D2.2

Requirements Analysis

v 1.2 / 2015-04-13

Noel Catterall (HWC), Symeon Papdopoulos (CERTH), Adrian Popescu (CEA), Timotheros Kastrinogiannis (VELTI)

This document defines the requirements of the technical aspects the USEMP software must adhere to. These requirements are determined from a user scenario role, technical aspects of the implementation and functional aspects of individual elements.

Project acronym	USEMP
Full title	User Empowerment for Enhanced Online Presence Management
Grant agreement number	611596
Funding scheme	Specific Targeted Research Project (STREP)
Work program topic	Objective ICT-2013.1.7 Future Internet Research Experimentation
Project start date	2013-10-01
Project Duration	36 months

Work Package	WP2
Deliverable Lead Org.	HWC
Deliverable Type	Report
Authors	Noel Catterall (HWC) Symeon Papadopoulos (CERTH) Adrian Popescu (CEA) Timotheros Kastrinogiannis (VELTI)
Reviewers	Tom Seymoens (iMINDS) David Lund (HWC)
Version	1.2
Status	Final
Dissemination Level	PU: Public
Due date	2014-04-01
Delivery date	2014-05-01
Revision date	2015-04-13

Version	**Changes**
0.1	Initial Draft Template
0.2	Partner Contributions
0.3	Final Draft
1.0	Final
1.1	First revision after EC review
1.2	Revised release after EC review

Table of Contents

1. Executive Summary

This document defines the requirements that the implementation of the USEMP system should adhere to. These requirements are defined from the use case perspective, i.e. requirements relating specifically to each defined use case, the system perspective, i.e. requirements relating to the design of the system as a whole, and the functional requirements, i.e. requirements relating to individual components of the system.

The document defines the methodology used in order to define the system requirements, arising from the use case scenarios outlined within D2.1; moreover functions available from technical partners are listed with the constraints on their use.

This document will constitute the initial setting from which the overall architecture design will be derived.

This document will be updated throughout the duration of the project accounting for further requirements as the need arises.

2.Requirements Design Methodology

Prior to work starting, a number of technologies that are available to be provided by consortium partners were outlined and use cases built around these technology concepts. These were constituted as tech card definitions, which allowed for all members of the project to understand the capabilities of the technology, how such technologies could be used to extract information about the user, and legal and social issues that could be derived from such extraction.

Generation of these tech cards lead to the creation of scenarios and then onto user stories as defined within deliverable D2.1. These user stories were discussed and elements of the feasibility of what elements are possible with the technology, and what is possible with regard to integration within OSNs was discussed.

Teleconferences were established in order to map these user stories to both requirements placed upon individual functions, and requirements placed on backend components and integration within OSNs in terms of technically feasible components, and where appropriate, specifications for a mock-up of components that cannot be placed within the existing eco system of API availability provided by OSNs.

The two user stories defined were:

OSN Empowerment Tool

A use case to give OSN users control and awareness of the content they are sharing online, whether that be observed or inferred from their data. This will aim to flag up sensitive content, and allow the user to control the sensitivity level of their exposed content.

Information Value Awareness Tool

This tool will allow a user to gain insight into the relative value of their personal data, such as the relationship with their profile to that of advertisers and OSN network operators.

In total three different software tools will be created within the USEMP project to address the above scenarios:

- For user data collection:
    - a browser (e.g. Firefox) plug-in (e.g. DataBait Plug-in)
    - an OSN-enabled application (e.g. DataBait Facebook App)
- For enabling users to access USEMP features and services:
    - a web application (e.g. DataBait Web App)

3.User Stories

The following constitutes the user stories as defined within deliverable D2.1. This section will outline the user requirements for the workflow, which will then be mapped to technical system requirements. These stories are already present within D2.1 in terms of feasibility, but are annotated with references here for mapping to technical system requirements.

3.1. OSN Empowerment Tool

The user searches online for a presence control tool	Feasible
The user finds the DataBait tool online	Feasible
The user reads about the DataBait tool online	Feasible
The user downloads the DataBait data collector plug-in on her computer (DataBait plug-in for free)	Feasible
The user installs the plug-in in all her web browsers (Chrome, Firefox) on (all) her computer(s) (Windows OS, MAC)	Feasible, but initially target one browser, one OSN (resource constraint)
The user accesses the DataBait web App and logs-in to her DataBait account (if a new user then she/he creates a new DataBait account).	Feasible
The user logs-in via DataBait web App (which also is an OSN-enabled app) with her OSN(s) account (e.g., FB).	Feasible
The user gives permission to DataBait to access her OSN (e.g., FB) profile and Graph API information.	Feasible
The user accesses the DataBait Mobile Web App on her smart phone (Android, iOS)	Feasible
The user logs in on DataBait Mobile Web App from her smart phone	Feasible for visualization and interaction but not for the browsing data collection
The user arrives at the homepage of DataBait. She sees a high-level	Feasible

visualization of her personal data trails	
The user zooms in on the visualization of her personal data trails, being able to filter the latter in various dimensions (e.g., time, sensitivity, etc.)	Feasible
The user sees the different companies and organizations (that FB has a contract with) that use/have access to her observed, behavioural and inferred data in a visual attractive way	Only possible with synthetic data or unless Facebook is forced to give us access to the list of third parties they work with
The user swipes the screen in order to navigate through DataBait features such as: ‘Profile’, ‘PD control’, ‘Future Control’, ‘Settings’, …	Feasible
The user selects ‘PD control’	Feasible
The user is able to visualize her digital trail in intuitive info graphics	Feasible
The user gets a list of different types of Privacy-sensitive Dimensions: Sexual orientation, Political preferences, Religion, etc.	Feasible
The user selects ‘sexual orientation’ and gets an overview of a) a probabilistic estimation of profiles used by different actors to target or accommodate her and potentially b) different categories of institutions/organisations that may be interested in such privacy-sensitive dimensions.	Feasible for point a), not feasible for point b) unless Facebook is forced to give us access to the list of third parties they work with
The user sees which parties that have access/use her personal data are also tracking (profiling) on this privacy-sensitive dimension	Not feasible at the moment. Would become feasible if Facebook is forced to give us access to the list of third parties they work with
The user can -for every party that have access/use her personal data- dis/enable to get tracked with respect to that privacy-sensitive dimension	Under current regulation, feasible as a simulated functionality. The other option is for the user to remove/change visibility of content for each item
At the bottom she also defines to apply these settings as generic policies that will affect (be applied) all future	Under current regulation, feasible as

institutions/organisations belonging to one of the categories.	a simulated functionality.
The user clicks on the home button and returns to the a high-level visualization of her personal data trails	Feasible
The user clicks on the ‘profile’ button	Feasible
The user gets insights of the different companies and organizations that are profiling her	only possible as mock-up
The user clicks on the icon of the shop named ‘Colruyt’ (where she does her weekly shopping)	only possible as mock-up
The user sees the probabilistic estimation of the way she is profiled by ‘Colruyt’	only possible as mock-up
The user wants to change what personal data is available for Colruyt and clicks on the ‘Data control’ button	only possible as mock-up
The users sees an overview of the different options she has to control her digital trails: permission revocation, PD data removal, Copy settings	only possible as mock-up
The user clicks on ‘Permission Revocation’ and sees a graph of what personal data she is sharing with Colruyt	only possible as mock-up
The user creates in a visual programming way rules (privacy policies) to define what type of information may be used by Colruyt and under what conditions.	only possible as mock-up
The user can choose in the next window which data may be available for Colruyt	only possible as mock-up
The user returns to the estimates profile by Colruyt	only possible as mock-up
The user now sees how her estimated profile has changed	only possible as mock-up
The user clicks on the home button and returns to the visualization	Feasible
The user clicks on ‘Show permissions’ and sees the rules that she created related to the use of her personal data	Feasible
The user selects one rule and changes it to another setting (e.g. from ‘free use’ towards ‘free use for non-commercial’	Feasible
The user saves the rule	Feasible
The user looks at the overview of the rules	Feasible

The user clicks on ‘future control’. Here she can allow DataBait to notify her when she’s about to release sensitive information of her choosing (Auditing)	Feasible
The user creates in a visual programming way rules to define when she wants to get notifications (real-time) when her online behavior influences the way she is profiled on different privacy-sensitive dimensions	Feasible
The user creates rules (where, when, how long the rules have to be active)	Feasible
The user saves the rules	Feasible
The user checks if the rules are saved in the defined way	Feasible
The user submits a picture on FB	Feasible
The user sees a pop-up, in a non-intrusive manner in terms of frequency, from DataBait that tells her she is going to submit information that will influence one of her privacy dimensions (e.g., sexual orientation), based on privacy priority, sensitivity and importance rules	Feasible
The user is proposed to select between ‘not submitting the post’, ‘obscuring the post’, ‘post picture from other application’	Feasible
The user selects ‘obscure the post’ and submits the picture	Feasible
The user looks at how the picture was submitted and is happy	Feasible

3.2. Value Awareness Tool

The user arrives at the homepage of DataBait web application and logs-in with his DataBait credentials. He sees a high-level visualization of his personal data trails.	Feasible
The user zooms in on the visualization of his personal data trails, being able to filter the latter in various dimensions (e.g., time, sensitivity dimensions, etc.)	Feasible
The user is able to see the estimations of profiles (and/or profile segments/categories) he is placed into by different actors in the network.	Feasible with simulated actors
Another feature he likes is the games that he can play. By playing the game he helps people who want to obscure their post, or cuts out part of the picture. He earns points with it.	Feasible in a simple manner
Another game is a quiz where he can guess what 3rd parties could infer from his online data. By playing the game he learned about his digital traits and could potentially implicitly intervene in the way he was profiled by others.	Feasible
His friend has sent him a request for re-use of his privacy configuration settings. As he is known as somehow ICT skilled and privacy aware, friends wanted to take over his configurations. He accepts the request.	Feasible
On the Homepage of the DataBait Web App the user clicks on ‘Your personal data value’	Feasible
The user sees enhanced DataBait visualisation via which a) he could gain useful insights on the value of his digital data and social footprint that he either directly shared in social networks (e.g., Likes of FB) or were indirectly collected by various network actors that track his activities on his web browser.	Feasible with simulated actors
He looks at it and requests more insights	Feasible
The user is shown some profile categories that DataBait thinks he is interested in (e.g., brands or activities). The user can click on these categories and delete specific profile attribute topics or acknowledge/refine his interest in a topic.	Feasible with simulated actors
The user can also search on the brands and topics he is interested in to find his personal interests.	Feasible with simulated actors
The user selects the brands and topics he is more interested in.	Feasible with simulated actors
The user gets presented a list of possibilities to validate his data: brand ambassadorship, scientific research, citizen engagement.	Feasible with simulated actors
The user clicks on Brand ambassadorship and gets presented again with his interest lists: technology, music, sailing, whiskey, photography, clothing, etc.	Feasible with simulated actors

The user clicks on whiskey and sees the brands that have contacted DataBait to get access to its database	Feasible with simulated actors
The user sees whiskey brands and can choose one for which he can become a brand ambassador	Feasible with simulated actors
The user highlights Johnny walker	Feasible with simulated actors
The user returns to the Brand Ambassadorship page and sees that he can still become BA for two more brands of his choice	Feasible with simulated actors
The user clicks on cameras and is delighted to see that he can become a BA for his favourite brand: Sony.	Feasible with simulated actors
The user clicks on ‘Profiles’	Feasible
The user gets an overview of different topics on which he might be tracked (Socio-demographics, Personality traits, Interests, …)	Feasible
The user is very interested to see how his socio demographics are estimated and how close they are to the reality and clicks on this topic.	Feasible
He gets estimated insights for his age, gender, ethnicity, nationality, sexual preference, professional background, political preference, …	Feasible with the features defined in WP6
He returns to the previous screen and clicks on Personality traits	Feasible
He gets estimated percentages for openess, neuroticity, …	Feasible with the features defined in WP6
The user returns to the Brand Ambassadorship page and sees that he is now presented with two new buttons: one for each brand.	Feasible
The user clicks on the Sony Camera-button	Feasible
He is presented with the value of his BA for Sony Camera’s, a settings button, …	Feasible
He can enter his email address in order to receive surveys where he can participate in, advertisements for workshops on Sony camera’s, vouchers for a free SD card.	Feasible in principle but the simulation will no go as far as that. Beyond the immediate scope of USEMP.
The user returns to the homepage of the DataBait-web application	Feasible
The user clicks on PI control	Feasible
The user clicks on Personal Data Licensing	Feasible with simulated actors
He gets a list of actors that might be interested in his data: non-profit organizations, research institutions, commercial organizations	Feasible with simulated actors

He clicked on the Open Knowledge Foundation NGO	Feasible with simulated actors
He decided that they could use the sensor data from his smart phone whenever he was connected to a Wi-Fi network with it for free and returned to the previous screen.	Not feasible
He clicked on the research institution on Alzheimer disease and there he configured that they always ask him to share certain types of information if they needed some	Not feasible
He clicked on Commercial organizations	Feasible with simulated actors
The user gets an overview of different topics on which these organizations might find useful (Socio-demographics, Personality traits, Interests, …)	Feasible with simulated actors
The user clicks on socio-demographics	Feasible with simulated actors
He gets estimated insights for his age, gender, ethnicity, nationality, sexual preference, professional background, political preference.	Feasible with simulated actors
He decides for each of these traits for which actors (that have a contract with DataBait) they could become available and looks at how much value this data received.	Feasible with simulated actors
When the user encounters his dad, his dad tells him he was spammed with ads for spirits and that he didn't want this.	Feasible with simulated actors
The user browses to the Personal Audience Management Panel of the DataBait Web App.	Feasible with simulated actors
He gets graphics on who is influenced by his posts, how his audience evolves, the segmentation of his audience. He sees his father was divided in the group of 'having interest in drinks'.	Feasible with simulated actors
The user takes his father out of this group.	Feasible with simulated actors
The user uploads a picture with a bottle of Johnny Walker	Feasible
DataBait user gets a pop-up from DataBait that this was his x-th post about whiskey in one week and that this commodification of his PI was not being encouraged through more value.	Feasible

4.Use Case Perspective

This section will lay out requirements specific to each use case. Requirements pertaining to individual functionality or system operations will be outlined within the latter sections as such requirements may have features relating to both use cases. This will focus on user requirements derived from the previous section, whereby specifications are drawn from elements marked as feasible for implementation.

User interactions are recorded in the previous section as to actions a user is to perform, so this section will focus on technical elements in order to perform primary functions.

4.1. Use Case 1: OSN Empowerment Tool

In terms of the OSN empowerment tool, the system must allow the user to interact with the system, and then pull the user's OSN data. For initial implementation Facebook will be chosen due to be being the largest social network, and providing typically more private information shared with friends, rather than networks such as twitter, for which data tends to be shared more publicly.

The system shall be able to access the users of OSN(s) and be able to process all data connected to such networks. This will focus on textual and imagery data for the capability of determining brands the user is interested in, and any privacy related actions such as smoking or other 'anti-social' behavioural actions.

The system may be able to process the information within one second such that the user can make informed decisions on their past data without long delays. In the event data processing is to take longer, a progress bar should be presented. A maximal extent of 10 seconds will be aimed for. This is in order to provide for a acceptable user experience. It may be the case dependent on the amount of data that initial results relate to a user's latest posts, and processing of full historic will require the user to re-login at a later date.

The system shall provide an interface such that newly created content or feedback (e.g. likes, ratings, etc.) can be vetted prior to upload. This interface should provide options to block the post, or offer alternatives to which the implications to the profile can be determined (see next point).

The system may be able to make best effort associations between data placed onto OSN(s) and the profile attributes, which can be inferred from such data.

These points are related to user requirements, which will need to be processed by the browser plugin, as such actions are unable to be intercepted directly within an OSN. This means such features will only have the ability to provide limited feedback.

The system may be able to provide suggestions for alterations regarding the visibility of parts of the posted content in order to allow the user to make informed changes on how the profile will be outwardly perceived. This is related to how textual and imagery content is scored from the elements defined within the latter functions section. For historical data this will be scored after the fact and analysed after posting, requiring the user to first login for data collection, and later login to see the results, which may affect the work flow defined in the prior section.

4.2. Use Case 2: Information Value Awareness Tool

The information value awareness tool will give an indication on the value of content – no direct monetary value will be assigned to data, as no direct correlation can be made as such value will vary between OSN, and an exacting value is hard to determine. Moreover it will focus on brand determination and the type of organisations, which may have an interest in a user's data.

The system shall provide a way/methodology to estimate the value of various end-users personal data/information (activities) shared and accessed via OSNs, this is a key element of this tool, and will make use of both a user's textual and imagery data.

The system shall enable the visualisation of end-users digital trails and thus, the estimations of profiles (and/or profile segments/categories) she/he is placed into by different actors in the network. Thus, the system should provide information from profiling in order to show the user which entities have the greatest interest in their data. This will be provided through the visualisation tools, and can make use of profiling data provided by the browser plugin, which will track information on which sites a user visits, and moreover the trackers those sites contain, which can be traced back to the user (such as Google ad-words, or delegated Facebook components).

The system shall enable to provide end-users useful insights on the value of their digital data and social footprint that are either directly shared in social networks (e.g., likes on Facebook) or are indirectly collected by various network actors that track their activities on web browsers. This is a key part of the user requirements on giving direct feedback of information they have posted to an OSN.

The system shall avoid situations, which would commoditise user's data. The focus is to alert the user to such situations, rather than make use of such situations itself. This is a key part of this project, and on resale of data and privacy concerns are to be addressed in the legal framework, and the user is be notified that data will only be used to inform themselves.

The system shall facilitate end users to efficiently manage the value of their personal data. That is to say the system will provide a means to alter posts, or identify posts such that a user has control over the impression their profile gives in terms of value.

The system may be able to get fruitful insights on how relevant a user's profile is for different stakeholders, this is in relation to user requirements with regard to brand ambassadors, and determining a user actions from their data.

The system may be able to enable end-users to manage their audience either explicitly, via allowing the monitoring and collection of their actions/profile by specific stakeholders, or implicitly, via their postings (and social actions in general). Such an action is dependent upon actions allowed by an OSN, and is therefore something, which would have to be provided through a mock-up (as per the prior section)

From user interactions, the system shall provide easily comprehensible cues on how the monetary value will change based on future actions prior to posting. Limited functionality is likely here as such an element will have to pass through the browser plugin, which can provide limited capturing of posts prior to arriving at an OSN.

The system may provide historic value data so the change in value can be tracked over time. This will constitute part of the DataBait GUI tool.

5.System Perspective

This section will lay out requirements from the interoperability of elements within the system and the interaction of the user with the system. This will give a brief overview as such elements will be defined through the technical architecture which will define the interfaces of the system in detail.

5.1. System Interfaces

There shall be two primary system interfaces. One interface will provide direct access to OSN data via publicly known APIs made available by the associated OSNs. The second interface will be via a browser plugin that can monitor user activity in the browser and provide for the ability to process data prior to posting. An additional interface may be required in order to enable the browser plugin interface to also feed and retrieve data from the OSN.

5.2. User Interfaces

There shall be two main user interfaces facilitating two modes of operation towards proficiently enabling end users to interact with and exploit USEMP features and innovation.

The first mode of operation will be via a web app (with an elastic HTML5 enable design for supporting both desktop and mobile browsers) used towards facilitating USEMP platform end-users:

a) to visualise their OSN digital trails (e.g., actions, personal data creation/access/tracking) and profile, via advanced visualizations aiming at providing a unique end-user experience as well as intuitive insights.
b) to easily navigate among USEMP platform enabled features such as: ‘Profile’, ‘Personal Data control’, ‘Future Control’, ‘Settings’, “Personal Data Value” etc.
c) to login via USEMP and OSNs towards providing USEMP platform permissions and accessing rights for getting OSNs historical and real-time data.

The second mode of operation will be via a browser plugin which will track and collect the end-user’s behavior and data (on his browser or OSN) and send them to USEMP’s back end for processing and analysis, towards informing the user of any implications as per section 4.1.

One of the key enablers of envisioned USEMP User interfaces will be 5ml, a tool that allows anyone with web design principles to create rich media applications for desktop web, Facebook, mobile web and mobile native applications (iOS, Android, Windows Phone). Thus, 5ml can be used to fast prototype applications and designs for pilot experiments and the UX experience for end-consumer.

5.3. Software Interfaces

Software interfaces are defined within Section 6. This section will provide the required input parameters for each software element and the associated output data returned.

5.4. OSN Constraints

OSNs limit access to data they contain through restrictions and limitations at the API level. There are two types of limits imposed by an OSN on the amount of API calls an application can make over time, either:

- Application Level (i.e. all users of an app)
- User Level (i.e. number of API calls – user's personal data – an application can access for a specific user

The above are very strict and vary from OSN to OSN. In reference to Facebook see 'https://developers.facebook.com/docs/reference/ads-api/api-rate-limiting/'.

The system must set priorities and limitations on the type/amount of a user's historical OSN data that is accessed per OSN.

The system must set limitations and throttling policies on the way we access the latter data.

Dependent on the amount of data required and given API rate limits, the process of collecting historical data may be time expensive.

Other OSNs such as twitter place no restrictions on the amount of data that can be pulled from a users profile, but does have rate limiting when pulling large amount of public data. This will be reviewed on a case-by-case basis as OSNs are added into the system.

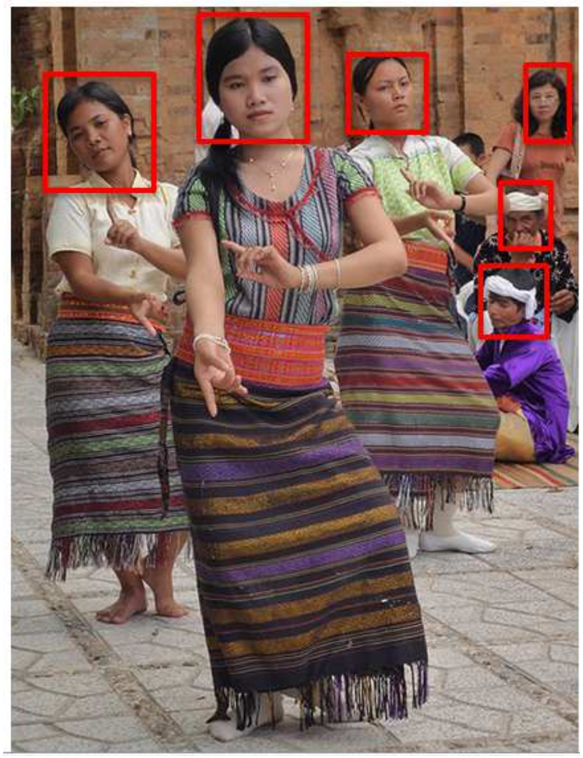
6.Functions

This section defines the technical integration requirements for the associated components. Specifically the input and operational parameters, and the resultant output. These elements will be used for the analytics of the OSN data used to determine social factors and extract additional data from a users profile. These elements were first described in order to determine user interactions with the system, and determine what actions are feasible within the system. These components will constitute the back end analytics, and therefore their capabilities are here noted.

6.1. Face Detection

The face detection tool spots faces within natural images, providing their localisation. This last is expressed as the coordinates of the upper-left corner and the width and height of the face's area. Good performances in the ideal case (frontal face, good resolution (50 pixels), no occlusion). Robustness should be improved for harder cases. The tool can be used as such or as an input for other modules such as face recognition. It is useful in order to detect images with faces in the multimedia stream of a user.

Example: Suppose that a user uploads the image from the following figure. The module outputs the list of faces in the image, their positions and their sizes.

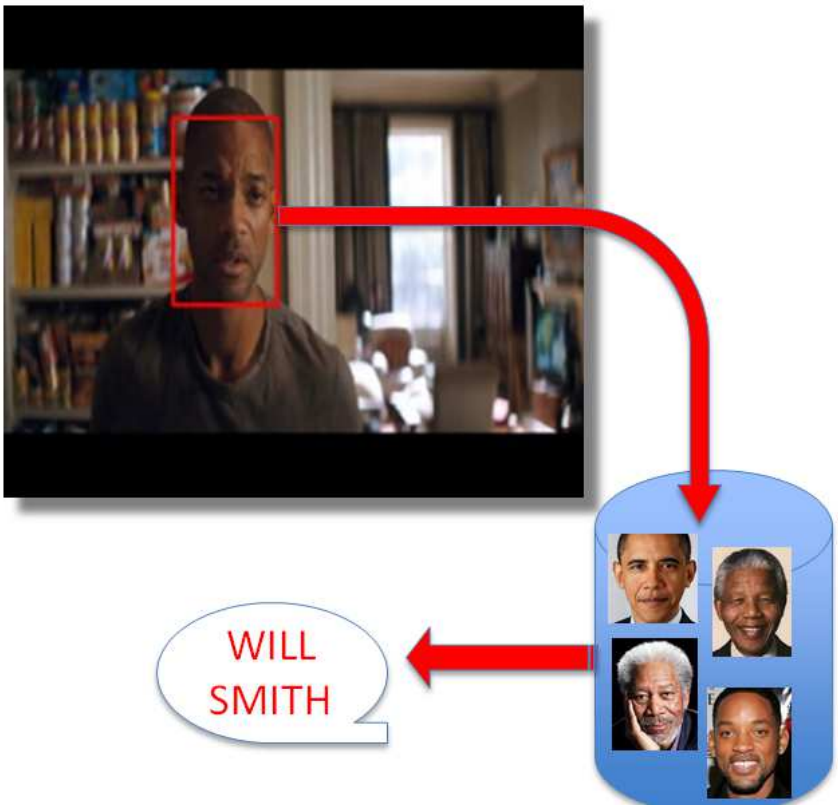


Inputs	Outputs
Image	The list of faces depicted in the image, along with their positions and sizes.

6.2. Face Recognition

The face recognition analyses faces within natural images, providing their identity. This is expressed as a textual string of the person's name. It performs well in the ideal case (frontal face, good resolution (50 pixels), no occlusion). The robustness should be improved for harder cases. It usually works from a pre-selected rectangular area around the face (see face detection). Given a closed list of possible identities, the closest one among them is chosen by the tool. It is important in order to detect which are the prominent persons in a user's social graph and can be used, for instance, in order to change the visibility or to remove photos of the user with one of his/her contacts.

Example: Suppose that the image from the following figure is uploaded by a user. The face in the image will be compared to a list of recognizable faces and the system will estimate that Will Smith is depicted in the image.

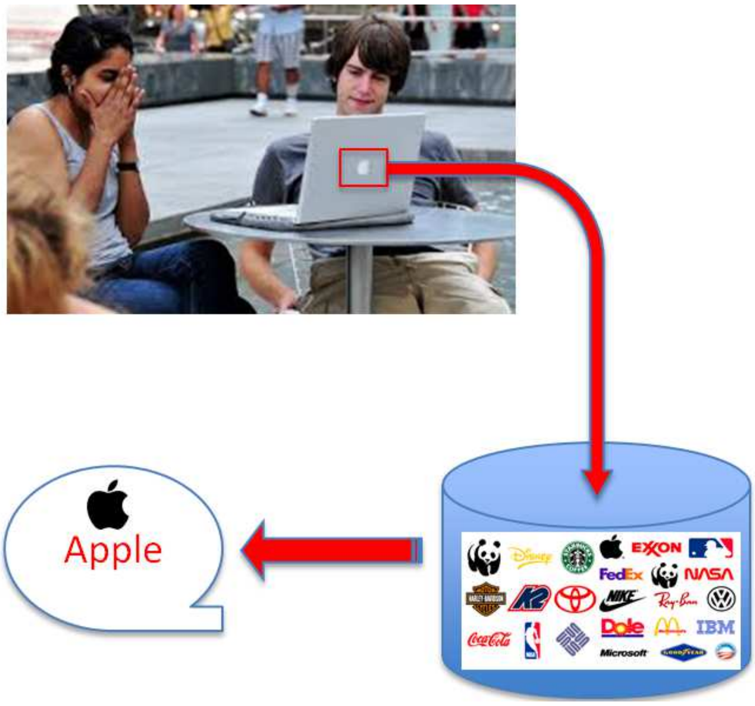


Inputs	Outputs
Image	The name(s) of the person(s) depicted in the image.

6.3. Logo Recognition

The logo recognition tool detects and recognizes logos within natural images. It identifies which logos are seen in a test image as well as its localization. This last is expressed as the coordinates of the upper-left corner and the width and height of the logo's area. Logos that can be recognised must be previously registered in the database of the technology and processed. It is useful mostly for associating a person to product/brands and for building a consumer profile through the aggregation of information shared over time.

Example: Suppose that the image from the following figure is uploaded by a user. By confronting it to an existing dataset of logos/brands, the tool will detect that the depicted person is an Apple user.



Inputs	Outputs
Image	The name(s) of the logo(s) depicted in the image.

6.4. Multi-Modal Similarity

This technology provides a level of similarity between two multimedia contents (text+image). By definition, such content may be composed of both textual and visual information. This information is processed separately then merged to result into a unique score of similarity. Multimedia similarity allows estimating the degree of similarity between two multimedia documents, even when they are composed of both textual and visual content. Domains of interest of the user can be obtained through the aggregation of individual estimations.

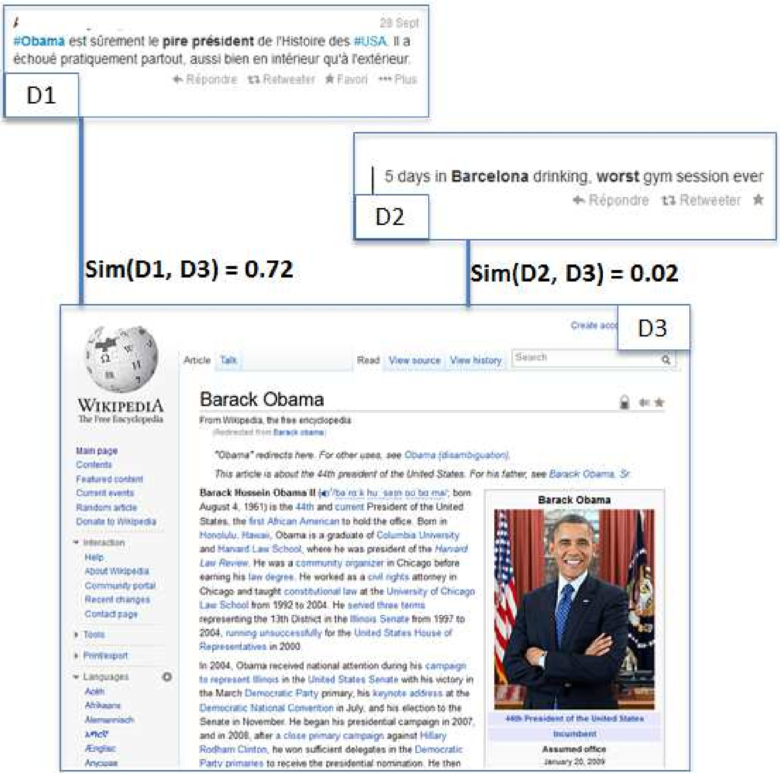
Example: Suppose that two blog articles (D1 and D2) describing cake recipes need to be compared. One of them includes mostly text and the other one mostly images. The tool represents D1 and D2 in a common conceptual space and will be able to detect that they convey information about the same topic (i.e. food).

Inputs	Outputs
Text (String)	Similarity between two individual documents.
Text (String) + Image	Domains of interest of the user.

6.5. Text Similarity

The text similarity leverages Wikipedia knowledge in order to improve classical text representations which processes only the words that appear explicitly in the texts. It copes with two important problems: (a) variable document length – single words, short OSN updates, news articles etc. – can be processed transparently; (b) multilingualism – documents in different languages – can be compared seamlessly. This technique automatically represents text documents via Wikipedia concepts and then exploits these representations to compute their similarity. It is useful in order to determine which are the domains of interest (i.e. sports, politics, religion etc.) of an OSN user. These domains of interest are obtained by aggregating individual contributions over time.

Example: Suppose that documents D1 (tweet in French), D2 (tweet in English), D3 (biography in English) are shared by the user. Their analysis will allow attaching D1 to D3 and as such will be attached to the "politics" domain.



28 Sept
#Obama est sûrement le **pire président** de l'Histoire des #USA. Il a échoué pratiquement partout, aussi bien en intérieur qu'à l'extérieur.

5 days in **Barcelona** drinking, **worst** gym session ever

Sim(D1, D3) = 0.72

Sim(D2, D3) = 0.02

Barack Obama

From Wikipedia, the free encyclopedia
(Redirected from Barack obama)

"Obama" redirects here. For other uses, see Obama (disambiguation).

This article is about the 44th president of the United States. For his father, see Barack Obama, Sr.

Barack Hussein Obama II (/bəˈrɑːk huːˈseɪn oʊˈbɑːmə/; born August 4, 1961) is the 44th and current President of the United States, the first African American to hold the office. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he was president of the *Harvard Law Review*. He was a community organizer in Chicago before earning his law degree. He worked as a civil rights attorney in Chicago and taught constitutional law at the University of Chicago Law School from 1992 to 2004. He served three terms representing the 13th District in the Illinois Senate from 1997 to 2004, running unsuccessfully for the United States House of Representatives in 2000.

In 2004, Obama received national attention during his campaign to represent Illinois in the United States Senate with his victory in the March Democratic Party primary, his keynote address at the Democratic National Convention in July, and his election to the Senate in November. He began his presidential campaign in 2007, and in 2008, after a close primary campaign against Hillary Rodham Clinton, he won sufficient delegates in the Democratic Party primaries to receive the presidential nomination. He then

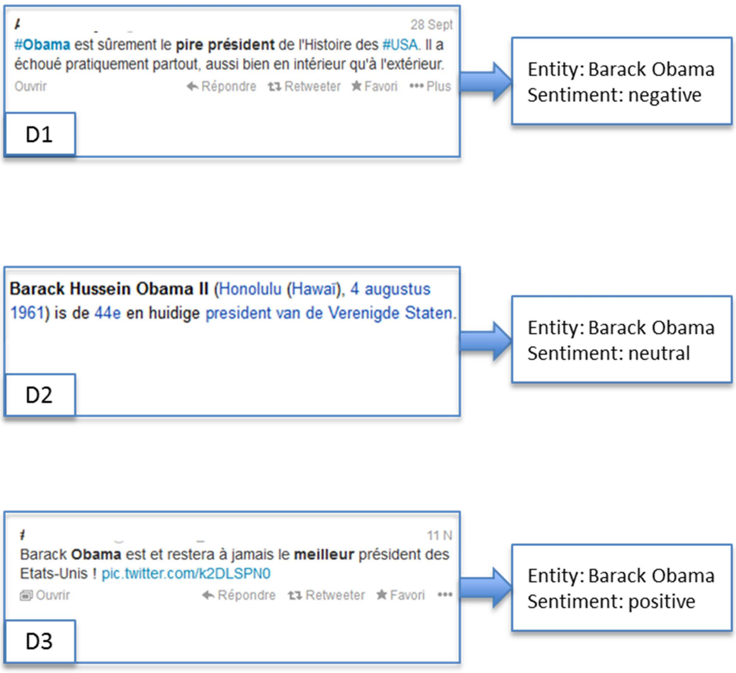
Barack Obama
44th President of the United States
Incumbent
Assumed office
January 20, 2009

Inputs	**Outputs**
Text (String)	Domain(s) the text refers to. Domains of interest of the user.

6.6. Opinion Mining

The opinion mining tool exploits lists of opinionated terms and/or structural features (such as punctuation, smileys etc.), combined with machine learning in order to attribute a sentiment to a given text. In a simple version, this sentiment can be one of positive/neutral/negative. It is used after entity recognition in order to estimate the sentiment that is expressed about a target entity. It is well studied for English and less mature for other languages. Opinion mining is useful to mine her/his opinions about potentially interesting topics of interest (politics, religion, sexual orientation etc.).

Example: Suppose that document D1, D2 and D3 are provided to the system. The outputs will be an entity and an associated sentiment, which is predicted from the user contributed text.



28 Sept

#Obama est sûrement le **pire président** de l'Histoire des #USA. Il a échoué pratiquement partout, aussi bien en intérieur qu'à l'extérieur.

Entity: Barack Obama
Sentiment: negative

Barack Hussein Obama II (Honolulu (Hawaï), 4 augustus 1961) is de 44e en huidige president van de Verenigde Staten.

Entity: Barack Obama
Sentiment: neutral

11 N

Barack **Obama** est et restera à jamais le **meilleur** président des Etats-Unis ! pic.twitter.com/k2DLSPN0

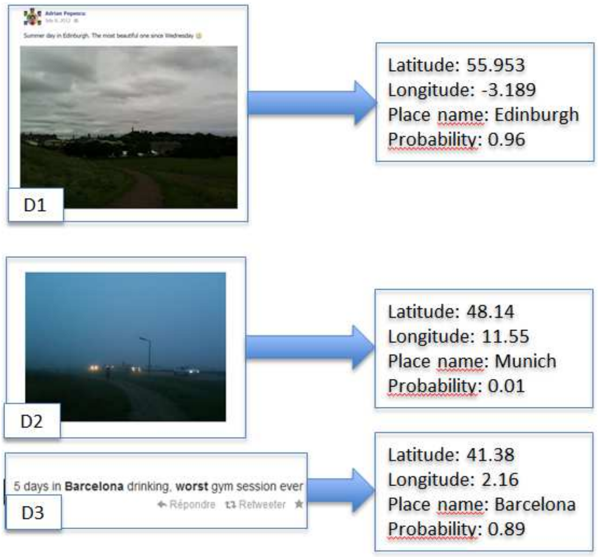
Entity: Barack Obama
Sentiment: positive

Inputs	Outputs
Text (String) Text + image	Positive/neutral/negative characterization of the entity/entities in the text.

6.7. Content Location

The content location tool exploits statistical models of places that are extracted either from text or image content in order to predict the most probable location a document refers to. The technique also associates a probability to the correctness of the prediction which makes it possible to filter out documents whose predicted location is unsure. The aggregation of predictions for individual documents of a user allows the extraction of a detailed location profile.

Example: Suppose that the following three documents (D1 - text+image; D2 - image alone and D3 - text alone) are provided to the system. The outputs will be coordinate pairs, corresponding place names and a probability for the location prediction to be accurate.


Adrian Popescu
July 6, 2012

Summer day in Edinburgh. The most beautiful one since Wednesday

5 days in **Barcelona** drinking, **worst** gym session ever

Latitude: 55.953
Longitude: -3.189
Place name: Edinburgh
Probability: 0.96

Latitude: 48.14
Longitude: 11.55
Place name: Munich
Probability: 0.01

Latitude: 41.38
Longitude: 2.16
Place name: Barcelona
Probability: 0.89

Inputs	Outputs
Image (URL pointing to image or Image object) Text (String) Text + image	List of locations which are associated to the user.

6.8. Personal Attribute Multimedia Predictor

The personal attribute multimedia predictor function takes an image as input and optionally a text-based description of the image (e.g. a title or a set of tags) and produces a set of prediction scores about personal attributes depicted in the image. For instance, such attributes could be smoking, drinking, partying, doing extreme sports, etc. These should be defined ahead of time so that appropriate classifiers are trained and tested before the actual deployment. The prediction scores could be provided in the form of float/double values (expressing the degree to which the concept of interest was detected) or boolean decisions or both.

Example: Suppose that we have selected the concepts smoking, drinking, extreme sports as concepts of interest and the function is provided with the following image object as input.

One possible output of the function could be:

[(S, 0.77, TRUE), (D, 0.32, FALSE), (XS, 0.01, FALSE)]

with S: Smoking, D: Drinking, XS: Extreme Sports

Inputs	Outputs
Image (URL pointing to image or Image object) Text (String) [accompanying image]	Attribute Prediction List (ArrayList) [for a pre-specified number of attributes]

6.9. Personal Attribute Behavioral Predictor

The personal attribute behavioral predictor function takes a set of user activities as input and produces a set of prediction scores about personal attributes that could be inferred by these activities. These attributes do not necessarily have to be the same as the ones of 4.8. In fact, they could be completely or largely non-overlapping, since there are attributes that are better inferable by visual content and others that can be inferred by behavioural features. User activities could refer to online activities associated with online resources, for instance likes of particular Facebook pages, visits of particular sites, retweets of particular articles, etc. The kind of online activities we want to use and the set of personal attributes that we want to predict need to be defined ahead of deployment so that appropriate models can be built and tested. An additional issue regarding this function has to do with the "completeness" of the user activity list that is provided as input. For instance, would this function operate in the same way independent of whether the input activities refer to the full user history or only a very small subset (in the extreme case only a single item)?

Example: Suppose that the concepts of interest are homosexual, diabetic, and liberal and that for a given user we have the following list of likes as input:

https://www.facebook.com/pages/Diabetes-Support/119703998103102

https://www.facebook.com/DiabeticLifestyle

https://www.facebook.com/michelleobama?rf=106042606102911

https://www.facebook.com/johnkerry

https://www.facebook.com/justsaynow

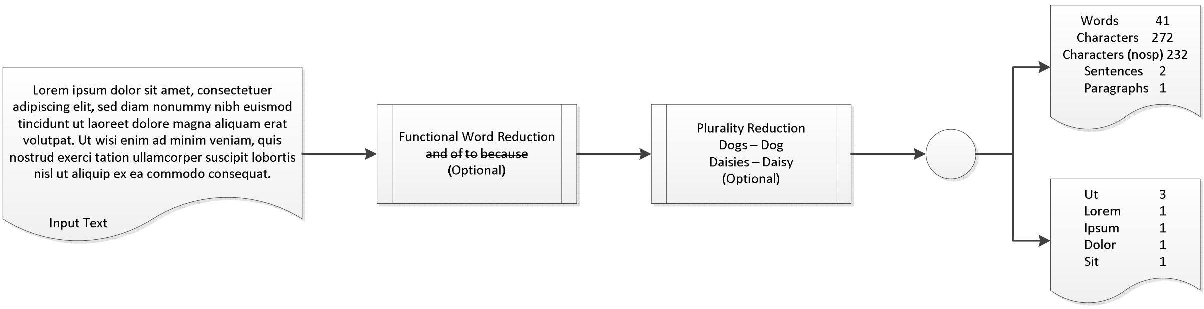
Then, the function output could be: [(H, 0.04, FALSE), (D, 0.82, TRUE), (L, 0.72, TRUE)]

with H: Homosexual, D: Diabetic and L: Liberal.

Inputs	Outputs
User Activity List (ArrayList) [elements of User Activity could be Likes, Visits to websites, etc. and in the simplest case could be represented as URLs]	Attribute Prediction List (ArrayList)

6.10. Word Count

The word count function takes raw text input either from a command line file uri, or via passing text directly as a formatted string. Additional parameters are an optional word blacklist, or enabling default functional word reduction, as well as a boolean parameter to reduce word plurality to singular form and count all such words as identical. The function output is an arraylist of words with the associated word frequency. Statistics about the input text makeup can be queried from the object but are not returned by default.


Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.
Input Text
Functional Word Reduction
~~and of to because~~
(Optional)
Plurality Reduction
Dogs – Dog
Daisies – Daisy
(Optional)
Words 41
Characters 272
Characters (nosp) 232
Sentences 2
Paragraphs 1
Ut 3
Lorem 1
Ipsum 1
Dolor 1
Sit 1

Inputs	Outputs
Textual Data File / Raw Text (String)	Word Frequency List (ArrayList)
Blacklist (Comma Separated String)	Text Statistics (Various)
Functional Word Reduction (Boolean)	
Plurality Reduction (Boolean)	

6.11. Tracking and Analytics Function

Tracking and analytics function (based on open source piwik) allows to the use of 1st party and 3rd party cookies and privacy control to track end users behaviour across mobile web/web/native applications, via a light weighted mobile client SDK. The function allows tracking of behavioural information on how end-users interact with web applications, Facebook applications and mobile applications and send the information to the USEMP platform back-end. This allows for further processing and analysis of OSN historical data collected by the USEMP OSN-enabled app. The communication interface will use HTTP APIs and jscript library for posting tracking information to the service (e.g., USEMP back end).

Inputs	Outputs
End-Users interactions on a specific web, mobile web or native application.	Cross-channels tracking information towards end-user digital trail profiling.