futuramente seja possível inserir dados na Base de Dados.

Com esta parte do projecto concluída e após conversa com o orientador, foi esquematizada a base de dados que iria albergar a informação obtida nos passos anteriores. O esquema da Base de Dados inicialmente apresentava a estrutura abaixo (Figura 7).

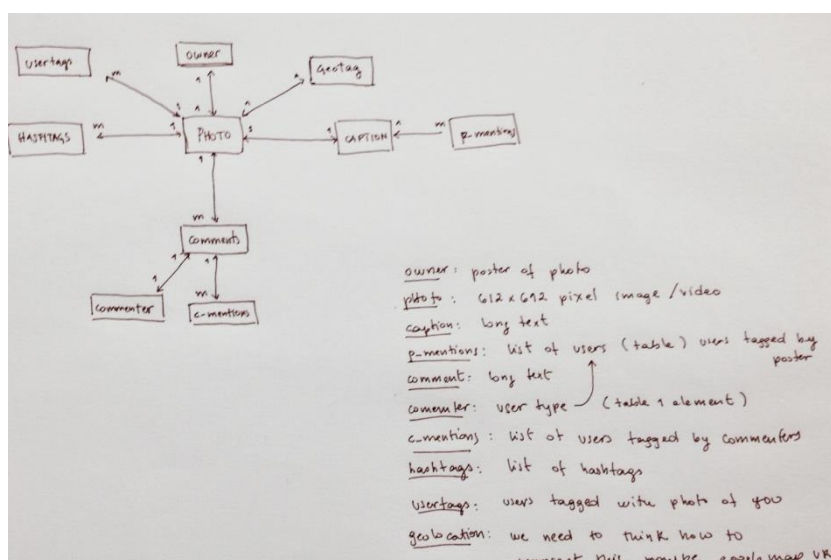


Figura 7 - Esquema inicial da base de dados

Com base no esquema apresentado em cima, foi construída a Base de Dados que iria conter a informação necessária para a sua análise.

Após análise cuidadosa do mapeamento da informação que foi recolhida da API, procedeu-se à análise e à reestruturação da Base de Dados com base nos objectivos posteriormente traçados.

Inicialmente a informação recolhida dos posts era apenas dos que apareciam pela primeira vez na nossa pesquisa, sendo que esta filosofia foi alterada, passando posteriormente a serem incluídos todos os posts (new and updated).

Juntamente com a informação recolhida, foram aparecendo erros ligados à rede do Instagram e à forma como a mesma rede social devolve a informação sobre o objecto.

Para complemento da informação anteriormente recolhida foi adicionada à Base de Dados, toda a informação dos comentários e de todos os utilizadores que realizaram comentários. Esta informação foi guardada na Base de Dados através da interface PDO de uma forma não persistente mas criando uma conexão a cada interacção.

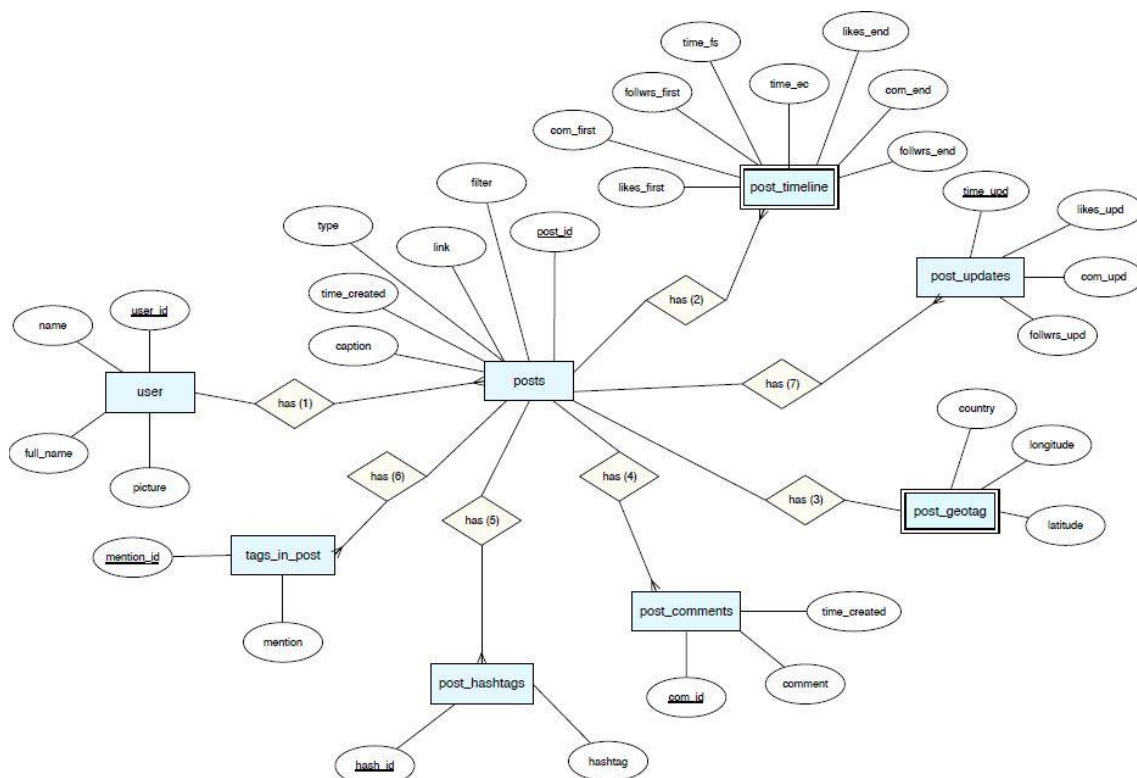


Figura 8 - Esquema de relação entre entidades da base de dados

Posteriormente o código foi testado num período de vinte e quatro horas, onde foram assinalados erros e correcções que deveriam a serem implementadas.

Sendo que um dos erros recorrentes apresentados *“too many connections”*, era proveniente da forma como a ligação à base de dados estava estabelecida, foram criadas conexões persistentes que solucionaram o problema e que mantêm a ligação constante à base de dados.

Tendo em conta todo o trabalho realizado anteriormente e com a intenção de perceber qual o impacto que um post padece aquando da sua entrada na página popular e de que forma os seus utilizadores são influenciados, a base de dados criada anteriormente, foi alterada de forma a comportar e possibilitar a análise a toda a informação vinda deste contexto.

Procedimento de recolha de dados:**INÍCIO:**

1. A cada segundo o programa pede 1 post, sendo que, este processo se realiza num período de 24 horas; portanto para cada resposta da API:
 - 1.1. O programa tenta decodificar a string num objecto passível de ser utilizado, que por vezes é impossível de ser transformado;
 - 1.2. Se conseguir:
 - 1.2.1. Para cada post vai verificar se o id já se encontra na DB
 - 1.2.2. Caso não exista
 - 1.2.2.1. Vai buscar e mapear todos os elementos relevantes (hashtags, comentários, utilizadores identificados no post, geolocalização), que podem ou não existir e todos os elementos fundamentais (dados do utilizador e dados do post) que obrigatoriamente têm de existir.
 - 1.2.3. Se existir
 - 1.2.3.1. Actualiza o número de likes, de comentários do post e de seguidores do utilizador, adicionando um elemento agregado à tabela update, fazendo igualmente uma actualização de todos os comentários.
 - 1.3. Se não conseguir fazer o decode passa ao próximo pedido.
2. No final do ciclo de 24 horas de recolha de dados, vai-se transferir uma última actualização de todos os posts existentes na DB em termos de likes, comentários do post e seguidores do utilizador.

Foi criado com recurso à linguagem PHP, um webservice que retorna todas as localizações presentes na base de dados (podendo ou não) serem limitadas por tempo.

Foi então recomendado pelo orientador a criação de um mapa, que mostraria as localizações dos posts em todo o mundo, que é obtido através do webservice.

VI. Resultados

Os resultados deste projecto serão agregados e apresentados conforme os métodos utilizados, tendo em conta que os produtos finais de cada etapa foram a API access, a Base de Dados e o Mapa dos posts da página popular.

1. API access

Os resultados obtidos deste protocolo são, de alguma forma, dinâmicos, uma vez que os dados totais obtidos são recolhidos, não só dos posts mas toda a informação dos utilizadores que comentaram e de quantos seguidores o utilizador principal do post tem no momento em que o post é encontrado pela pesquisa.

A API access permite o acesso a toda a informação agregada a um post e através do armazenamento de apenas alguns registos, consegue-se criar um “life cycle” do post. Permite também o acesso de forma dinâmica à informação do post, tornando possível, se soubermos o nome ou número do utilizador, aceder a toda a informação que se encontra em todo o Instagram sobre esse utilizador.

| ph_id | type | likes | comments | followers | time_created |
|------------------------------|--------|-------|----------|-----------|--------------|
| 683337874841538698_32310003 | new | 19505 | 268 | 586589 | 1395691266 |
| 683341615553391588_13864937 | new | 38440 | 298 | 2431801 | 1395691155 |
| 683341615553391588_13864937 | update | 38538 | 298 | 2431802 | 1395691200 |
| 683344601683070366_624720002 | new | 26888 | 416 | 707834 | 1395691226 |
| 683344962185685030_11957137 | new | 9485 | 113 | 498161 | 1395691872 |
| 683346594256810414_178522459 | new | 23480 | 355 | 607588 | 1395691519 |
| 683346594256810414_178522459 | update | 23689 | 362 | 607611 | 1395691725 |
| 683348251550005485_51275365 | new | 24331 | 445 | 328901 | 1395691943 |
| 683352663165297440_39989663 | new | 15804 | 139 | 397264 | 1395692226 |
| 683352663165297440_39989663 | update | 16117 | 143 | 397277 | 1395692565 |

Figura 9 - Alteração de vários posts na base de dados

É ainda possível ir buscar a imagem relativa a um post assim como a imagem relativa a um utilizador, sendo apresentado o resultado abaixo (Figura 9).



Figura 10 - Imagens obtidas da API do Instagram

Para chegar a estes resultados foi preciso ultrapassar barreiras como por exemplo: os servidores do Instagram estarem em baixo (o que impossibilitou a chamada à API); pessoas que bloquearam o acesso da API ao seu conteúdo; não existir informação em larga escala sobre o acesso à API. Estas foram no entanto ultrapassadas com alguma facilidade.

De todos os problemas relatados, aquele que mais limitou os resultados foi o facto do objecto devolvido aparecer mal formado e não passível de ser transformado em objecto utilizável. Verificamos que este problema é mais provável de acontecer quando se recolhe uma grande quantidade de posts ao mesmo tempo, e foi ultrapassado através de ajudas *online* o que resultou num processo dispendioso, em termos temporais.

2. Base de Dados

A base de dados criada resulta do agrupamento de oito tabelas (Tabela 1), que contêm registos capturados em um período de tempo compreendido entre 00:00 do dia 2 de Abril de 2014 até 23:59 do dia 3 de Abril de 2014 (24 horas). Cada tabela tem um significado diferente e nelas podemos encontrar os seguintes conteúdos:

Tabela 1 - Campos da base de dados

| Nomes | Definições |
|-----------------------|---|
| post_comments: | O comentário, a ligação ao post e a hora em que foi comentada e a ligação ao user |
| post_timeline: | Permite criar um life clycle para o post guardando o registo de quando um post foi encontrado pela primeira vez e o seu estado quando acaba o ciclo de recolha de dados |
| post_geotag: | Referencia latitude e longitude |
| post_hashtags: | Texto associado aos hashtags |
| tags_in_post: | Ligação aos utilizadores identificados num post |
| post: | Informação única do post |
| user: | Informação sobre o utilizador |
| post_update: | Cada vez que a pesquisa encontra um determinado post permite guardar o estado do post e do utilizador naquele momento. |

Através da imagem disponibilizada acima, podemos verificar que os utilizadores não fornecem toda a informação pedida aquando do preenchimento dos vários campos pedidos pelo Instagram. Só com esta informação é possível vermos esta falta de preenchimento dos campos.

Tendo em conta cada tabela individualmente, os resultados advêm do número de registos que até ao momento cada uma tem. Assim a tabela 2 (ver abaixo) apresenta o número identificado em cada campo.

Tabela 2 - Números reais da base de dados

| Designação da tabela | Número de registo |
|----------------------|-------------------|
| post_comments | 111734 |
| post_timeline | 4488 |
| post_geotag | 1406 |
| post_hashtags | 4665 |
| tags_in_post | 1150 |
| post | 4488 |
| user | 232392 |
| post_update | 6872 |

Foram descobertos indícios de que, por cada post na página popular, o seu número de links e comentários aumenta, assim como o número de pessoas a seguirem aquele utilizador. No exemplo apontado abaixo verificamos que a diferença num mesmo post desde que entrou na página popular é de 313 likes, 4 comentários e 13 novos seguidores para o utilizador (Figura 11).

| ph_id | type | likes | comments | followers | time_created |
|------------------------------|--------|-------|----------|-----------|--------------|
| 683337874841538698_32310003 | new | 19505 | 268 | 586589 | 1395691266 |
| 683341615553391588_13864937 | new | 38440 | 298 | 2431801 | 1395691155 |
| 683341615553391588_13864937 | update | 38538 | 298 | 2431802 | 1395691200 |
| 683344601683070366_624720002 | new | 26888 | 416 | 707834 | 1395691226 |
| 683344962185685030_11957137 | new | 9485 | 113 | 498161 | 1395691872 |
| 683346594256810414_178522459 | new | 23480 | 355 | 607588 | 1395691519 |
| 683346594256810414_178522459 | update | 23689 | 362 | 607611 | 1395691725 |
| 683348251550005485_51275365 | new | 24331 | 445 | 328901 | 1395691943 |
| 683352663165297440_39989663 | new | 15804 | 139 | 397264 | 1395692226 |
| 683352663165297440_39989663 | update | 16117 | 143 | 397277 | 1395692565 |

Figura 11 - Alteração de valores em um post da página popular

Foram detectados problemas de acesso, de performance, relativos a constantes alterações e revisões e consequentes iniciações da base de dados. Os problemas de acesso, erro “too many connections”, observados através das ligações PDO, foram resolvido criando conexões persistentes.

Observou-se que o MySQL apresentava dificuldades ao nível da performance sendo que as *queries* realizadas à base de dados prolongavam-se por um extenso período de tempo.

Com vista aos melhoramentos necessários à base de dados e por motivo das sucessivas revisões da mesma, os dados presentes na base tiveram de ser constantemente iniciados.

3. Mapa

Através do mapa de geolocalização dos posts (Figura 12) verificou-se que a maioria dos posts que chegam à página popular e que têm localização, distribuem-se maioritariamente por zonas como a Europa e os Estados Unidos da América. Verificou-se também um crescimento de posts em zonas como o Médio Oriente, o Sudeste da Ásia e o Brasil. Não foram verificados quaisquer post em zonas como por exemplo a Venezuela, a Columbia, o Perú, a Bolívia, o Iémen, o Irão, o Iraque e a maior parte do continente africano. Não ocorreram problemas significativos no desenvolvimento desta componente do trabalho.



Figura 12 - Mapa com localização dos posts da página popular

VII. Métodos estatísticos

Recolheram-se dados sobre as mensagens na página popular em um ciclo de 24 horas, recuperando em média um post da página popular por segundo. O conjunto de dados completo contém 4.489 entradas, das quais 1.407 correspondem às mensagens em que o utilizador inclui geolocalização. Este subconjunto dos dados reunidos é a utilizada para as investigações relatadas.

Análises estatísticas descritivas básicas foram realizadas para olhar para o comportamento de mensagens na página popular em termos de (1) número de gostos; (2) comentários recebidos, bem como (3) aumento seguidores para o utilizador que introduziu a foto ou vídeo. Para estas três variáveis-chave, e para o (4) país, nós olhámos para as suas distribuições. Todas estas distribuições foram consideradas distribuições altamente assimétricas. Isto significa que as medidas regulares associadas com meios e distribuições padrão para distribuições normais ou quase normais não se aplicam às nossas observações. Para essas variáveis que foram classificadas como 1/4, mediana e 3/4 de Quantis para gostos, comentários e seguidores como medidas de tendência central e de dispersão (Tabela 3).

Tabela 3 - Quantis de Gostos, Comentários e Seguidores

| | Quantiles | | |
|------------------|-----------|------|------|
| | 1/4 | 1/2 | 3/4 |
| Likes | 2926 | 5397 | 9965 |
| Comments | 39 | 77 | 156 |
| Followers | 1053 | 2508 | 5170 |

Uma tabela (tabela 4) com os 10 países mais relevantes também é relatada. Além disso, observou-se para alguns dos valores atípicos distantes observados nestas variáveis. Finalmente, apresentam-se também distribuições classificadas por país/hora do dia.

Tabela 4 - 10 países mais significativos na página popular

| | |
|-------------------------|-----|
| 1. USA | 27% |
| 2. Thailand | 16% |
| 3. Turkey | 6% |
| 4. Kuwait | 6% |
| 5. Australia | 6% |
| 6. Brazil | 4% |
| 7. United Arab Emirates | 4% |
| 8. Russia | 4% |
| 9. Saudi Arabia | 3% |
| 10. Sweden | 3% |

Top ten countries in the popular page

No período entre as 00:00 do dia 2 de Abril de 2014 até 23:59 do dia 3 de Abril de 2014, foram recolhidos 1407 entradas, correspondendo a posts, que continham associado geolocalização.

Os países destacados a bold (Estados Unidos da América e Tailândia) são os países com a maior presença na página popular, dividindo o aparecimento dos post conforme o fuso horário. A Turquia (país a itálico) é o país que se destaca por ter uma presença mais constante na página popular. É de notar, a forte presença de países com raízes árabes.

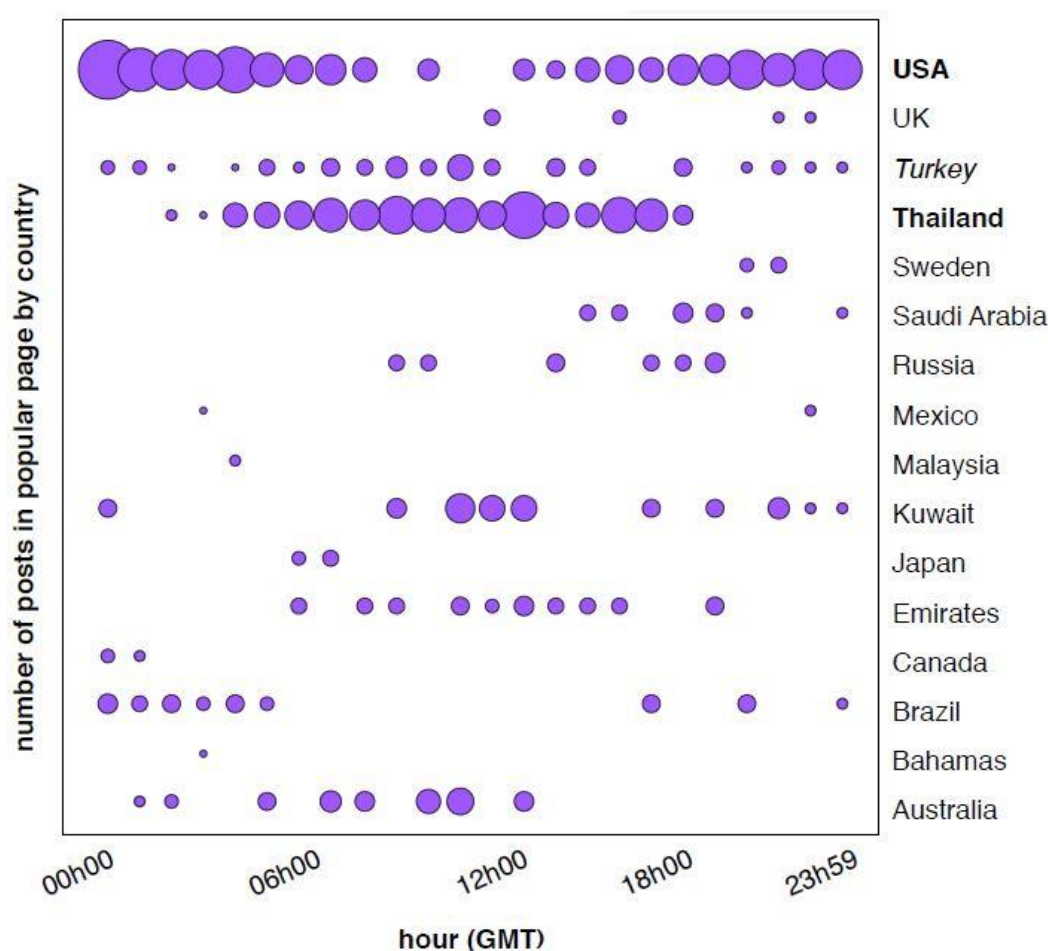


Figura 13 - Número de posts na página popular por país/hora

VIII. Conclusões e trabalho futuro

As conclusões que se podem retirar deste projecto são relativas à API access, base de dados e mapa.

Em relação à API access podemos concluir que os objectos disponibilizados na informação de retorno continuam a ser mal construídos pelo Instagram, pelo que alguns dos objectos que chegam revelam-se impossíveis de serem transformados e mapeados.

De todos os dados recolhidos é possível verificar que determinados utilizadores bloqueiam a própria API access não possibilitando assim o acesso as suas informações. Para o “developer” obter dados mais concretos é preferível realizar mais pedidos de uma quantidade de informação menor do que um pedido de uma grande quantidade de informação, pois os dados serão mais fáceis de aceder.

Em relação à base de dados foram também retiradas algumas conclusões como, por exemplo, que existem pessoas que têm mais de que um post a aparecer na página popular e também que um determinado número de pessoas não fornece (embora não nos tenha sido possível perceber se propositadamente ou não) todos os dados que podem estar presentes num post. Podemos também observar alterações em posts/utilizadores que tiveram um post seu a aparecer na página popular, como por exemplo o referido na figura 11.

No mapa conclui-se que uma percentagem significativa dos posts que aparecem na página popular e que têm associado a sua geolocalização, estão dispersos por diferentes zonas do globo, em países desenvolvidos ou em desenvolvimento. Verifica-se também que em determinadas regiões do mundo não se registam quaisquer ocorrências de posts, o que pode acontecer por um número variado de razões, que não iremos analisar por estar à margem dos objectivos do presente estudo.

Terminamos por fim com a conclusão que muitos utilizadores não têm conhecimento de quanta informação pessoal disponibilizam nesta rede social, e da facilidade de acesso por qualquer pessoa, sem que disso tenham conhecimento

As conclusões do presente estudo levam-nos a acreditar na necessidade de aprofundamento do tema em análise, o que pretendemos que aconteça em Tese de Mestrado, onde nos propomos a analisar os dados de forma qualitativa.

IX. Bibliografia

- [1] J. Horchert and C. Stöcker, “Mapping the Internet: A Hacker’s Secret Internet Census,” *Spieg. Online*, Mar. 2013.
- [2] “10 Ways Brands Use Instagram for Business,” *SumAll*. [Online]. Available: <http://blog.sumall.com/journal/10-ways-brands-use-instagram-for-business-2.html>. [Accessed: 30-Mar-2014].
- [3] “Instagram Is the Best Social Media Service for Businesses: Report,” *The Next Web*. [Online]. Available: <http://thenextweb.com/facebook/2013/12/05/sumall-annual-report-finds-instagram-best-social-media-service-businesses/>. [Accessed: 30-Mar-2014].
- [4] “Instagram,” *Wikipédia, a enciclopédia livre*. 30-Mar-2014.
- [5] “Internet,” *Wikipedia, the free encyclopedia*. 30-Mar-2014.
- [6] “Panorama da Internet no Mundo,” *Jornalismo & Comunicação*. .
- [7] “Internet World Stats - Usage and Population Statistics,” *Internet World Stats*. [Online]. Available: <http://www.internetworldstats.com/stats.htm>. [Accessed: 30-Mar-2014].
- [8] “Uso da Internet no mundo,” *Wikipédia, a enciclopédia livre*. 15-Mar-2014.
- [9] “Enemies of the Internet,” *Enemies of the Internet*. [Online]. Available: <http://12mars.rsrf.org/2014-en/>. [Accessed: 30-Mar-2014].
- [10] D. Bamman, B. O’Connor, and N. Smith, “Censorship and deletion practices in Chinese social media,” *First Monday*, vol. 17, no. 3, Mar. 2012.
- [11] “Social Networking,” *Mashable*. [Online]. Available: <http://mashable.com/category/social-networking/>. [Accessed: 30-Mar-2014].
- [12] “Social networking service,” *Wikipedia, the free encyclopedia*. 30-Mar-2014.
- [13] M. Storck, “The role of social media in political mobilisation: a case study of the January 2011 Egyptian uprising,” *Univ. St Andrews Scotl.*, 2011.
- [14] “Twitter Revolution: How the Arab Spring Was Helped By Social Media,” *PolicyMic*. [Online]. Available: <http://www.policymic.com/articles/10642/twitter-revolution-how-the-arab-spring-was-helped-by-social-media>. [Accessed: 30-Mar-2014].
- [15] “Advantages and Disadvantages of Social Networking: Should You Spend More Time Connecting Online?” .

- [16] “Top5 Vantagens e Desvantagens das Redes Sociais,” *Gap Consulting*. [Online]. Available: <http://gapconsulting.org/top5-vantagens-desvantagens-das-redes-sociais/>. [Accessed: 30-Mar-2014].
- [17] “Social media research raises privacy and ethics issues.” [Online]. Available: <http://www.usatoday.com/story/news/nation/2014/03/08/data-online-behavior-research/5781447/>. [Accessed: 30-Mar-2014].
- [18] Instagram, “Embedding Endpoints.” [Online]. Available: <http://instagram.com/developer/embedding/>. [Accessed: 30-Mar-2014].
- [19] D. H. <dick.hardt@gmail.com>, “The OAuth 2.0 Authorization Framework.” [Online]. Available: <http://tools.ietf.org/html/draft-ietf-oauth-v2-12>. [Accessed: 30-Mar-2014].
- [20] “PHP,” *Wikipédia, a enciclopédia livre*. 28-Mar-2014.
- [21] PHP Group, “php,” 2014-2001. [Online]. Available: <http://php.net/>. [Accessed: 30-Mar-2014].



X. Anexos



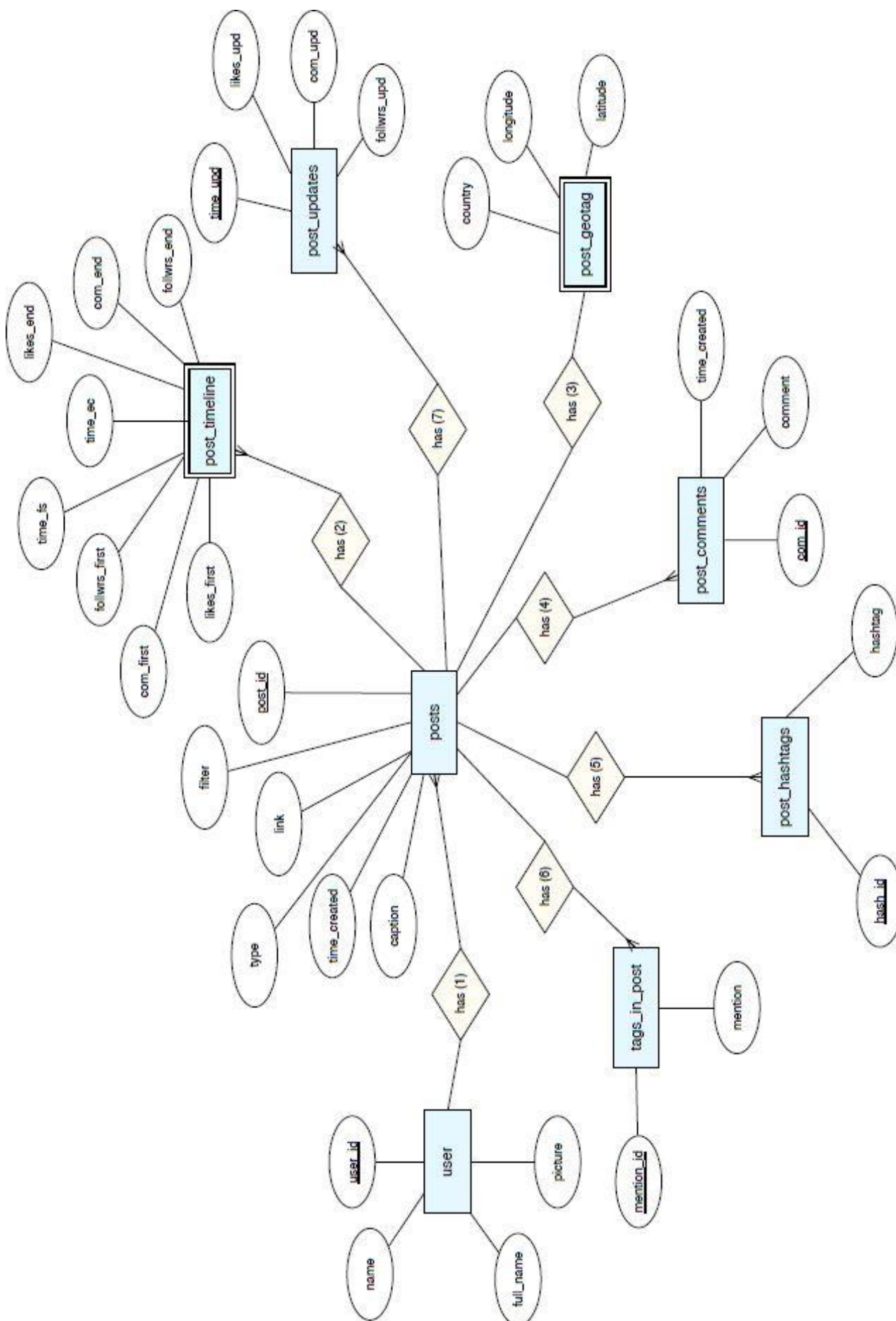
1. Manual Técnico da Aplicação

Procedimentos

- 1 – É necessário para o funcionamento correcto da recolha de dados uma máquina com uma arquitectura de 64-bits;
- 2 - Instalar Wamp disponível em: <http://www.wampserver.com/en/> na directoria C: ;
- 3 - Com a tarefa descrita em 2 finalizada, alterar permissões do php.ini para permitir openssl e do httpd.conf para permitir conexões de todos;
- 4 - Copiar a pasta www entregue no CD para a directoria de C:/wamp substituindo a pasta que lá se encontra;
- 5 - Ligar Wamp;
- 6 - Criar Base de dados em localhost -> phpmyadmin através de browser com a querie presente em BD_final.txt;
- 7 – Copia a pasta Eclipse entregue no CD para o computador para a directoria C: ;
- 8 – Correr o programa Eclipse através de aplicativo presente na pasta no passo 7;
- 9 - Correr no Eclipse o programa através da configuração presente no mesmo;
- 10- 24 horas depois através da página localhost->instagram->map.php pode começar a ver a distribuição de pontos no mapa.



2. Base de dados





XI. Glossário

API - Application Programming Interface

HTTPS - HyperText Transfer Protocol Secure

iOS - Sistema Operacional Móvel da Apple Inc

Json - JavaScript Object Notation

MYSQL - Structured Query Language

PDO – PHP Data Objects

PHP - Hypertext Preprocessor originalmente Personal Home Page

USA – United States of America

Wamp - Windows Apache MySQL PHP

XII. Índice Remissivo

| | |
|-----------------------------|--|
| Análises estatísticas | 25 |
| API..... | 8, 9, 15, 16, 17, 20, 35 |
| base de dados | 8, 9, 17, 18, 19, 22, 23 |
| censura | 9, 12 |
| developer | 15, 16, 29 |
| dinâmica | 5, 7, 8, 16, 20 |
| dinâmicos..... | 5, 8 |
| <i>engagement</i> | 5, 6, 7 |
| erros | 18 |
| geolocalização | 8, 24, 25, 26 |
| informação | 7, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22 |
| Instagram | 1, 5, 6, 7, 8, 9, 14, 15, 16, 17, 18, 22, 28, 29 |
| internet..... | 7, 10, 11 |
| life cycle | 20 |
| mapa | 8, 12, 19, 24, 32 |
| objecto | 8, 17, 18 |
| partilhar..... | 13, 14 |
| PHP..... | 8, 9, 15, 16, 17, 19, 29, 35 |
| popular..... | 5 |
| programação | 8, 15 |
| redes sociais..... | 5, 8, 9, 13, 14 |
| tempo | 7, 8, 16, 19, 22 |
| utilizadores | 5, 7, 9, 10, 11, 12, 13, 14, 16, 18, 19, 20, 22 |
| workspace | 16 |