Outline

Outline

- Explain the outline of the presentation

Outline

- Current Section

Introduction

What is Twitter:
- Twitter is a Microbloging Platform.
- Users share what is happening in less than 140 characters.

Why do Topic Detection on Twitter?
- 19% of all tweets are about brands and 78% of Internet users put their trust on other users.
- Before events hit the news reports, they are being commented on Twitter.

- Twitter is a Microbloging Platform.
- Users share what is happening in less than **140**
- **19%** of all tweets are about brands and 78% of Internet users put their **trust** on other users.
- Before **events** hit the news reports, they are being **commented on Twitter.**

Topic Detection on Twitter

Hard to Detect Topics on Individual Tweets
- Slang words.
- Typos.
- Small body of text.
- General *topic detection* mechanisms rely heavily on *TF-IDF*.

More Information Than Just Words
- Hashtags; Replies; Timestamps; Geo Coordinates.
- Social network behind the author of the tweet.

- **Hard to Detect Topics on Individual Tweets**
- Slang words.
- Typos.
- Small body of text.
- General *topic detection* mechanisms rely heavily on *TF-IDF*.
- **More Information Than Just Words**
- Hashtags; Replies; Timestamps; Geo Coordinates.
- Social network behind the author of the tweet.
- **TF-IDF** tf–idf, short for term frequency–inverse document frequency, **division** between how many times a word appears in the document, divided by how relevant the word can be on the entire set.

Document Clustering

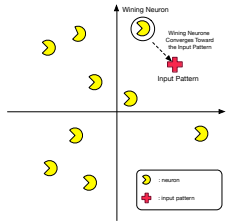
Cluster analysis or *clustering* is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

- Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).
- can be **supervised** or **unsupervised**

The Self Organizing Map
- A self-organizing map is a type of artificial neural network that is trained using *unsupervised* learning to produce a low-dimensional representation of the input space of the training samples, called a map.
- Self-organizing maps use a neighborhood function to preserve the topological properties of the input space.
- Mimics the way the cortex of highly developed animals brain (are supposed to) work.

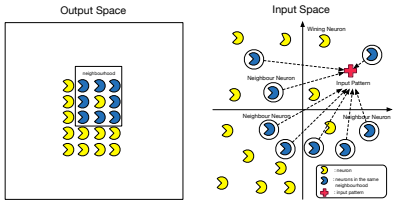
- **The clustering algorithm used during this thesis**
- A self-organizing map is a type of artificial neural network that is trained using *unsupervised* learning to produce a low-dimensional representation of the input space of the training samples, called a map.
- Self-organizing maps use a neighborhood function to preserve the topological properties of the input space.
- Mimics the way the cortex of highly developed animals brain work.

SOM Input Space



- Explain the input space

SOM Output Space



- Explain the output space space

Outline

- Current Section

GEO SOM
- Applies the first law of geography "Everything is related to everything else, but near things are more related than distant things".
- Defining a variable k which is used as a "geographical tolerance" that forces the winning neuron to be geographically near the input pattern.
- selection of the winning neuron is done in two steps. First, geographic neurons inside the tolerance k with the input data as a center are selected. Only after that, comparisons are made with the rest of data present in the input data.

- Applies the first law of geography "Everything is related to everything else, but near things are more related than distant things".
- Defining a variable k which is used as a "geographical tolerance" that forces the winning neuron to be geographically near the input pattern.
- selection of the winning neuron is done in two steps. First, geographic neurons inside the tolerance k with the input data as a center are selected. Only after that, comparisons are made with the rest of data present in the input data.

WEBSOM

The WEBSOM self organizing maps

- First SOM is called *word category map* used to find words that have similar meaning.
- Second SOM, called *text*, used to cluster the documents.

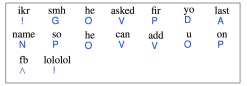
- by Honkela
- Websom is used to cluster webpages
- has two soms

Twitter Natural Language Processing

ARK Tweet NLP:

- Built using a maximum entropy Markov model.
- Tags words, such as nouns, verbs, etc..
- Can tag words, such as abbreviations, emojis and spelling errors.

Example:



- Twitter sucks with graphical errors
- NLP no twitter antes deste trabalho nao era possivel
- Tweet automatically tagged with ARK Tweet NLP. ! stands for interjection, while V stands for verbs and D for determiner.

Miller et al, 2001

- Similarity breeds connection.
- Homophily means *"people like us.*
- In diverse societies, race, and race-like ethnicity create the most stark divides.
- Sex, age, religion, and education strongly structure our relations with others.

- Not web social, like real social

Outline

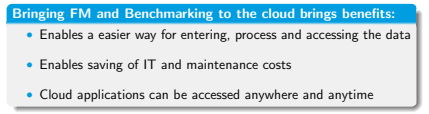
- Current Section

Goals

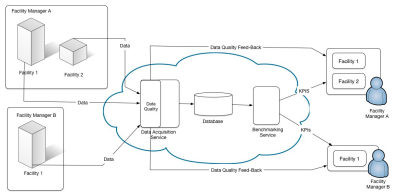
- Display KPIs through graphics
- Have a ranking between organizations
- Use authentication service to authenticate the users of a organization
- Have a cache on the database for better performance

- A aplicação vai fazer:
- A aplicação vai permitir uma vizualização de indicadores de performance, por forma a que cada organização consiga ter uma percepção global do seu estado face à concorrência em tempo real
- Display KPIs through graphics
- Mostrar os KPIs das diferentes fcilities através de gráficos
- Ter um ranking entre as diferentes facilities
- Ter um serviço de autenticação de utilizadores, onde apenas conseguem aceder à informação relativa à sua facility. Além disso, cada uma das organizações não saberá a identidade das restntes no sistema.
- Teremos, também, uma cache na base de dados para melhor performnce da mesma
-
- para isso, vamos usar cloud, porque..muda slide

Cloud Computing



- porque a cloud está em crescimento e permite os beneficios para a FM como:
-
- Permite uma forma simples e rápida de processamento e acesso a dados
-
- Permite uma redução de custos de IT e manutenção
-
- Podem ser acedidas em qualuqer lado e a qualquer momento
-
- Sendo assim, muda de slide e explica a arquitectura

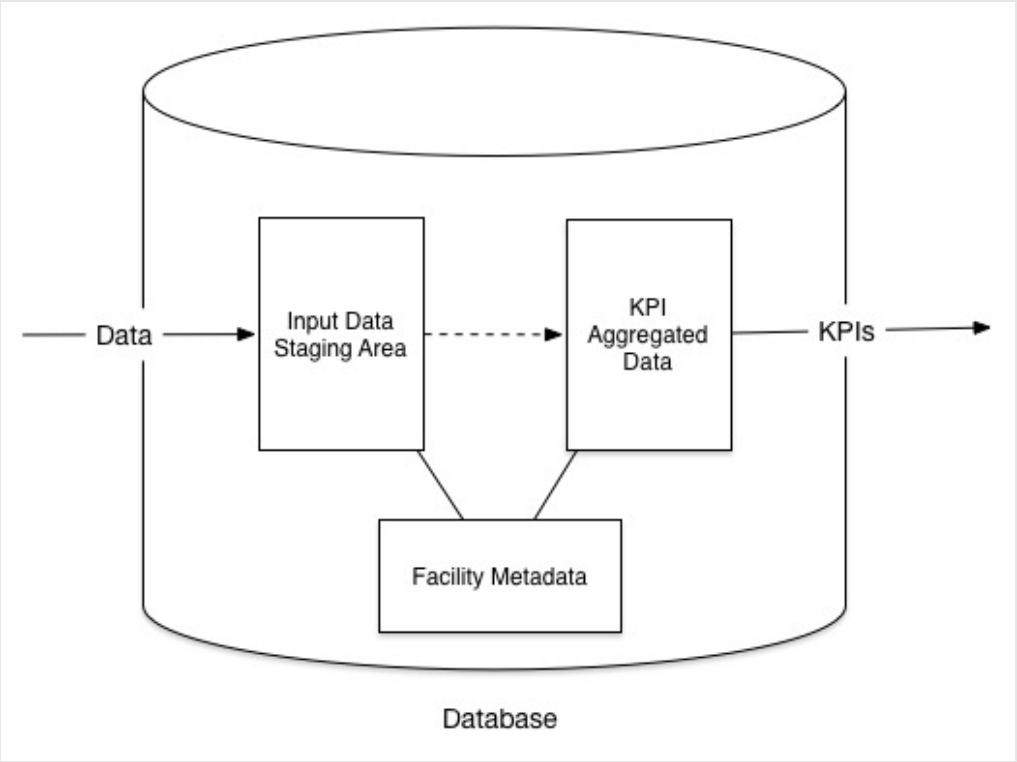


- **Client Side:**
-
- Running on the browser of the user connecting to the website
-
- Bootstrap Framework
-
- Javascript library highcharts or D3.js
-
- **Server Side:**
-
- Will be running the application
-
- Will be the responsible for the processing and storage of the data sent by the Organization to the DB
-
- Play Framework

Architecture

Database
- Relational Database
- Theoretically divided in three

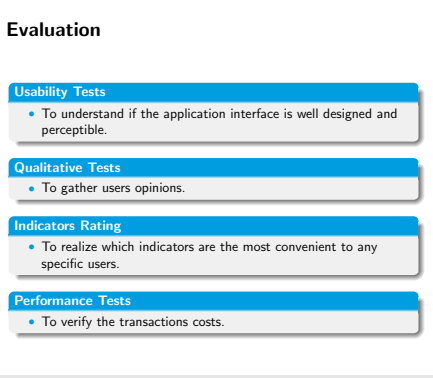
- la



- Heroku is a cloud platform as a service (PaaS) supporting several programming languages. Heroku was acquired by Salesforce.com in 2010.[1] Heroku, one of the first cloud platforms, has been in development since June 2007, when it supported only the Ruby programming language, but has since added support for Java, Node.js, Scala, Clojure, Python and PHP and (undocumented) Perl. The base operating system is Debian or, in the newest stack, the Debian-based Ubuntu.[2]

Outline

- Current Section



- la

Usability Tests

- Para perceber se a interface está bem desenhada e compreensivel

Qualitative Tests

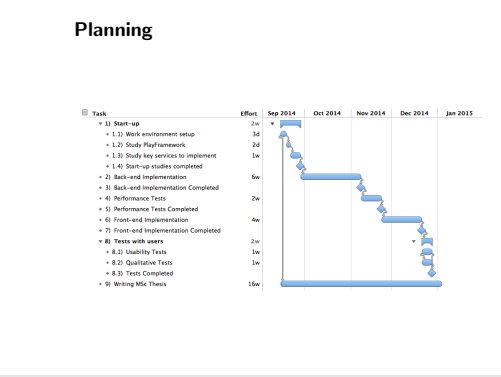
- Para colher as diferentes opiniões dos utilizadores relativamente à aplicação

Indicators Rating

- A aplicação vai ter um sistema de rating de indicadores, onde os utilizadores poderão ajudar na selecção dos mesmos

Performance Tests

- Para verificar os custos por transacção



- Esta tese tem de ser entregue no inicio de Janeiro,
- Estou a contar terminá-la antes do final de Dezembro já devido às festas do Natal e Passagem de Ano
- O trabalho vai começar por um set-up do sistema para desenvolvimento, seguido do desenvolvimento de back-end e só depois do fron-end e interface do sistema.
- Por fim, serão realizados os diferentes testes aos utilizadores.

Outline

- Current Section

Conclusions

- There is no commonly agreed set of metrics to compare facilities
- Analysis of existents standards
- Proposal of a set of KPIs
- Validation through the cloud proposal solution

- Não existe ainda uma forma acordada de fazer benchmarking nem um conjunto de métricas de comparação de facilities.
-
- Neste trabalho fizemos uma analise aos diferentes standards existentes, literatura cientifica e softwares de FM
-
- Propusemos uma lista de KPIs a serem utilizados por todas as organizações
-
- E uma forma de os validar através de uma solução de software cloud

Conclusions

-

Thank you!

Questions?