

Gary Bennett, Rawand Hawiz, Jeansy Molanza, Jiefei Wei

Group Project

Project Initiation Document

October 30, 2015

Loughborough University

Contents

List of Figures	iii
List of Tables	iv
1 Introduction	1
1.1 The Client	1
1.2 The Project	1
1.3 The Team	1
1.3.1 Jeansy Molanza – Project Management/Business Analyst	1
1.3.2 Gary Bennett – Front End Development/Business Analyst	1
1.3.3 Jifei Wei/Rawand Hawiz – Back End Development	2
2 Project Plan	3
3 Software Development Approach	4
3.1 Choice of Development Approach and Rationale	4
3.2 Sprint Cycle	4
3.3 Version Control	5
3.4 Automated Unit Testing	6
3.5 Test/Behaviour Driven Development	6
3.6 Continuous Integration	7
3.7 Team Structure	7
4 Risk Assessment Plan	9
5 Project Management	10
6 Technical Considerations	11
6.1 Sentiment Analysis	11
6.1.1 Step 1: Data collection & pre-processing	11
6.1.2 Step 2: Classifier	12
6.1.3 Step 3: Classify	12
7 Derp	14
8 Ideas	15
8.1 Overview	15
8.2 Business Model	15
8.2.1 Tier 1: Personal	15
8.2.2 Tier 2: Premier	16
8.2.3 Tier 3: Enterprise	16
9 Stakeholder Feedback	18
9.1 Stakeholder 1	18
9.2 Stakeholder 2	18
9.2.1 Overview and Business Model	18
9.2.2 Personal Offering	18
9.2.3 Premier Offering	19
9.2.4 Enterprise Offering	19
9.3 Stakeholder 3	19
9.4 Stakeholder 4	20
9.5 Stakeholder 5	20
9.6 Stakeholder 6	20

9.7 Stakeholder 7	20
10 Mathematical Algorithms	22
10.1 Pattern Matching	22
10.1.1 Description	22
10.1.2 Application	22
10.2 Image Recognition	22
10.2.1 Description	22
10.2.2 Application	23
11 Literature Review	24
References	25

List of Figures

1	Project Plan	3
2	Sprint Cycle	4
3	Distributed Version Control	5
4	Development Time for a Software Project	6
5	Behaviour-Driven Development	6
6	Continuous Integration Process	7
7	Collect Content	11
8	Content Preprocessing	11
9	Word Frequency	12
10	Feature Extraction	12
11	Training Classifier	12
12	Classify	13
13	derp	14
14	Flipping Card User Interface	15
15	Bitap Algorithm Pseudocode	22
16	Comparison of Image Recognition Algorithms	22

List of Tables

1	Project Plan Key	3
2	Project Risk Assessment	9

1 Introduction

1.1 The Client

Atos S.E. (ATO) is a France-based company that provides IT services (Reuters, 2015). The company provides an assortment of services, including: bespoke enterprise software, e-payment services, business integration solutions and cloud services. In addition, the company serves clients in various industries, including: energy and utilities, financial services, manufacturing, retail, public sector, healthcare and transport. Over time, Atos has developed strategic alliances with Cisco, Dell, EMC, Hitachi, Microsoft, Oracle, Samsung, SAP, Siemens and VMWare; and strategic business partnerships with HP and IBM (Bloomberg, 2015). This has allowed them to significantly leverage their service offerings and develop the credibility required to attract new trade. For example, the company was recently awarded a £6m contract to run Transport for London's service desk, a move that will see them responsible for logging, tracking and managing requests such as e-mail configuration, printing and remote access for 25 000 users across multiple locations (Jee, 2015).

1.2 The Project

The group has been tasked with the job of creating a solution to track and identify ones online personal data on behalf of Atos. The solution must be capable of securing, filtering or bringing added value to the data. Consequently, the group must start thinking about actions, functionalities or objectives that might be useful in the solution to protect personal data, for example: automatic de-subscription, notifications, deletion, or ensuring results do not appear in search engine top hits.

1.3 The Team

The following section outlines the expected roles and responsibilities for each individual involved in this project. However, due to the dynamic nature of the project, roles may be re-assigned to one or more individuals throughout the duration of the project. As a result, individuals may need to play a different role than that which was initially outlined. This approach to team development correlates to Ancona and Bresman (2007) ideas on adaptive team structure within organisations:

1.3.1 Jeansy Molanza – Project Management/Business Analyst

As Project Manager, responsible for the successful initiation, planning, execution, monitoring, controlling and closure of the project. As Business Analyst, responsible for the design of solutions that will deliver value to stakeholders – acting as an agent of change whenever necessary.

1.3.2 Gary Bennett – Front End Development/Business Analyst

As Front End Developer, responsible for implementing interactive elements, translating UI/UX wireframes to actual code that will produce the solution's interface that the end user engages with.

As Business Analyst, responsible for the bridging the gap between technical and project requirements.

1.3.3 Jifei Wei/Rawand Hawiz – Back End Development

As Back End Developer, responsible for implementing server-side application logic; creating the application's API to facilitate database interaction and intelligent agents tasked with performing a service on a regular schedule. In addition, will be in charge of writing unit tests and source control.

2 Project Plan

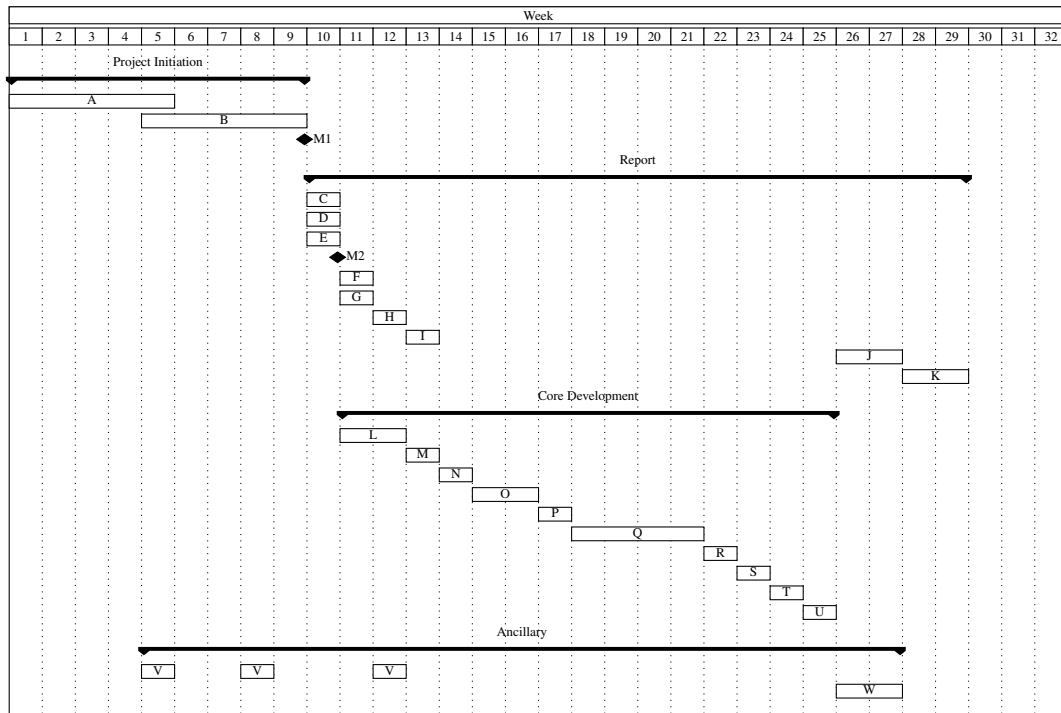


Figure 1 Project Plan

Table 1 Project Plan Key

Code	Task	Code	Task
A	Project Initiation Document	L	Map Application Flow
B	Atos Entry Form	M	Create Application UI
M1	Complete Project Initiation Document	N	Create Application Database
C	Overview	O	Create Application API
D	Problem Definition	P	Create Application Unit Tests
E	Feasibility Study	Q	Create Intelligent Agents
M2	Complete Introduction	R	Test Functional Usability
F	Functional Requirements	S	Detect Security Holes
G	Non-Functional Requirements	T	Correct Technical Issues
H	Architectural/Logical Design	U	Backlog Grooming
I	Physical/Process Design	V	Get Feedback from Client
J	Implementation Section	W	Prepare Client Documentation
K	Evaluation Section		

3 Software Development Approach

3.1 Choice of Development Approach and Rationale

The group has decided to follow an agile development methodology throughout the software development stage. According to Boehm (1981), the cost of removing a software defect or implementing substantial change rises exponentially for each stage of the development cycle when it remains undiscovered. Consequently, the group must start considering the best way to effectively handle inevitable change throughout the projects lifecycle. If the group opted to combat this change by following the traditional waterfall approach and anticipating the complete set of requirements early on then it could quickly be accused of being unresponsive to changing business conditions. Highsmith and Cockburn (2001) assert that a much better strategy can be espoused from the agile method, in which the group can attempt to reduce the cost of change throughout the project by:

- Producing weekly deliveries in order to achieve an early win and acquire rapid feedback
- Inventing simple solutions, so there is less to change and making those changes is easier
- Testing constantly in an automated manner, for earlier, less expensive, defect detection

Rizwan and Hussein (2008) assert that the main disadvantages of agile models are poor quality product, poor design and improper documentation. Fortunately, the group can effectively tackle the abovementioned issues in the following manner: disciplined sprint cycles, version control, automated unit testing, test/behaviour driven development, continuous integration, and adequate team structure.

3.2 Sprint Cycle

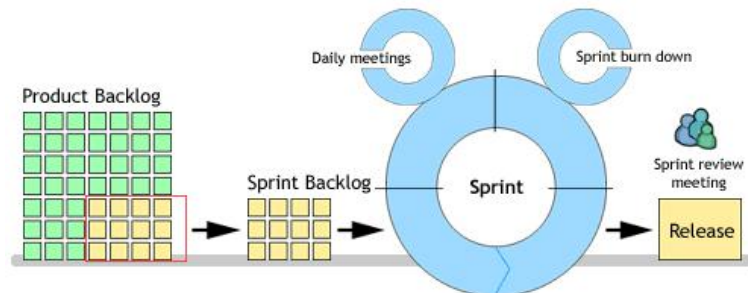


Figure 2 Sprint Cycle

A sprint is “a regular, repeatable work cycle in scrum methodology during which work is completed and made ready for review” (Techopedia, 2015). It is a core part of the agile approach and will see the group working to 1 week sprints over a 10 week period, consequently completing 10 sprints throughout the duration of the project. In addition, it will generate a sense of accountability and ownership within the group, as it promotes transparency, frequent review and constructive discussion.

Software requirements identified at the outset of the project will be maintained in a product backlog, where they are prioritised by business value (Cohen, 2010). The group will receive feedback on which items to prioritise after discussions with the target audience who have a vested interest in the product. Prior to commencing a sprint, the group will have a sprint planning meeting. During this meeting, the group will decide which items to add to the sprint, and identify the tasks necessary to complete each item. If possible, the group will estimate how many hours each task will take someone

to complete (Techopedia, 2015). As team engagement and morale is likely to fluctuate from sprint to sprint, it will be important to get verbal approval from everyone about the commitments being made at the time.

The sprint will commence immediately after the planning meeting. Throughout the sprint, the group will have daily meetings (Monday to Friday and excluding public holidays) on the WhatsApp messaging platform or face-to-face if possible. During this meeting, everyone will be asked:

- What did you do yesterday?
- What are you doing today?
- Are there any currently impediments in your way?

At the end of each sprint, the group will have a sprint review meeting. During this meeting, the group will show what they have accomplished during the sprint and demo any newly added features. In addition, the current state of the project is assessed against the goal determined during the sprint planning meeting. Ideally, the team would have completed each product backlog item brought into the sprint, but if they have not then the item will be carried over into the next sprint (Techopedia, 2015).

3.3 Version Control

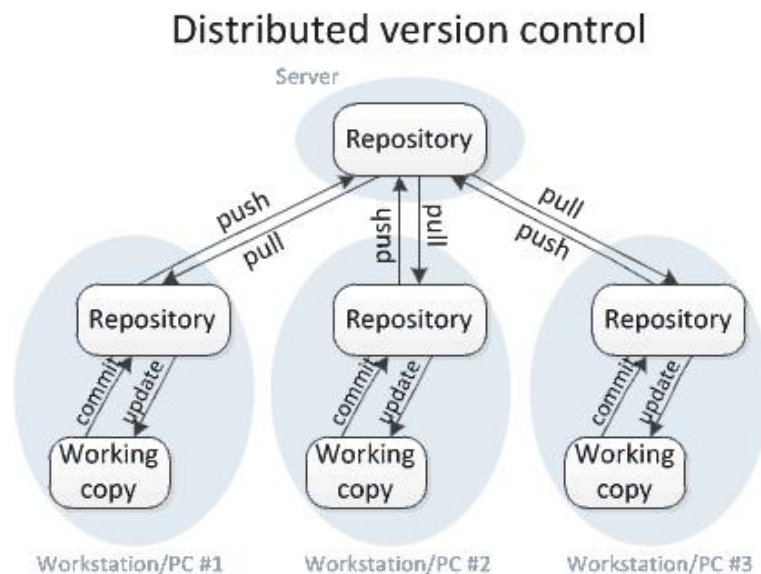


Figure 3 Distributed Version Control

Version control is “a system that records changes to a file or set of files over time so that you can recall specific versions later”. Consequently, a version control system will allow the group to revert files to a previous state, compare changes over time, and see who last modified something that might be causing a problem (Git, 2015). The group will be using GitHub, a distributed version control system. As a result, each group member will need to commit their changes to local repository on their computer before pushing them to a server, instead of directly checking them into the server repository (Jean and Brady, 2014). By splitting up the process into two steps, group members can carry out the necessary due diligence to ensure that their code does not break the solution by testing their changes locally.

3.4 Automated Unit Testing

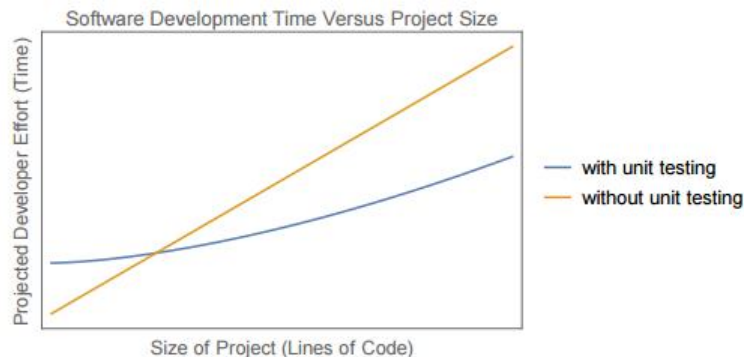


Figure 4 Development Time for a Software Project

According to Sharma, et al (Sharma et al., 2015), a unit test is a function that compares an input-output pair and then returns a Boolean value. A result of True indicates that the code is behaving as intended, but a result of False indicates that it is not, and consequently, that any program relying on that code cannot be trusted to have as intended. To reduce the overall development time, the group will need to create unit tests in order to reduce the number of errors in the application. A well designed unit tests will allow the group to make sweeping changes to the code quickly, as everyone will better understand the design of the code being worked on. This will be helpful when the group starts to document the solution, as it will be easier to define what each part of the application is supposed to do.

3.5 Test/Behaviour Driven Development

```
Feature: Sign up
  Sign up should be quick and friendly.

  Scenario: Successful sign up
    New users should get a confirmation email and be greeted
    personally by the site once signed in.

    Given I have chosen to sign up
    When I sign up with valid details
    Then I should receive a confirmation email
    And I should see a personalized greeting message

  Scenario: Duplicate email
    Where someone tries to create an account for an email address
    that already exists.

    Given I have chosen to sign up
    But I enter an email address that has already registered
    Then I should be told that the email is already registered
    And I should be offered the option to recover my password
```

Figure 5 Behaviour-Driven Development

The group will not write unit tests after all of the software is written. Instead, each unit test will be written beforehand in order to reduce the likelihood of producing bad code – employing a test-driven development (TDD) approach which involves the following steps as suggested by Osherove (2009):

- Write a failing test to prove code or functionality is missing from the solution. For example, if a group member wanted to add a new feature to the solution that remembers the users age, then

they would need to write a test that verifies that the age is indeed a number. At this point, the test should not pass because the solution currently lacks that functionality.

- Make the test pass by writing code that meets the expectations of the test.
- Refactor the code to ensure that a piece of code is easier to read and maintain, while still passing all of the previously written tests.

Alternatively, the group can also experiment with a behaviour-driven development (BDD) approach, in which the intent of the system is tested. This is different to the TDD approach, as the group will be focused on writing scenarios that prove code or functionality is missing from the solution by writing a failing customer acceptance test that describes the behaviour of the system from the customer's point of view (Oshero, 2009). A scenario must be written in manner that can easily be read or written by any group member. North (North, 2015) affirms that this can be done by using a common vocabulary that spans the divide between business and technology. As a result, it becomes a communication and collaboration tool. Consequently, business stakeholders can be drawn into the software development process, helping the group better understand their requirements (North, 2015).

3.6 Continuous Integration

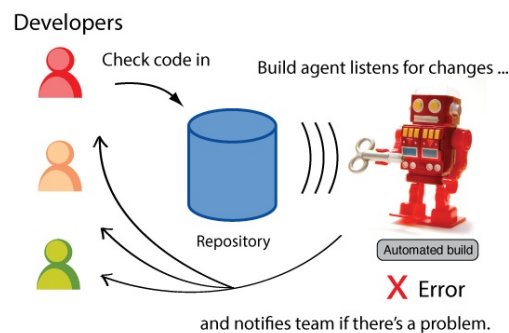


Figure 6 Continuous Integration Process

Continuous integration (CI) is a software development practice in which updates to the solution are immediately tested and reported on when they are added to the main codebase. In using CI, the group will receive rapid feedback so that if a defect is introduced, it can be identified and corrected as soon as possible (North, 2015). The group will be using Travis CI, an open-source continuous integration service to automatically build and test the solution hosted at GitHub. According to Kawalerowicz and Berntson (2011), if the group implements and maintains relatively good CI processes, they are more likely to create a better solution, as they would have done testing and integration consistently earlier in the process, reducing the chances of catching bugs later. Furthermore it promotes transparency, as each group member will be able to see the results of the build and know where the problems are.

3.7 Team Structure

The group is comprised of “generalising specialists” – individuals who have one or more technical speciality coupled with general knowledge of the business domain, so they can deliver direct value to the project in a numerous ways and easily change roles if required at a later stage. Furthermore, as a changing group structure is detrimental to achieving any set objectives, it is expected that the group

will remain as stable throughout the course of the project (Scott and Lines, 2015).

4 Risk Assessment Plan

This section identifies any risks associated with the project, determines the likelihood of their occurrence and hypothesis their overall impact on the project. Furthermore, an attempt is made to categorise each as either technical or non-technical risk. (Dawson, 2009) asserts that technical risks refer to any risk associated with the software development process, whereas non-technical risks are associated with the project management process.

Table 2 Project Risk Assessment

Risk Item	Risk Management Technique	Risk Impact	Risk Category
Group loses sense of urgency lost during earlier stages. It could be because the project is not clearly understood yet.	Group must meet regularly and report back to one another on task progress to generate a sense of accountability.	$1 \times 5 = 5$	Non-Technical
As the project progresses, scope creep begins to threaten deadlines. The group over-promises but under-delivers.	No new item can be added to the product backlog or current sprint unless it is deemed critical to the projects success.	$2 \times 4 = 8$	Technical
The solution proposed by the group does not meet the end users need as the business model is fundamentally flawed.	Group must thoroughly analyse and gather solution requirements during the early stages of the project.	$2 \times 5 = 10$	Non-Technical
Project files (reports, databases or code) are corrupted or lost, as a result the group needs to recreate the affected files.	Group must use a version control system to track code changes and backup documents on a cloud storage service.	$2 \times 5 = 10$	Technical
Group realises it lacks the technical skills required to develop certain features in the solution, so they are omitted.	Group must give itself enough time at the start of the project to properly understand the development environment.	$1 \times 5 = 5$	Technical
Group regularly moves onto the next sprint without finishing all items picked up in one or more previous sprints.	Group must ensure that they have enough time to complete assigned items per sprint and not over-promise anything.	$1 \times 5 = 5$	Non-Technical
Group does not meet regularly, and due to a lack of communication it is hard to know what everyone is doing.	Group must endeavour to consistently adopt working practices that promote transparency and accountability.	$2 \times 3 = 6$	Non-Technical
Group member accidentally overrides another members work because they didnt have the latest version of the code.	Group must use a version control system to track code changes and backup documents on a cloud storage service.	$1 \times 3 = 3$	Technical
Group spends too much time discussing what needs to be done, and consequently has no time to work on solution.	Group leader must have an agenda prepared prior to every meeting and draft a project plan that has everyones support.	$2 \times 4 = 8$	Non-Technical
Group does not make it to the competitions next round.	Make adjustments if necessary and continue development.	$2 \times 0 = 0$	Non-Technical

5 Project Management

6 Technical Considerations

6.1 Sentiment Analysis

Once the user's online content has been extracted, it will then have to be placed into one of the following categories: negative, positive, neutral. This can be achieved by performing sentiment analysis on the content, which first has to be trained by supplying training data.

The language used for the sentiment analysis will be Python, along with the Natural Language Toolkit (NLTK) Python library.

6.1.1 Step 1: Data collection & pre-processing

The first step is to collect existing positive, negative and neutral content and store them in an array.

```
positive_content = [
    ("I love this car", "positive"),
    ("This view is amazing", "positive"),
    ("I feel great this morning", "positive"),
    ("I am so excited about the concert", "positive"),
    ("He is my best friend", "positive")
]

negative_content = [
    ("I do not like this car", "negative"),
    ("This view is horrible", "negative"),
    ("I feel tired this morning", "negative"),
    ("I am not looking forward to the concert", "negative"),
    ("He is my enemy", "negative"),
    ("John is so annoying", "negative")
]

neutral_content = [
    ("He is alright", "neutral"),
    ("Today's weather is just right", "neutral"),
    ("She is alright", "neutral")
]
```

Figure 7 Collect Content

These words are then collected into a single list of tuples, each of which containing two elements.

```
content = []

for (words, sentiment) in positive_content + negative_content + neutral_content:
    words_filtered = [e.lower() for e in words.split() if len(e) >= 3]
    content.append((words_filtered, sentiment))
```

Figure 8 Content Preprocessing

6.1.2 Step 2: Classifier

A list of each word extracted from all the content needs to be collected and then ordered based on frequency of occurrence. This can be done by initially collecting all words and associating a frequency of occurrence to each and then ordering the list based on the frequency value.

```
def get_words_in_content(content):
    all_words = []
    for (words, sentiment in content):
        all_words.extend(words)
    return all_words

def get_word_features(wordlist):
    wordlist = nltk.FreqDist(wordlist)
    word_features = wordlist.keys()
    return word_features

word_features = get_word_features(get_words_in_content(content))
```

Figure 9 Word Frequency

To create the classifier, relevant features needs to be captured via a feature extractor. Below shows the function implementation of this.

```
def extract_features(document):
    document_words = set(document)
    features = {}
    for word in word_features:
        features['contains(%s)' % word] = (word in document_words)
    return features
```

Figure 10 Feature Extraction

A training set can then be created using the nltk library and a classifier object can be instantiated.

```
training_set = nltk.classify.apply_features(extract_features, content)
classify = nltk.NaiveBayesClassifier.train(training_set)
```

Figure 11 Training Classifier

6.1.3 Step 3: Classify

Now the classifier has been created and trained, the sentiment analyser can be tested.

```
content = "You're so annoying"
print classifier.classify(extract_features(content.split()))
# $ negative

content = "I really love you"
print classifier.classify(extract_features(content.split()))
# $ positive
```

Figure 12 Classify

7 Derp

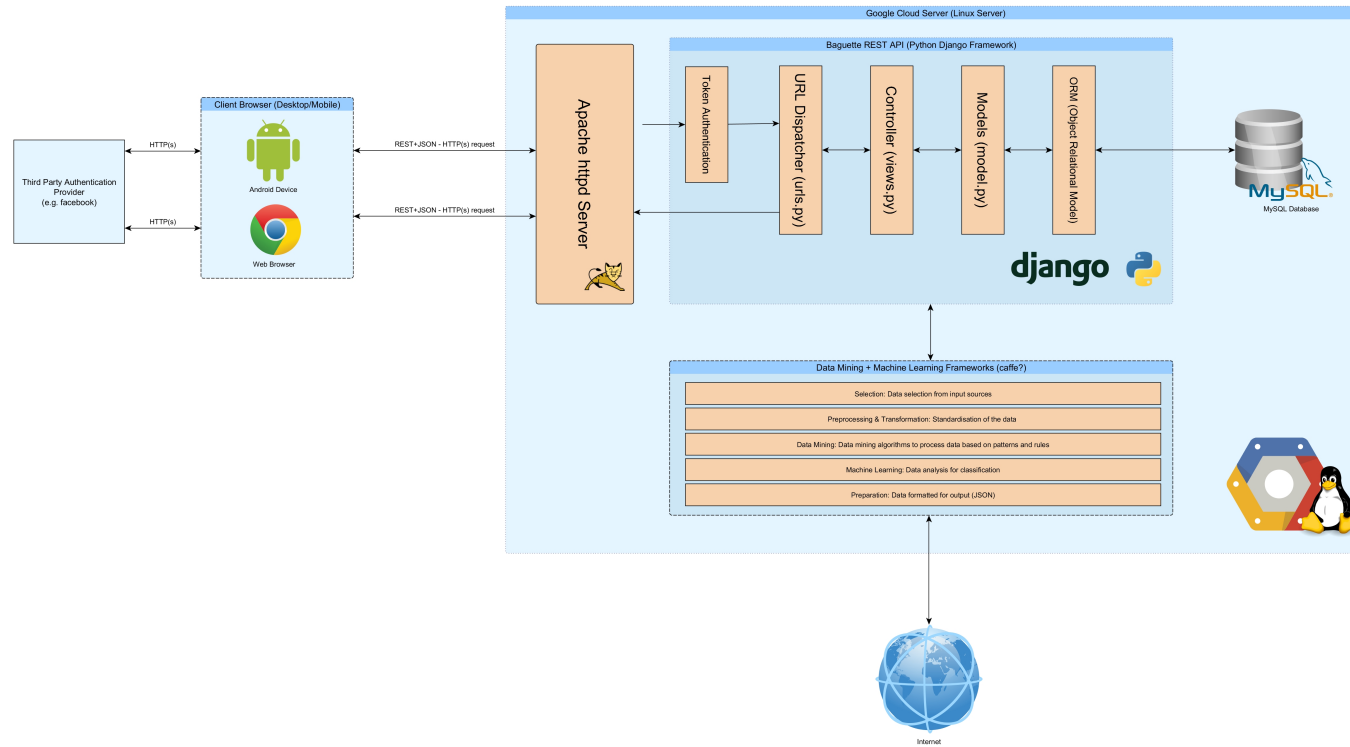


Figure 13 derp

8 Ideas

8.1 Overview

Breadcrumb, the title of group's proposed solution, is an impression management tool that helps individuals and corporates positively manage and optimise their online information. The name comes from the story of Hansel and Gretel, who left a trail of breadcrumbs as they walked through the forest so they could trace their way home. In the same way, the group noted individuals and corporates often leave a digital footprint that makes it easy for others to develop an opinion about their reputation from what they see online. Unlike footprints left in the sand at the beach, an online "breadcrumb trail" often sticks around long after the tide has gone, and can have negative implications for all involved.

8.2 Business Model

Breadcrumb will come in the form of three-tiered structured offering, which will provide clients with a customised level of service at different price points in order to appeal to a wider segment.

8.2.1 Tier 1: Personal

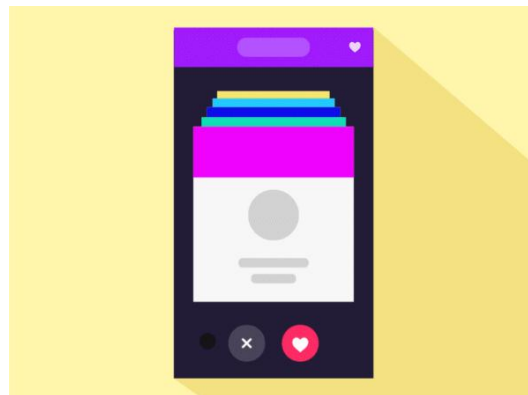


Figure 14 Flipping Card User Interface

The Personal offering is geared towards individuals wanting to monitor and control their online information – for instance, prior to making a job application. In this case, when Breadcrumb is run for the first time, the user will be asked to connect all their social media accounts (i.e. Facebook, Instagram and Twitter) and provide their personal information. Breadcrumb then scans each social media account for content it initially considers to be significantly positive or negative and collates a primary list. Further to this, it trawls the internet to find material that it initially considers to be significantly positive or negative and collates a secondary list. As Breadcrumb is an impression management tool, it is important for it to understand the entirety of someone's online profile, so that it can recommend ways in which the user can improve their digital footprint. Once both lists are collated, they are combined and presented to the user as a "graphical deck of cards". Each card contains information on the content Breadcrumb identified as significantly positive or negative. At this point, the user can swipe through each card and decide whether they personally deem the content to be positive or negative. Consequently, Breadcrumb is trained to understand what the user wants

their online profile to look like, so that in future scans, it will get better at flagging relevant material. Over time, this increases Breadcrumb's value or stickiness, making it rather difficult for users to switch.

As the group expects the Personal offering to be used infrequently, it will employ a freemium pricing strategy for this portion of the service. As this service is new to market, offering a freemium version to individual users can be a quick way to drive trial usage and generate feedback before scaling up because of the low barrier to entry.

In order to convert free users into paying users, Breadcrumb will only display a limited number of results but charge users a small fee in order to unlock the rest. Moreover, Breadcrumb will regularly provide the user with recommendations on how to improve their online profile, which can be unlocked for another small fee. These recommendations will be derived from what Breadcrumb understands about the user and may involve ideas around content production. All of these features will be presented to the user in a gamified manner, often with the promise of "rewards" that motivate the user to keep returning due to their perceived fairness of the transaction. Moreover, such a system will also give the user a sense of control over their continued investment.

8.2.2 Tier 2: Premier

The Premier offering is geared towards individuals wanting to understand and shape public opinion, create compelling narratives and develop initiatives that persuade people to act – for instance, a agent aiming to maximise the market value of their client (in this case, a football player) would want to have them conveyed in the media as a skilful player in order to inspire a bid from a particular football club.

The Premier offering contains the same features as the Personal offering, but also provides users the opportunity to review multiple online profiles at once and derive a sentiment score which measures the emotional tone of the information found, along with current public perception. This allows for a variety of interesting use cases – for example, a spin-doctor coordinating a political campaign in the build-up to a presidential election can compare the sentiment score between two competing candidates in order to better understand how the public is reacting to their clients recent activities.

Further to this, Breadcrumb can provide the spin-doctor or agent with a unique, in-depth impression management strategy that can help improve their clients sentiment score and overall public image. The group envisions that the Premier offering could also be used by corporates seeking to maintain the online reputation of their executives, and by celebrities in order to identify instances of copyright infringement, wherein their work or likeness have been used without prior permission.

As the group expects the Premier offering to be regularly used, it will employ a subscription pricing strategy for this portion of the service. Users have the chance to use Breadcrumb free for 30 days in order to determine whether it is suitable for them, but beyond that they will need to pay a monthly fee.

8.2.3 Tier 3: Enterprise

The Enterprise offering is geared towards small, medium and large corporates wanting to aggregate relevant content, analyse the sentiment of each mention and derive information about the individuals hosting discussions about the brand. Further to this, Breadcrumb will provide corporates with tailored, brand amplification strategies. The group also envisions that the Enterprise offering could also be used by corporates screening job applicants, subsequently providing human resource departments

with insights into whether or not an individual is a fit for the organisation based on their online profile.

As the group expects the Enterprise offering to be regularly used, it will employ a subscription pricing strategy for this portion of the service identical to the Premier offering, as it builds upon that package.

9 Stakeholder Feedback

In order to determine the business viability of the proposed ideas, the group drafted a proposal and sent it to multiple stakeholders in an attempt to get feedback. Their responses are as follows:

9.1 Stakeholder 1

I think your solution is rather complex. Implementing context/grammar checking and analysing the intent of language statements sounds very difficult. How are you going to go about doing it?

Why not simply provide a service that will simply scan the Internet for all traces of your existence through your multiple identities and then automatically remove them. I think that is a sellable service. The Breadcrumb concept has the connotations of a CV service which still requires manual intervention to make things fit. I cannot see this being automated as I think you are hoping for.

9.2 Stakeholder 2

I gave this a read and found it really interesting. Some of the examples have some relevance to me so I really liked the idea. As someone coming to the business proposal with completely fresh eyes, I'll try to highlight the areas I particularly liked and the areas that needed a bit more explanation.

9.2.1 Overview and Business Model

Firstly, I really liked the name and the explanation for this. I think that you make a really good link to technology and the internet. With regards to the business idea, I liked the idea of having packages for different users. I also liked the pricing structure proposed, along with the add-ons available.

9.2.2 Personal Offering

I understand that this is geared at someone like me that wants to understand and control their online presence on social media? I think that here you probably need to expand on the selling point. Why would I really need this? There needs to be a reason why this will really benefit an ordinary person like me. For example, if I wanted to improve my presence and performance on social media, maybe the app would help to collate my previous posts which have been most successful or received the most likes so that I know what style of posts are popular and where my profile has been 'positive'.

With the idea of it creating 2 lists, I don't understand the difference between the primary and secondary list that the app creates. I don't understand the difference between the two lists or the need for these.

In addition, I think you could perhaps explain a bit more how it decides what information is significantly positive or negative. Does it scan for trigger words or consider the number of likes?

9.2.3 Premier Offering

I really liked the celebrity example and the idea of them wanting to maintain a reputation or check for copyright. I think this example differentiates it from just an advanced search engine. I feel there is danger in the other examples that it is just a search engine. I could be wrong. I think with the football agent situation you need to again tighten up the reasons why they are using it and the reason it would benefit them. Are they using it to find out about what is said on a player or to find information to put together a profile? Again, is this just a search engine? What makes it less of a search engine is the 'sentiment score' feature. How does it pick up sentiment? Again, is this from trigger words? You might want to explain.

9.2.4 Enterprise Offering

I like this tier. My sister is setting up her own business so when I think about the app from her perspective, I can see how it would be useful. I think it would be good for a small business to be able to see where people are talking about your brand and the 'sentiment' of their feedback.

I think that in this tier, there should be differentiation between small businesses and larger organisations using it as HR screening. The idea of large organisations using this app is huge and a separate tier to small businesses. And, are you suggesting that we should be searching someone's social media when they apply for a job? That's scary!

9.3 Stakeholder 3

- It's an impression management tool which, if I'm correct, aims to alter the perception others have of you based on what can be found on the internet. Could the software consider the presentation of statistical data of regularity at which the deemed positive and negative situations appear to subconsciously make the user more conscious of their online behaviour. I suggest this because if people try hard enough the information can still be found online so to reduce this reduces the risk.
- I quite like the name Breadcrumb. Great marketing by the team here. Fairy tales are generally told to children, however, the idea relates a story known by many, from the young to the old, so makes the purpose of the product easy to understand for all and applicable to the majority. I like how it takes present day issue where people are becoming more aware of the digital footprint.
- Not sure about following statement: Breadcrumb then scans each social media account for content it initially considers to be significantly positive or negative and collates a primary list. Further to this, it trawls the internet to find material that it initially considers to be significantly positive or negative and collates a secondary list. It appears to be a very subjective statement. Can the specification as to what this is be initiated by the user or will this be purely general and on the basis of parameters set by the app?
- It's not been stated what is then done with the information. What does the group actively do to ensure the better presentation i.e. ensuring the information cannot be found on web searches?
- With regards to the Premier offering, how would you ensure that the sources of review were unbiased, relevant and respected?

9.4 Stakeholder 4

I love the name - seriously cool. I'm not sure what stage in the development you're at, but I'm not entirely sure how the application will allow users to be forgotten' from what I'm reading in the document. This may be intentional, but thought it was worth mentioning.

9.5 Stakeholder 5

- Have you thought about scanning LinkedIn profiles for the Enterprise offering?
- Will the application be available on desktop? In the example of a HR department using it to screen job applicants this may incur further IT costs if it can only be used on smartphones.
- For the enterprise offering are you able to ensure corporates that their data will be protected? I think this would be important with investment banks as regulators are on everyone's back.

9.6 Stakeholder 6

It's difficult to understand the landscape that this has been set in a business plan would need financials, market research, action plan etc., which I assume is not part of the scope of this, so I'll ignore. If that isn't the case, let me know and I'll try to help. Also, there is no mention of any challenges foreseen with the implementation of this along with how they would be overcome. This would tell the prospective stakeholders that you are realistic with expectations and have thought through the proposal.

Also, the style of writing is individual I personally get frustrated with the marketing spiel, wanting a more direct black and white description, but I doubt the judges are looking for the same style. I had to read several times to fully understand the idea.

Having said that, I like the idea. I don't understand the need to capture the personal account details, as (I understand) the added value is identifying what is in the public domain, and logging in to a restricted account is only looking at a restricted area.

I think you need to think about the added value for example, ability to automatically request removal of information, subscribe to credit checking or police checking etc., extending the checking to contacts, memberships, directorships everything that would give confidence that an employer check (both ways, for employee and employer) is comprehensive.

Full corporations would have teams of people responsible for preserving the company's image the growth would be for smaller companies who do not have the resources to do it themselves.

9.7 Stakeholder 7

- Depending on the exact brief, it could be worth adding figures, as a means of providing rough estimates over the commerciality of your app. If the brief said short business plan, highlighting commerciality could score you extra brownie points. Possibly expanding upon the potential revenue generated by each tiered offering. Whereas if the brief simply says create an app, then commerciality won't be needed. Furthermore if it is a business plan and you're worried about word limit, simply make a table outlining your estimates, and then stick it in as an image.

- The brief states, that your app plan should cover the perspectives of the app from the view of different stakeholders. I don't feel your current plan directly addresses these stakeholders. I advise that you include a stakeholder perspective section in your finalised business plan e.g.:
 - An individual - better understanding of their digital footprint
 - A citizen - collate public records on them, and exercise right to be forgotten if old websites keep coming up, like old social media pages
 - An employee - improve and clean up their digital footprint to make themselves more employable

10 Mathematical Algorithms

This section contains information on the mathematical algorithms that could be utilised in the solution in order to help individuals devise a strategy for which data linked with them could be used for.

10.1 *Pattern Matching*

10.1.1 Description

The bitap algorithm is an approximate or exact string matching algorithm that is one of the underlying algorithms of the UNIX “agrep” utility. It determines whether a given text contains a substring which is equal to a given pattern, where approximate equality is defined in terms of the Levenshtein distance – a complementary algorithm which determines how many changes must be made to a string or phrase in order to turn it into another string or phrase. In comparison to other algorithms, bitap does most of the work with bitwise operations, thus making it run super-fast.

Figure 15 Bitap Algorithm Pseudocode

10.1.2 Application

The solution will require access to two predefined datasets for classifying the information returned by search engine. The first dataset contains some sensitive keywords that will make a significant influence on public impression, such as drug abuse or achievement. The other dataset stores the words with a neutral, negative or positive emotion, and each word has a corresponding score, for example: disappointment with score -5, humiliation with score -8, honour with score +8 and success with score +6.

The steps for implementing this algorithm are as follows:

1. Search website articles for sensitive keywords and their relative location
2. Search for “emotional words” that appear within the spatial locality of the sensitive keyword
3. Calculate the score of the article or website based on the number of emotional words found

10.2 *Image Recognition*

10.2.1 Description

Affine scale-invariant feature transform (or ASIFT) is a computer vision algorithm that is used to detect and describe local features in images. It is regarded as an efficient method for determining matches between two arbitrarily-selected points within two widely separated views. Moreover, ASIFT is capable of finding a correspondence, even for pixels within an area with certain uniform properties, such as similar colours and textures.

Figure 16 Comparison of Image Recognition Algorithms

10.2.2 Application

During the text processing stage, a pattern matching algorithm can attempt to detect the purpose of an article or website. If it is determined that the page is an advertisement, then the image and face recognition algorithm can attempt to match the image with the relevant person's face. Once matched, a check can be carried out to determine whether or not the website has the right to use the portraiture.

11 Literature Review

The concept of the “*right to be forgotten*” is based on the European ideal of an individual determining the development of their own life in an autonomous manner, without being periodically stigmatised as a consequence of a specific action performed in the past, especially when these events do not have any relationship with the contemporary context (Mantelero, 2013). However, does the “*right to be forgotten*” really have a sound basis? Mayes (2015) argues that the right to being forgotten is a figment of our imaginations, describing it as an antisocial, nihilist act which could eventually signify the degradation of our power to act in the world. Fleischer (2015) extends this argument by asking the difficult question: who should be responsible for what should be remembered or forgotten? For example, if Italian courts decide that Italian murderers should be able to delete all references to their online convictions after a period of time, would this Italian standard apply to the entire Web or only be applied to *.it domains? Under the ruling in Article 12 of the Directive 95/46/EC, Europeans who feel that they are being misrepresented by search results that are no longer accurate or irrelevant can ask Google to delink the material. If the request was approved, the information would remain online at the original site, but no longer come up under certain search engine queries (Manjoo, 2015).

It can be argued that certain ideas pertaining to the “*right to be forgotten*” are synonymous with those pertaining to self-presentation and self-disclosure. Goffman (1959) ideas on self-presentation asserts the ways in which an individual may partake in strategic activities “*to convey an impression others which it is in their best interests to convey*”. Online self-presentation is more malleable and subject to self-censorship, allowing individuals to express or make salient multiple aspects of their identity. Consequently under certain conditions, individuals may wish to express themselves more openly and honestly than in face to face contexts (Ellison, Heino, and Gibbs, 2006) or engage in misrepresentation.

Within the assigned context, the group will need to consider incorporating impression management features in order to create a marketable commercial solution. Impression management can be defined “*as the goal-directed activity of controlling information about a person, object, entity, idea or event*” (Connolly-Ahern and Broadway, 2007). As discussed above, the group considers impression management to be an expansion of the “*right to be forgotten*” concept, as it presents an opportunity for individuals and corporates to correct any misrepresentations of themselves online. Relationships are increasingly initiated and maintained online, providing individuals and corporates with an opportunity to create strategic images of themselves for social purposes (Rosenberg and Egbert, 2011).

Moreover, according to Pak and Paroubek (2010) it creates opportunities for data extracted from these sources to be used in opinion mining and sentiment analysis. Sentiment analysis can be defined as “*the computational study of opinions, sentiments and emotions expressed in texts*” (Liu, 2010). For example, during an election campaign, political parties may be interested to know if people support their candidate or not by analysing messages written across multiple social media networks in order to extract an aggregated sentiment which can be used to support decision-making. However in this case, complications may arise as the disclosure domain is boundless and consequently all messages will need to be filtered as they can refer to any subject (Ortigosa, Martin, and Carro, 2014).

References

- Ancona, D. and H. Bresman (2007). *X-teams: How to Build Teams That Lead, Innovate and Succeed*. Harvard Business School.
- Bloomberg (2015). *ATOS SE (ATO:Euronext Paris): Stock Quote & Company Profile – Businessweek*. URL: <http://www.bloomberg.com/research/stocks/snapshot/snapshot.asp?ticker=ATO:FP>.
- Boehm, B. (1981). *Software Engineering Economics*. Prentice Hall.
- Cohen, G. (2010). *Agile Excellence for Product Managers: A Guide to Creating Winning Products with Agile Development Teams*. Superstar Press.
- Connolly-Ahern, C. and C. Broadway S (2007). “The importance of appearing competent: An analysis of corporate impression management strategies on the World Wide Web”. In: *Public Relations Review*, pp. 343–345.
- Dawson W, C. (2009). *Projects in Computing and Information Systems*. 2nd ed. Addison Wesley.
- Ellison, N., R. Heino, and J. Gibbs (2006). “Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment”. In: *Journal of Computer-Mediated Communication*.
- Fleischer, P. (2015). *Foggy thinking about the right to oblivion*. URL: <http://peterfleischer.blogspot.co.nz/2011/03/foggy-thinking-about-right-to-oblivion.html>.
- Git (2015). *Git – About Version Control*. URL: <https://git-scm.com/book/en/v2/Getting-Started>About-Version-Control>.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Anchor Books.
- Highsmith, J. and A. Cockburn (2001). “Agile software development: the business of innovation”. In: *Computer*, pp. 120–127.
- Jean Steven, S. and D. Brady (2014). *Professional Team Foundation Server 2013*. John Wiley & Sons.
- Jee, C. (2015). *Transport for London awards Atos 6 million service desk contract*. URL: <http://www.computerworlduk.com/news/it-management/transport-for-london-awards-atos-6-million-service-desk-contact-3596762/>.
- Kawalerowicz, M. and C. Berntson (2011). *Continuous Integration in .NET*. Manning Publications.
- Liu, B. (2010). “Handbook of Natural Language Processing”. In: ed. by N. Indurkha and F. Damerau J. Chapman and Hall. Chap. Sentiment Analysis and Subjectivity, pp. 627–661.
- Manjoo, F. (2015). *‘Right to Be Forgotten’ Online Could Spread*. URL: <http://www.nytimes.com/2015/08/06/technology/personaltech/right-to-be-forgotten-online-is-poised-to-spread.html>.
- Mantelero, A. (2013). “The EU Proposal for a General Data Protection Regulation and the roots of the ‘right to be forgotten’”. In: *Computer Law & Security Review*, pp. 229–235.
- Mayes, T. (2015). *We have no right to be forgotten online*. URL: <http://www.theguardian.com/commentisfree/libertycentral/2011/mar/18/forgotten-online-european-union-law-internet>.
- North, D. (2015). *Behaviour-Driven Development*. URL: <http://behaviourdriven.org/BehaviourDrivenDevelopment>.
- Ortigosa, A., J. M. Martin, and R. M. Carro (2014). “Computers in Human Behavior”. In: *Sentiment analysis in Facebook and its application to e-learning*, pp. 527–541.
- Osherove, R. (2009). *The Art of Unit Testing: with Examples in .NET*. Manning Publications.
- Pak, A. and P. Paroubek (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. Tech. rep. LREC.
- Reuters (2015). *Atos SE (ATOS.PA)*. URL: <http://www.reuters.com/finance/stocks/overview?symbol=ATOS.PA>.
- Rizwan M.J., Q. and S. Hussein (2008). “An adaptive software development process model”. In: *Advances in Engineering Software*, pp. 654–658.

- Rosenberg, J. and N. Egbert (2011). “Online Impression Management: Personality Traits and Concerns for Secondary Goals as Predictors of Self-Presentation Tactics on Facebook”. In: *Journal of Computer-Mediated Communication*, pp. 1–18.
- Scott W, A. and M. Lines (2015). *Introduction to Disciplined Agile Delivery: A Small Team’s Journey from Scrum to Continuous Delivery*. CreateSpace Independent Publishing Platform.
- Sharma, G. P. et al. (2015). *Unit testing, model validation, and biological simulation*. Tech. rep. Cornell University.
- Techopedia (2015). *What is Scrum Sprint? – Definition from Techopedia*. URL: <https://www.techopedia.com/definition/13687/scrum-sprint>.