International Journal on Advances in Software, vol 12 no 1 & 2, year 2019, http://www.iariajournals.org/software/

Trend Discovery and Social Recommendation in Support of Documentary Production

Giorgos Mitsis, Nikos Kalatzis, Ioanna Roussaki, Symeon Papavassiliou
Institute of Communications and Computer Systems
Athens, Greece
e-mails: {gmitsis@netmode, nikosk@cn, ioanna.roussaki@cn, papavass@mail}.ntua.gr

***Abstract*—Recent market research has revealed a globally growing interest on documentaries that have now become one of the biggest content-wise genre in the movie titles catalog, surpassing traditionally popular genres such as comedy or adventure films. At the same time, modern audiences appear willing to immerse into more interactive and personalized viewing experiences. Documentaries, even in their linear version, involve high costs in all phases (pre-production, production, post-production) due to various inefficiencies, partly attributed to the lack of scientifically-proven cost-effective Information and Communications Technology tools. To fill this gap, a set of innovative tools is delivered that focus on supporting all stages of the documentary creation process, ranging from the documentary topic selection to its final delivery to the viewers. This paper elaborates on two specific tools that primarily focus on the interests and satisfaction of the targeted audience: the Integrated Trends Discovery tool and the Social Recommendation & Personalization tool. It presents their design, functionality and performance, discusses the extended evaluation and validation that has been carried out and concludes with exploring the future plans and potential regarding these tools.**

Keywords-documentary production; social-media analytics; Integrated Trends Discovery tool; Social Recommendation & Personalization tool; evaluation; validation; benchmarking.

I. INTRODUCTION

From the earliest days of cinema, documentaries have provided a powerful way of engaging audiences with the world. They always had social and market impact, as they adapted to the available means of production and distribution. More than any other type of films, documentarians were avid adapters of new technologies, which periodically revitalized the classical documentary form. The documentary is a genre that lends itself straightforwardly to interaction. People have different knowledge backgrounds, different interests and points of view, different aesthetic tastes and different constraints while viewing a programme. Therefore, it becomes evident that some form of personalized interactive documentary creation will enhance the quality of experience for the viewers, facilitating them to choose different paths primarily with respect to the documentary format and playout system. The convergence between the documentary production field and of digital media enables the realization of this vision.

As the range of Information and Communications Technology (ICT) platforms broadens, documentary makers need to understand and adopt emerging technologies in order to ensure audience engagement and creative satisfaction, via the use of personalization and interactive media. One of the major challenges for stakeholders in the arena of documentary creation is the development of processes and business models to exploit the advantages of those technical achievements, in order to reduce the overall cost of documentary end-to-end production, to save time and to deliver enhanced personalized interactive and thus more attractive documentaries to the viewers.

This paper is based on [1] that has been prepared within PRODUCER [2], an H2020 EU project that aims to pave the path towards supporting the transformation of the well-established and successful traditional models of linear documentaries to interactive documentaries, by responding to the recent challenges of the convergence of interactive media and documentaries. This is achieved via the creation of a set of enhanced ICT tools that focus on supporting all documentary creation phases, ranging from the user engagement and audience building, to the final documentary delivery. In addition to directly reducing the overall production cost and time, PRODUCER aims to enhance viewers' experience and satisfaction by generating multi-layered documentaries and delivering more personalized services, e.g., regarding the documentary format and playout.

In order to provide the aforementioned functionality, the PRODUCER platform implemented 9 tools, each focusing on a specific documentary production phase. These tools are: Integrated trends discovery tool, Audience building tool and Open content discovery tool (that support the documentary pre-production phase), Multimedia content storage, search & retrieval tool and Automatic annotation tool (that support the core production phase), Interactive-enriched video creation tool, 360° video playout tool, Second screen interaction tool and Social recommendation & personalization tool (all four focusing on the documentary post-production phase). The architecture of the PRODUCER platform is presented in more detail in [3].

As already mentioned, this paper is based on [1], where an initial prototype implementation was described for two of the PRODUCER tools: the Integrated Trends Discovery tool and the Social Recommendation & Personalization tool. In the current paper, the final version of the prototypes is presented along with a thorough evaluation.

In the rest of the paper, Section II elaborates on the design & functionality of the Integrated Trends Discovery tool while Section III focuses on the description of the Social Recommendation & Personalization tool. In Section IV, the results of the evaluation of the tools are presented. Finally, in

Section V, conclusions are drawn and future plans are presented.

II. Integrated Trends Discovery Tool

This section elaborates on the ITD tool, i.e., its innovations, architecture, user demographics inference mechanism and respective evaluation.

A. Rationale and Innovations

In recent years, there is an increasing trend on utilizing social media analytics and Internet search engines analytics for studying and predicting behavior of people with regards to various societal activities. The proper analysis of Web 2.0 services utilization goes beyond the standard surveys or focus groups. It is a valuable source of information leveraging internet users as the largest panel of users in the world. Analysts from a wide area of research fields have the ability to reveal current and historic interests of individuals and to extract additional information about their demographics, behavior, preferences, etc. One of the valuable aspects of this approach is that the trial user base consists of people that have not participated in the user requirement extraction phase.

Some of the research fields that demonstrate significant results through the utilization of such analytics include epidemiology (e.g., detect influenza [4][5] and malaria [6]) epidemics), economy (e.g., stock market analysis [7], private consumption prediction [8], financial market analysis and prediction [9], unemployment rate estimation [10]) politics (e.g., predicting elections outcomes [11]).

On the other hand, there are limitations on relying only on these information sources as certain groups of users might be over- or under-represented among internet search data. There is a significant variability of online access and internet search usage across different demographic, socioeconomic, and geographic subpopulations.

With regards to content creation and marketing, the existing methodologies are under a major and rapid transformation given the proliferation of Social Media and search engines. The utilization of such services generates voluminous data that allows the extraction of new insights with regards to the audiences' behavioral dynamics. In [12], authors propose a mechanism for predicting the popularity of online content by analyzing activity of self-organized groups of users in social networks. Authors in [13] attempt to predict IMDB (http://www.imdb.com/) movie ratings using Google search frequencies for movie related information. In a similar manner, authors in [14] are inferring, based on social media analytics, the potential box office revenues with regards to Internet content generated about Bollywood movies.

The existing research approaches mainly focus on the post-production phase of released content. Identifying the topics that are most likely to engage the audience is critical for content creation in the pre-production phase. The ultimate goal of content production houses is to deliver content that matches exactly what people are looking for. Deciding wisely on the main documentary topic, as well as the additional elements that will be elaborated upon, prior to engaging any resources in the documentary production process, has the potential to reduce the overall cost and duration of the production lifecycle, as well as to increase the population of the audiences interested, thus boosting the respective revenues. In addition, the existence of hard evidence with regards to potential audience's volume and characteristics (e.g., geographical regions, gender, age) is an important parameter in order to decide the amount of effort and budget to be invested during production.

There are various social media analytics tools that are focusing on generic marketing analysis, e.g., monitoring for a long time specific keyword(s) and websites for promoting a specific brand and engaging potential customers. These web marketing tools rely on user tracking, consideration of user journeys, detection of conversion blockers, user segmentation, etc. This kind of analysis requires access to specific websites analytics and connections with social media accounts (e.g., friends, followers) that is not the case when the aim is to extract the generic population trends. In addition, these services are available under subscription fee that typically ranges from 100 Euros/month to several thousand Euros/month, a cost that might be difficult to be handled by small documentary houses.

The ITD Tool aims to support the formulation, validation and (re)orientation of documentary production ideas and estimate how appealing these ideas will be to potential audiences based on data coming from global communication media with massive user numbers. The ITD tool integrates existing popular publicly available services for: monitoring search trends (e.g., Google Trends), researching keywords (e.g., Google AdWords Keyword Planner), analyzing social media trends (e.g., Twitter trending hashtags). In more details, the ITD tool innovations include the following:

- Identification and evaluation of audience's generic interest for specific topics and analysis/inference of audience's characteristics (e.g., demographics, location)
- Extraction of additional aspects of a topic through keyword analysis, quantitative correlation of keywords, and association with high level knowledge (e.g., audience sentiment analysis)
- Discovery and identification of specific real life events related to the investigated topic (e.g., various breakthroughs of google/twitter trending terms are associated with specific incidents)
- Utilization of data sources that are mainly openly accessible through public APIs, which minimizes the cost and increases the user base.

B. Architecture & Implementation Specifications

A functional view of ITD tool's architecture is provided in Fig. 1. Its core modules are described hereafter.

RestAPI: This component exposes the backend's functionality through a REST endpoint. The API specifies a set of trend discovery queries where the service consumer provides as input various criteria such as keywords, topics, geographical regions, time periods, etc.

Trends Query Management: This component orchestrates the overall execution of the queries and the processing of the replies. It produces several well formulated queries that are forwarded to the respective connectors/wrappers to dispatch the requests to several existing TD tools/services available online. Given that each external service will reply in different time frames (e.g., a call to Google Trends discovery replies within a few seconds while Twitter stream analysis might take longer) the overall process is performed in an asynchronous manner, coordinated by the Message Broker. The Query Management enforces querying policies tailored to each service in order to optimize the utilization of the services and avoid potential bans. To this end, results from calls are also stored in ITD tool's local database in order to avoid unnecessary calls to the external APIs.

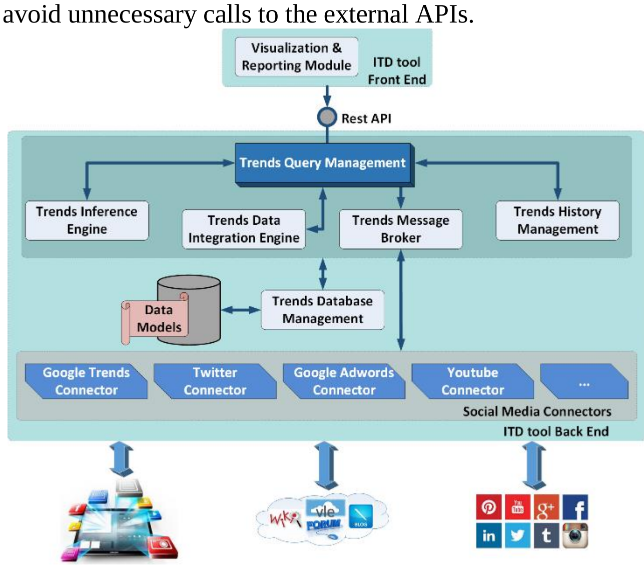


Figure 1. Architecture of the Integrated Trends Discovery Tool.

Trends Message Broker: This component realizes the asynchronous handling of requests. It is essentially a messaging server that forwards requests to the appropriate recipients via a job queue based on a distributed message passing system.

Social Media Connectors: A set of software modules that support the connection and the execution of queries to external services through the provided APIs. Connectors are embedding all the necessary security related credentials to the calls and automate the initiation of a session with the external services. Thus, the connectors automate and ease the actual formulation and execution of the queries issued by the Query Management component. Some example APIs that are utilized by the connectors are: Google AdWords API, Twitter API, YouTube Data API v3.

Trends Data Integration Engine: This module collects the intermediate and final results from all modules, homogenize their different formats, and extracts the final report with regards to the trends discovery process. The results are also modelled and stored in the local data base in order to be available for future utilization.

Trends Database Management & Data model: The ITD tool maintains a local database where the results of various calls to external services are stored. The Database Management module supports the creation, retrieval, update and deletion of data objects. This functionality is supported for both contemporary data but also for historic results (Trends History Management). Hence, it is feasible for the user to compare trend discovery reports performed in the past with more recent ones and have an intuitive view of the evolution of trend reports in time.

Trends Inference Engine: In some cases, the external services are not directly providing all information aspects of the required discovery process and the combination and analysis of heterogeneous inputs is required. To this end, the application of appropriate inference mechanisms on the available data allows the extraction of additional information escorted by a confidence level that expresses the accuracy of the estimation. Details on the rationale and mechanisms of this module are presented in the following section.

The technologies used for the implementation of the ITD tool can be found in Table I.

TABLE I. ITD SOFTWARE SPECIFICATIONS

Licensing	Open source
Core Implementation Technologies	Python 2.7
Additional technologies utilised	Nginx server Django 1.10 (Python framework) djangorestframework 3.5.1 Celery RabbitMQ Redis
Database details	MySQL 5.x
Exposed APIs	REST
Exchanged data format	JSON
GUI description	HTML5, Javascript, CSS3, Angular JS 1.6, Angular-material 1.1.3

The tool is developed as an open source project and the source code can be found at [15].

C. Knowledge Extraction Approaches

As discussed in the previous section, during the preproduction phase of a documentary, producers are highly interested in estimating audiences' interests in correlation with high level information like the gender, the age and the sentiment of potential audiences. In a similar manner, after a show has been aired, useful results can be inferred through the analysis of the Internet buzz that the show has created. In other cases, merging information from different, previously unrelated, sources may provide a higher confidence on the final outcome. To this end, various data processing and inference mechanisms are deemed necessary. The ITD tool follows a modular approach with regards to this aspect. The ITD tool provides the necessary means for collecting all relevant data at one place and then different data analytics algorithms can be applied allowing the extraction of additional knowledge according to the scope of the user. As a proof of concept and for supporting the needs of the production teams within the scope of the PRODUCER project, inference algorithms were developed for: i) extracting audience's characteristics through Twitter data and ii) analyze popularity of targeted TV shows by be complementary use of Google Trends service with Twitter. The design principles and the actual evaluation results of both approaches are presented in Section IV.

D. Graphical User Interfaces

The Front-End allows the user to create a new query and visualizes the respective results. The overall process consists of two steps supported by two pages (Fig. 2).

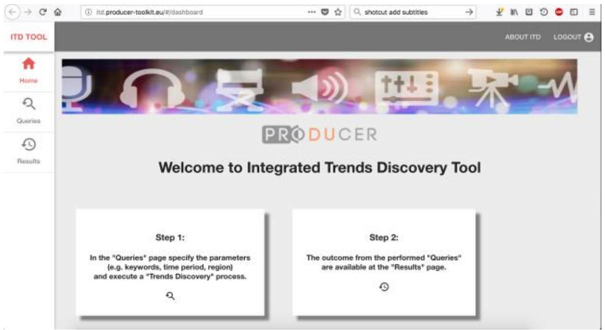


Figure 2. The "Home" page of the ITD tool.

First the query's parameters within the "Queries" (Fig. 3) page are specified and based on these parameters a discovery process is initiated.

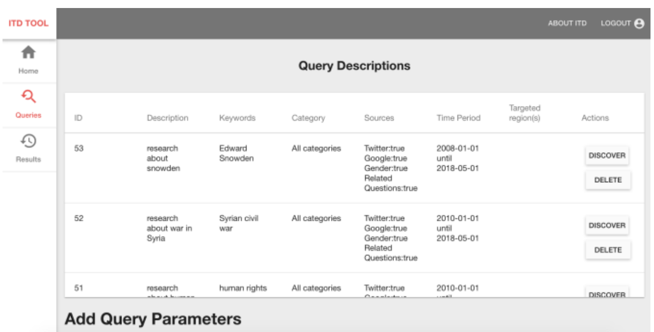


Figure 3. The "Queries" page of the ITD tool.

After a successful completion of the query the results are presented on the "Results" page (Fig. 4 and Fig. 5), which provides the following output: (i) a graph of terms (each term is escorted by a user's popularity metric and is correlated with other terms, where a metric defines the correlation level), (ii) interest per location (country/city), (iii) interest per date(s) allowing the identification if significant dates and seasonal habits, (iv) sentiment and gender analysis related with the researched topic and (vi) questions related to the topic.

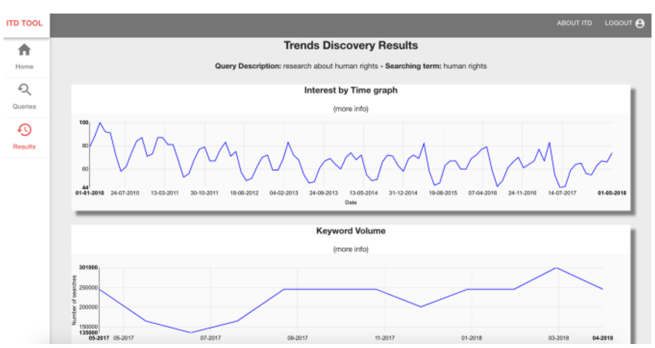


Figure 4. A snapshot of "Results" page focusing on "Interest by Time" and "Keyword Volume".

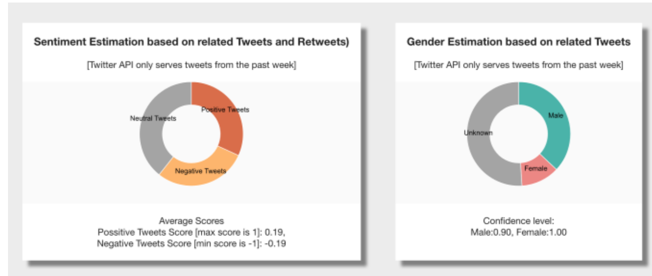


Figure 5. A snapshot of "Results" page focusing on "Sentiment and Gender Estimation".

Finally, the front-end allows the reviewing of results from past queries and the conversion and download of the query results in CSV format.

III. Social Recommendation & Personalization Tool

This section elaborates on the SRP tool, i.e., its functionality, architecture, recommendation extraction algorithm.

A. Rationale & Goal

Personalization & Social Recommendation are dominant mechanisms in today's social networks, online retails and multimedia content applications due to the increase in profit of the platforms as well as the improvement of the Quality of Experience (QoE) for its users and almost every online company has invested in creating personalized recommendation systems. Major examples include YouTube that recommends relevant videos and advertisements, Amazon that recommends products, Facebook that recommends advertisements and stories, Google Scholar that recommends scientific papers, while other online services provide APIs such as Facebook Open Graph API and Google's Social Graph API for companies to consume and provide their own recommendations [16].

The Social Recommendation & Personalization (SRP) tool of PRODUCER holistically addresses personalization, relevance feedback and recommendation, offering enriched multimedia content tailored to users' preferences. The tool's functionalities can be used in any type of content that can be represented in a meaningful way, as explained later. The application is thus not restricted to documentaries.

The recommendation system we built is not restricted to the video itself, but applies to the set of enrichments accompanying the video as well. Interaction with both video and enrichments is taken into consideration into updating the user's profile, thus holistically quantifying the user's behaviour. Its goal is to facilitate the creation of the documentary and allow the reach of the documentary to a wider audience. To do so, the SRP tool is responsible for proposing content to the user or to the producer of the film relevant to specific target groups, via a personalization mechanism.

B. Architecture & Implementation Specifications

SRP tool's architecture is presented in Fig. 6 and it consists of the following components:

RestAPI: This component is responsible for the exchange of information between the frontend of the SRP tool or any application willing to use the SRP tool's functionality, and its backend.

Frontend: This component is responsible for the Graphical User Interface via which the user interacts with the tool. More information on this component will be presented in subsection D.

User Interaction Monitoring: As the user interacts with the content and the frontend of the tool, interactions and data are being sent to the backend in order to be processed by the tool and perform the corresponding actions.

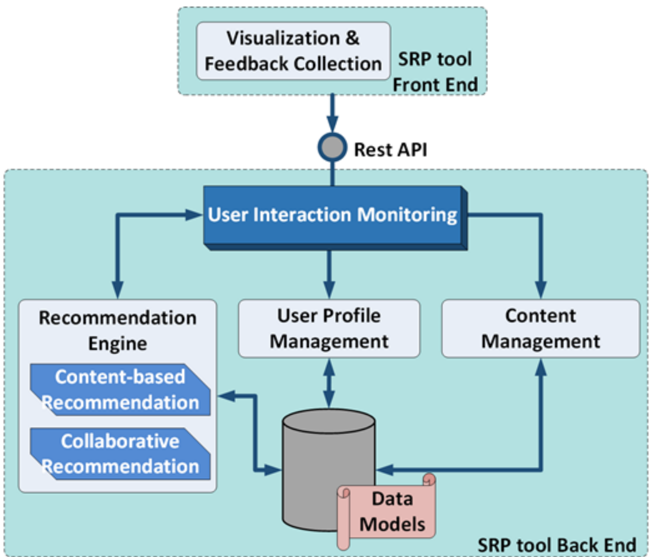


Figure 6. Architecture of the Architecture of the Social Recommendation & Personalization Tool.

Data Models: The database where all the data that the tool needs in order to operate seamlessly are stored.

Content Management: The module that processes the ingested content in order to provide a meaningful representation to the underlying algorithms.

User Profile Management: The module that keeps user profiles updated as far as their demographics and actual preferences are concerned, based on their interaction with the content and the platform.

Recommendation Engine: The core part of the tool where the recommendation process takes place and provides the users with the appropriate content.

Various state-of-the-art technologies were utilized in order to achieve the performance and security necessary for the optimal operation of the system. The software specifications for the SRP tool can be found in Table II.

TABLE II. SRP TOOL SOFTWARE SPECIFICATIONS

Licensing	**Open source**
Core Implementation Technologies	Python 3.5.2
Additional technologies utilised	Nginx server uwsgi Django 1.10 (Python framework) djangorestframework 3.5.1 gensim 0.13.4.1 Postgresql 9.5.7 Docker Docker-compose
Database details	PostgreSQL
Exposed APIs	REST
Exchanged data format	JSON
GUI description	GUI application communicating with the backend of the tool. Users have to signup/login to use the tool's backend functionalities.

The tool is developed as an open source project as well and the source code can be found in [17].

C. Functionality & Design

This section elaborates on the details regarding the features and mechanisms supported by the SRP tool. In order for the recommendation engine to work, the content must be properly indexed and the system should have information about the user's preferences. The Content Management module ingests the content's data and maps each content item to a vector as described later in this section. The interaction of the user with the content allows the creation of a similar vector for the user which later can be used to provide recommendations either on a personal level or for a specified target group. The rest of the section further elaborates on each of the functions performed by the SRP tool.

As already stated, the first process the SRP tool has to perform is to index the content in a meaningful way, an important step as also indicated in [18][19]. Each video/enrichment is mapped to a vector, the elements of which are the scores appointed to the video/enrichment expressing the relevance it has to each category of the defined categories. The categories used come from the upper layer of DMOZ (http://dmoztools.net/), an attempt to create a hierarchical ontology scheme for organizing sites, Since the videos in the PRODUCER project are of generic nature, a common ontology scheme seems fit. The feature terms used are presented on Table III.

TABLE III. FEATURE TERMS

Art	Business	Computer	Education	Game	Health	Home
News	Recreation	Science	Shopping	Society	Sport	Child

Each multimedia content item is therefore described as follows: $X_P = [X_{P_1}, X_{P_2}, \ldots, X_{P_N}]$, where P_i are the specified categories and X_{P_i} is the relevance the content has to the specific category. Each element of the vector X_P needs to be generated in an automatic way from the metadata accompanying the video since such a representation is not already available nor is manually provided by the content creators. To achieve this, a previous version of the tool used a naïve tf-idf algorithm while in the current version of the SRP tool, a more sophisticated approach is considered. More specifically, the X_P are appointed using the Word2Vec model [20] a model of a shallow two-layer neural network that is trained to find linguistic context of words. It takes as input a word and returns a unique representation in a multidimensional vector space. The position of the word in this vector space is such that words that share common contexts are located in close proximity with each other.

Since the multidimensional vector representation is not useful to us in the way it is, we apply the same procedure on the feature terms used in our vector representations. By doing so, each feature term also has a multidimensional vector representation on the same space as the words and the similarity between the word and each category can be computed. To calculate the overall similarity score, we use a linear combination between the maximum score from all words on the document and the average score of the words. The average score is used in order to reduce the chance that a word that appears few times in the text, but is very relevant to the category in question, skews the result too much in its favor.

In our algorithm we use a pre-trained model from the Wikipedia dataset which consists of millions of documents on a large variety of themes and as a result is a pretty generic dataset covering all the topics that are of interest.

In order to be able to identify content relevant to target audiences, the tool needs to collect information and preferences of viewers since user profiles constitute another integral part of a recommendation system. The representation of each user on the system follows the same principals as the content vector representation, where the vector's elements signify the importance each term has to the user. As a results each user is represented by a vector $U = [U_{P_1}, U_{P_2}, \dots, U_{P_N}], U_{P_i} \geq 0, \forall i$, where U_{P_i} is the value each user gives to each feature term.

Within the platform the SRP tool operates, the viewer registers and provides some important demographics (i.e., gender, age, country, occupation and education). This information is used in order to create an initial user vector for the user, based on the preferences of users similar to his demographics group. Alternatively, instead of providing this information explicitly, the viewer can choose to login with his/her social network account (e.g., Facebook, Twitter) and the information could be automatically extracted.

The user profile created via this process is static and is not effective for accurate recommendation of content since: a) not every user in the same demographic group has the same preferences and b) his/her interests change dynamically. Thus, in addition to the above process the SRP tool implicitly collects information for the user's behavior and content choices. Using information about the video he/she watched or the enrichments that caught his/her attention, the SRP tool updates the viewer's profile to reflect more accurately his/her current preferences.

The created user profile, allows the tool to suggest content to the viewer to consume, as well as a personalized experience when viewing the content by showing only the most relevant enrichments for his/her taste. Through a content-based approach, the user's profile is matched with the content's vector by applying the Euclidean similarity measure as:

$$sim_{up}^{cf}(i,j) = \frac{1}{1 + \sqrt{\sum_k \left(U_{i_k} - X_{j_k}\right)^2}} \tag{1}$$

where U_i is the user's profile vector and X_P^j is the content's vector. Other similarity metrics were also tested and will be presented in Section IV.

The collaborative approach is complementary with the content-based recommendation using information from other viewers with similar taste, to increase diversity. The idea is to use already obtained knowledge from other users in order make meaningful predictions for the user in question. To do so, the similarity between users is computed as follows:

$$sim_{uu}(i,j) = \frac{1}{1 + \sqrt{\sum_k \left(U_{i_k} - U_{j_k}\right)^2}} \tag{2}$$

where the H more similar users from the user's friends list are denoted as close neighbors. We then compute the similarity of the neighbors to the item:

$$sim_{up}^{cbf}(i,j) = \sum_{s=1}^{H} sim_{up}^{cf}(i,s) \cdot sim_{uu}(s,j) \tag{3}$$

The final similarity between the user and the item is calculated via a hybrid scheme by using the convex combination of the above similarities:

$$sim_{up}^{h}(i,j) = (1-\theta) sim_{up}^{cbf}(i,j) + \theta sim_{up}^{cf}(i,j) \tag{4}$$

where $\theta : 0 \leq \theta \leq 1$ is a tunable parameter denoting the importance of the content-based and the collaborative approach on the hybrid scheme. A value of $\theta = 0.5$ has been shown to produce better results than both approaches used individually [21].

Based on the collected data above and the constructed viewers' profiles, the producer of the documentary can filter the available content based on the preferences of the targeted audience. For this purpose, the k-means algorithm [22] is used to create social clusters of users. Based on the generated clusters, a representative user profile is extracted and is used to perform the similarity matching of the group with the content in question. The SRP tool assigns a score to each item and ranks the items based on that score.

After the creation of the documentary, the SRP tool can be used as an extra step in order to provide a filtering on the enrichments that are paired with the video, so that they do not overwhelm the viewer, filtering out less interesting ones. After specifying the target audience, the SRP tool can provide the list of suggested enrichments that the producer can either accept automatically or select manually based on his/her preferences, enabling the delivery of personalized documentary versions, tailored to audience interests.

D. Graphical User Interfaces

The Social Recommendation & Personalization tool provides a Graphical User Interface (GUI) in order to make it accessible to users willing to use the standalone version of the tool. In the integrated platform, the GUI is part of the platform in order to better exploit its potential by combining its services with that of the rest of the tools.

Since the tool needs some information about the users in order to efficiently provide its recommendations, a page where he/she can enter or alter his/her personal information is provided (Fig. 7). This information is used to initialize the user profile but will also be valuable when willing to gather information for a specific target group. When the user enters his/her information, the data is stored in the SRP tool database.

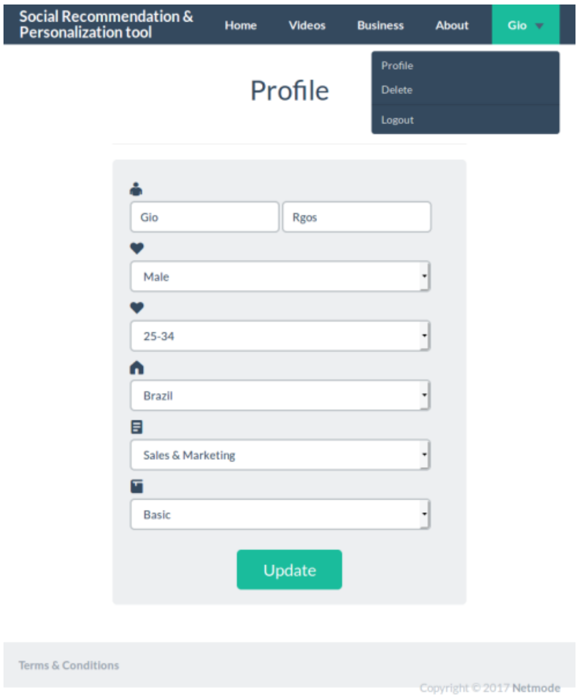


Figure 7. SRP tool login screen.

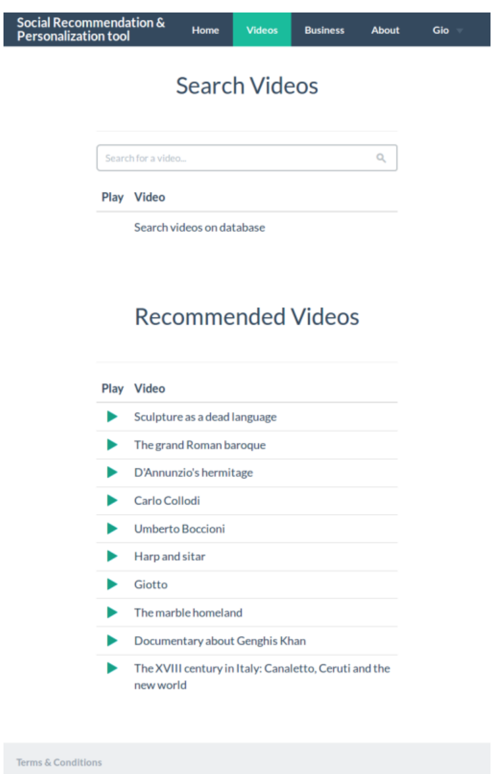


Figure 8. Recommended videos page of SRP tool.

By clicking on "Videos" from the navigation bar, a search bar for searching specific videos as well as a list of videos are presented to the user (Fig. 8). The list of the videos contains the top ten videos from the video database, ranked based on the profile of the user that requested the list by making use of the hybrid recommendation mechanism. It is thus subject to change every time the user interacts with the system, so that the top videos correspond to what the system believes are the most interesting videos for the user at any time.

The "Play Video" page contains more information about the video, as well as the video content itself (Fig. 9). From this page, the user can view the video, interact with it by sharing it to social media, like it or dislike it and watch the enrichments associated with the video. All information concerning the interactions of the user with the content is sent back to the SRP tool backend to update the profile of the user in order to be able to make more precise recommendations in the future.

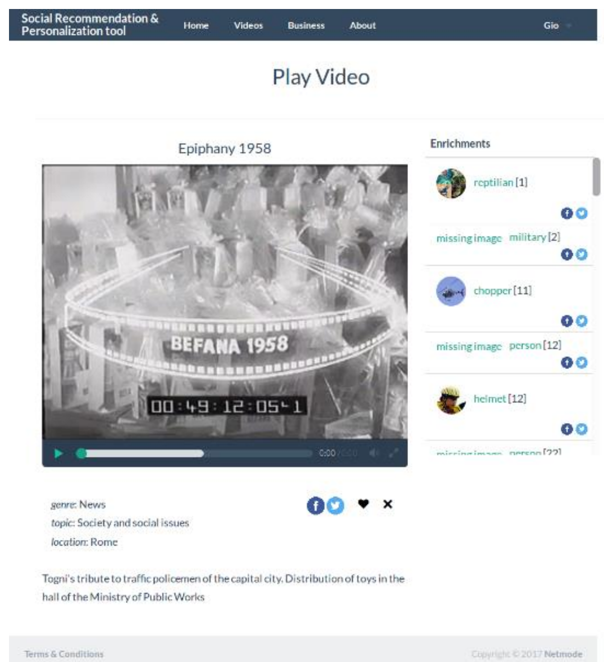


Figure 9. Play video page of SRP tool.

The last page provided by the GUI is to be used by the content providers or producers willing to use the services provided by the SRP tool (Fig. 10). The page is split in three columns. The leftmost contains a form where the user can select the audience group he/she wants to target in his/her documentary, so that the tool knows what recommendation to make. After choosing the appropriate values in the form, the user clicks on search and in the middle column, a list of the 10 most recommended videos for the target group appears. The list is ranked from most to least relevant. After selecting the appropriate video, the enrichments of the video appear on the right column. The tool gives the user the ability to select which ones of the suggested enrichments he/she finds appropriate for his/her documentary by toggling the slider at the top right of the enrichment. After making his/her selection, the user can export his/her choices for further use in the documentary creation process. In the integrated platform, the exported data could be used by the rest of the tools of the PRODUCER platform

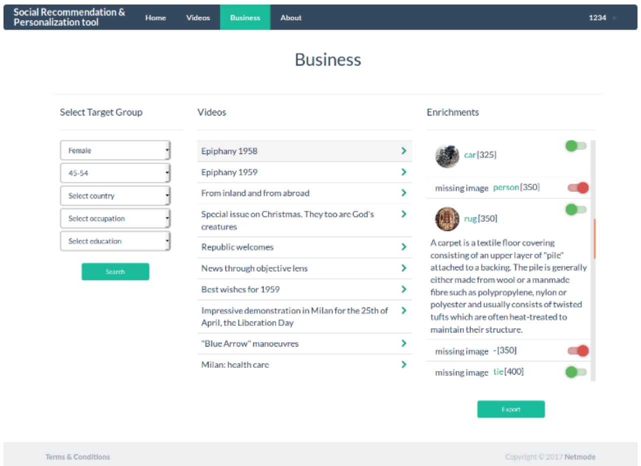


Figure 10. SRP tool page for Business users.

IV. EVALUATION & BENCHMARKING

In this section, an extensive evaluation of the two tools is presented in order to measure their performance and effectiveness on their corresponding tasks. In order to successfully evaluate the tools, both an objective benchmarking process via simulations on the underlying algorithms and a subjective benchmarking process by actual usage of the tools from real users were performed.

The reason for performing both offline and online evaluation techniques is that recommendation systems are relatively complex mechanisms and their performance cannot be holistically captured through their mathematical model representation. Offline benchmarking was used to configure the underlying algorithm and tweak the available parameters via measuring the effectiveness based on the state of the art metrics, while online experiments came as a confirmation to the above selections and captured the overall Quality of Service perceived by the users (e.g., cold start recommendations, over-specialization etc.)

A. Objective benchmarking

1) Integrated Trends Discovery Tool

The first set of evaluation actions for the ITD tool refers to the inference processes through data analytics approaches. In more details the inference algorithms were developed for: i) extracting audience's characteristics (gender) through Twitter data and ii) analyse popularity of targeted TV shows by the complementary use of Google Trends service with Twitter.

Extracting audience's characteristics

The rationale for identifying potential audiences' gender and age characteristics is that this kind of information is not freely available from social media services due to user privacy protection data policies. There are various state of the art attempts that focus on inferring user demographics though probabilistic approaches based on user related data freely available on social media (e.g., tweets content, linguistic features, followers' profile) [23][24][25][26]. With regards to the documentary preproduction phase, Twitter service proved to be the most appropriate one for extracting user profile information, as Twitter account data and content are openly available. The Facebook social media service recently updated the related data access policy and doesn't allow the access to user content if there is no direct relation with the user (e.g. friends). In a similar manner Google adwords service only provides access to user profile data strictly for mediating Google advertisements and doesn't allow the utilization of such data for other reasons to third parties.

The task of age and gender estimation is tackled by the ITD tool via the utilization of classification algorithms trained with ground-truth data sets of a number of tweeter users containing records of real Twitter profile information and the respective gender/ age. The core idea for the classification algorithm is that stylistic factors are often associated with user gender, so the Twitter profile colour that has been utilized in combination with the profile picture and the display name. The applied approach, which is presented in detail in [27], constitutes a scalable and fast gender inference mechanism, as a very limited number of features is being utilized for each user thus resulting to a low-dimensional space, in which the machine learning algorithms for gender detection operate. The core benefit of the proposed approach is that it is able to scale and process a very large dataset of Twitter users while is conclusive even in the case where only one of the three aforementioned profile fields used is specified.

The trained network is then utilized in order to generalize the training process and estimate missing information from wider networks of twitter users. The inference process is coordinated by the Trends Inference Engine. The engine uses the TwitterAPI to retrieve tweets where the keywords connected with certain topics are mentioned. Based on the respective Twitter Account ids, profile information is collected for each account. Based on profile attributes (e.g., "name", "screen_name", "profile photo", "short description", "profile_color") each user is classified to age & gender category and each classification is escorted by a confidence level.

To infer the gender of users based on their profile pictures, the Face++ Face Detection API (https://www.faceplusplus.com) is utilized. This service detects human faces within images and estimates the respective gender associated with a confidence level. To exploit the display name for determining the user's gender, a data matching technique is used comparing the names of Twitter users with the names stored in the datasets of Genderize (https://genderize.io/).

In order to exploit the theme color to infer the user gender, a hex color code has been obtained for each user via the Twitter API corresponding to the user's chosen color.

The obtained color codes have been converted to the corresponding RGB representation thus generating three features (capturing the respective Red, Green and Blue values of the theme color).

All aforementioned features were used to train three machine learning gender classifiers, namely a Photo Classifier, a Color Classifier and a Name Classifier, each exploiting the information gained from the features extracted from the corresponding field. The output of these classifiers is the inferred gender for each user, along with the respective estimation confidence level. In order to couple the outputs of all aforementioned standalone gender classifiers in a hybrid approach, three "gender numbers" have been assigned to each user, each capturing the output of one classifier.

The evaluation has been based on a public data set (https://www.kaggle.com/crowdflower/twitter-user-gender-classification) of ground truth data containing information of 10021 twitter users' profiles. The dataset contains the gender of distinct twitter users escorted by profile information.

In order to evaluate the gender inference algorithm, the initial dataset (~10000 records) has been divided into 40 parts each containing about 250 records. Each dataset part was gradually incorporated to the classifier, while the last 250 records were used for evaluation. The initial evaluation attempts did not provide high performance results. A data cleansing process was subsequently performed removing records that had the default predefined Twitter profile colors that resulted in a dataset of ~2000 records. The same evaluation process was then conducted where each of the 40 parts contained 50 records.

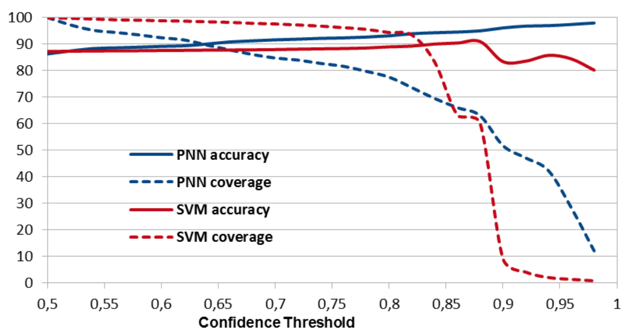


Figure 11. Accuracy and Coverage for PNN and SVM Hybrid Classifiers.

As it is presented in Fig. 11 and discussed in detail in [27], the evaluation process indicated that the utilization of two supervised learning algorithms namely the Support Vector Machines (SVMs) and Probabilistic Neural Networks (PNNs) perform excellent, resulting in ~87% accurate results. The evaluation process is planned to proceed with further testing of the proposed approach based on more datasets, originating from additional social media (not only Twitter), to compare with similar existing approaches and to incorporate additional user profile attributes, including text analysis of provided profile description and Tweets text.

Social Media and Google Trends in Support of Audience Analytics

One of the objectives of the ITD tool's inference engine is to improve the quality and reliability of the generated results by combining the outcomes of different sources of information. On the same time, there have been various research efforts aiming to investigate how social media are used to express or influence TV audiences and if possible to estimate TV ratings through the analysis of user interactions via social media. Based on the state of the art review [28], the research work conducted so far by various initiatives on this domain focuses mainly on the utilization of Twitter and Facebook. However, in certain occasions, the respective volume of information derived by these social media services is not enough resulting on low reliability outcomes. To this end, the second evaluation process of the ITD tool targets the case where the Twitter service is utilized in combination with Google Trends [29] towards the extraction of audience statistics for specific TV shows.

The analysis conducted for the Italian talent show "Amici di Maria de Filippi" that broadcasts for the last 17 years and lies among the most popular shows in Italy. The show airs annually from October until June, thus being appropriate for yearly examination of the data. In this study, data of the year 2017 have been used, split in two semesters as elaborated upon subsequently.

The keyword-hashtag that is utilized by audience is the '#amiciXX' where XX corresponds to the number of the consequent season that the show is aired. The analysis that was conducted by the ITD targeted the period January -June 2017 where the respective hashtag was '#amici16' and the period July to December 2017 where the respective hashtag was '#amici17'.

Using as keyword these hashtags and by utilizing the ITD tool, data were collected from Google Trends and Twitter. With regards to Google Trends, a time series of the relative search figures -search volume for the term divided by the total volume of the day- normalized between 0 and 100 were available by the service. Utilizing the Twitter API 882024 tweets collected for '#amici16' and 135288 for '#amici17' terms respectively. The collected data have been grouped based on date in order to acquire the daily volume.

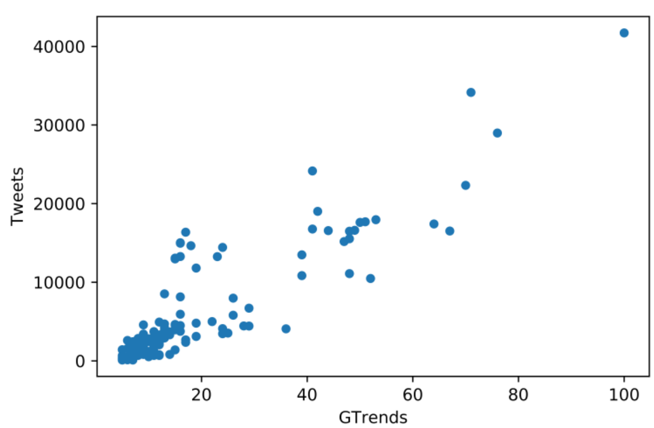


Figure 12. Correlation of Google Trends and Twitter data for the term '#amici16' targeting the first semester of 2017.

In order to verify the correlation between data originating from Google Trends and those originating from Twitter, the

Pearson correlation coefficient was utilized. The obtained results for the first semester of 2017 are illustrated in Fig. 12 and lead to coefficient of 0.893 and to significance of approximately 10-32. This indicates that the two datasets are strongly correlated, since we secured that the figures of each set are matched 1-1 and the low significance ensures that this result cannot be produced randomly. The respective outcomes for the second semester of 2017 are presented in Fig. 13 and lead to correlation coefficient of 0.816 and to significance of about 10-30. The slightly lower correlation demonstrated can be fully justified by the fact that the show does not broadcast during the summer and thus there is lower activity both on Twitter, as well as on Google, resulting in lower correlation results. Nevertheless, the findings indicate a strong relation between Twitter and Google Trends data. The aforementioned results confirm what the authors originally expected: Data obtained from Google Trends and Twitter at the same period are strongly (linearly) correlated and this of course can be further exploited in a variety of research purposes.

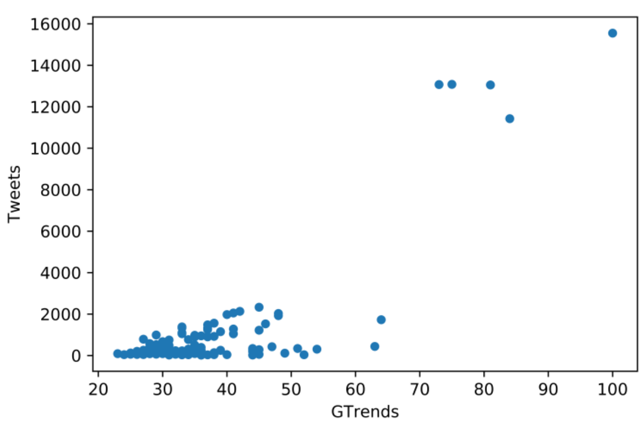


Figure 13. Correlation of Google Trends and Twitter data for the term '#amici17' targeting the second semester of 2017.

The described data homogenization and correlation evaluation mechanism has been integrated within the Inference Engine of the ITD tool allowing the dynamic deduction of whether the data from the two different information sources are converging or not for the utilized keywords that refer to the respective shows. The correlation level is then utilized as an additional value that is escorting the keyword presence volumes and presented to the end-user as an additional indication of the metrics' confidence.

The evaluation experiments conducted with regards to the overall utilization of the tool are encouraging and have allowed for the discovery of potential shortcomings early in the development phase. Such an issue is related to the volume of calls to external services. For example, Twitter API limits the allowed calls to 15 every 15 minutes per service consumer. As this issue was expected, a caching mechanism is utilized where results from each call to the Twitter API are also stored in the local database. Hence the ITD builds its own information store in order to avoid unnecessary calls. To this end, as the tool is utilized from various users, the local information store is getting richer.

2) Social Recommendation and Personalization tool

Concerning the evaluation of the Social Recommendation and Personalization tool, part of the benchmarking procedure was performed for the evaluation of the effectiveness of the algorithms used for the generation of the feature vectors of the content, that corresponds to the first process performed by our tool described in Section III, the indexing of the content in a meaningful way. In our tool, we represent the content as a vector, where each element is one of the 14 categories we have specified, and the value is the percentage to which the content is relevant to this category.

The models used in the evaluation process are four pre-trained models [30] on Wikipedia 2014 in glove representation [31] after we passed them from a transformation process to fit the Word2Vec representation, which contain a vocabulary of 400k words and 50 dimensions, a 100 dimensions, a 200 dimensions and a 300 dimensions vector representation respectively, as well as a pre-trained model on Google News with a vocabulary of 3 million words with a vector representation of 300 dimensions.

In order to test the efficiency of those models, in Section A.2.1, the default accuracy test of word2vec models questions-words [32] was performed while in A.2.2, the model was tested on the ability to effectively categorize content items on the 14 categories and a representative example from our dataset is presented. More examples can be found in [33].

Since the representation of the content is only part of the overall mechanism, an evaluation on the effectiveness of the recommendation algorithm as described in the rest of Section III was also performed. In Section A.2.3 the design of the evaluation process is described and in Section A.2.4 the state of the art metrics used for the evaluation are presented. Finally, in Section A.2.5 the results of the simulations are presented and discussed.

A.2.1. Question-words test

This test consists of 19544 sets of 4 words, and is used to test how well a generated vector model does with analogies of different kinds: For example, capital (*Athens Greece Baghdad Iraq)*, currency (*Algeria dinar Angola kwanza*) etc. The idea is to predict the 4th word based on the three previous ones.

Once vectors from a corpus with sentences containing these terms is generated, the question-words file can be used to test how well the vectors do for analogy tests (assuming the corpus contains these terms). So, given an example from question-words.txt (*Athens Greece Baghdad Iraq)*, the analogy test is to look at nearest neighbours for the vector

Vector(Greece) - Vector(Athens) + Vector(Baghdad)

If the nearest neighbour is the vector Iraq then that analogy test passes.

After running the question-words test for all five models, the successful and unsuccessful attempts of the algorithm have been recorded. The respective results are presented in Table IV.

TABLE IV. MODEL EVALUATION

Model	*Correct*	*Incorrect*
Wikipedia 50d	49.69%	50.31%
Wikipedia 100d	65.49%	34.51%
Wikipedia 200d	71.98%	28.02%
Wikipedia 300d	74.05%	25.95%
GoogleNews	77.08%	22.92%

All models perform pretty good with at least once in two successfully predicting the missing word for the smaller model (Wikipedia 50d 49.69%). What we notice is that the larger the model, the better the performance. Both larger vector representations and larger vocabulary contribute to the increase in the percentage of the correct predictions, as well as the quality and length of the corpus used to train the model.

As we can see from the results, the Google News model clearly performs the best with a success rate of 77% but due to its size, it is not very practical on small infrastructures such as the one used for our prototype.

A.2.2 Examples from our database

To test the efficiency of the Word2Vec model on the actual problem of finding the relevance that the video has in each of the 14 categories, we did some evaluations on the actual data we had in our video database. The idea behind the evaluation is to provide the title together with some tags and the description of the video, and the neural network should be able to successfully deduce this relevance. The more available metadata each video has, the better the result of the algorithm is expected to be. For this evaluation process, we used the Google News model which is the best performing one, and which we expected to have the most accurate representations.

A representative video example is presented in Table V.

TABLE V. PROPERTIES OF VIDEO EXAMPLE AND RESPECTIVE INDEXING DELIVERED BY SRP TOOL.

Title						
Documentary about Leonardo da Vinci						
Description						
Learn more about the life and the achievements of the Italian Renaissance polymath Leonardo da Vinci. His areas of interest included invention, painting, sculpting, architecture, science, music, mathematics, engineering, literature, anatomy, geology, astronomy, botany, writing, history, and cartography. He has been variously called the father of palaeontology, ichnology, and architecture, and is widely considered one of the greatest painters of all time. Sometimes credited with the inventions of the parachute, helicopter and tank, he epitomized the Renaissance humanist ideal						
Tags						
Sciences, History						
Art	Business	Computer	Education	Game	Health	Home
0.438	0.205	0.250	0.366	0.206	0.253	0.225
News	Recreation	Science	Shopping	Society	Sport	Child
0.168	0.253	0.753	0.132	0.319	0.194	0.339

In this example, a documentary provided by Mediaset is analyzed that concerns the life of Leonardo da Vinci. From the description provided we can see that he was a scientist as well as an artist, and so the algorithm gives a high score to "Science" and a lesser one but still high score to "Art" categories.

More details and examples of the multimedia content indexing delivered by the SRP tool are provided in [30].

A.2.3 Recommendation algorithm evaluation via simulations

In order to evaluate the performance of the algorithm used in the Social Recommendation and Personalization Tool, we also performed some offline experiments via simulations on MATLAB in a similar way as in [21]. In order to achieve this task, sets of content items are given a scoring on the 14 categories, and sets of users with a specified behaviour are created. Based on their behaviour, the users have different probabilities on performing actions on a content item, depending on the relevance and thus the likelihood that the user is interested in the item. Although the users are artificial, we make reasonable assumptions trying to emulate a real-life user behaviour.

In our simulation we have created 50 videos, having 8 enrichments and 8 advertisements each, and a feature vector of 14 categories. Videos are assigned into 5 classes, where in each class, $\left\lfloor \frac{F}{m} \right\rfloor = 2$ elements get a higher score, corresponding to different video topics (e.g., arts and science). 30 users are created to interact with the content and are again divided in 5 classes, in a similar way as the videos. Each user class implies different interests and preferences and so users that tend to select different videos and enrichments.

The simulation consists of 200 recommendation rounds where, in each round, a list of 6 most relevant videos according to the current profile of the user is presented him, in a ranked order. As already described in Section III, the hybrid recommendation approach we are using combines the content and the collaborative recommendation approach as follows:

$$sim_{up}^{h}(i,j) = (1-\theta) sim_{up}^{cbf}(i,j) + \theta sim_{up}^{cf}(i,j)$$

where θ is the tunable parameter.

For the collaborative part of the algorithm, we randomly assign 7 users as friends of each user and we use the 5 closest ones as his/her neighbours, which are the ones whose profile vectors are used to provide the collaborative recommendations.

As far as the similarity metrics are concerned, we perform a comparative evaluation between inner product, cosine and Euclidean similarities. More information on the similarity metrics and the respective results are presented in this section.

As mentioned, user behavioural vectors are used to simulate how users interact with the video, and more specifically 5 interactions are considered:

- Percentage of video watched
- Number of clicks on enrichments
- Number of share of enrichments
- Number of click on ads
- Explicit relevance feedback

These interactions are the same as the ones used in the actual tool.

Videos are watched by the user based on the video ranking the algorithm provides, and with a probability relevant to the video's rank and the user's behavioural vector, the user performs or not the above actions. The probabilistic nature of the process is used so that not all users perform all actions, as well as to capture the realistic tendency of users following particular behaviour based on their actual interest.

After the user has finished his actions, an update procedure follows, similar to the one described in [21]. It should be noted that most of the parameters have been chosen to provide the best results based on the work presented in [21], parameters that were also used on the implementation of the Social Recommendation and Personalization tool.

In order to reduce the randomness from our results, we run the experiment 10 times and calculated the average values on our figures.

A.2.4 Evaluation Metrics

The system is evaluated based on three metrics, in order to measure its effectiveness. The metrics used are the Profile Distance, the Discounted Cumulative Gain and the R-score [34] and are defined as in [21].

- ***Profile Distance***

The Profile Distance metric, measures the difference between the generated profile score of the users from the tool and the actual predefined profile score that corresponds to the actual interests and preferences of the user. In the simulations, this corresponds to the Euclidean distance of the user profile and the user behaviour vector. From the calculation of the metric we can see if the user vector converges to the actual interests through the constant update process based on the interactions of the user with the content and from its change over time, measure how fast, given a new user with no profile, this convergence takes place.

- ***Discounted Cumulative Gain***

Another method of evaluating the system is by measuring how "correct" is the ordering of the recommendations the tool provides to the specific user. Since actually knowing the correct ordering is impossible, we approximate it by assigning a utility score to the recommendations list, which is the sum of the utility score each individual recommendation has. The utility of each recommendation is the utility of the recommended item, as a function of the explicit feedback provided by the user, discounted by a factor based on the position of the recommendation on the list. This metric assumes that the recommendations on top of the list, are more likely to be selected by the user, and thus discount more heavily towards the end of the list.

In the Discounted Cumulative Gain, the discount, as we go down the list, follows a logarithmic function and more specifically,

$$DCG = \sum_i \frac{2^{r_i} - 1}{log_2(i+1)}$$

where i is the item position in the list and r_i is the user's rating on the item i. The base of the logarithm typically takes a value between 2 and 10, but base of 2 is the most commonly used [35].

- ***R-score***

The R-score follows the same idea of evaluating the "correct" ordering of the recommendations but instead of a logarithmic discount, it uses an exponential one. Since the items towards the bottom of the list are mostly ignored from the scoring, the R-score measure is more appropriate when the user is expected to select only a few videos from the top of the list.

The equation that is used for the calculation of the R-score is the following one,

$$R = \sum_i \frac{max(r_i - d, 0)}{2^{\frac{i-1}{a-1}}}$$

where i is the item position in the list, r_i is the user's rating on the item i, d is the neutral rating denoting the indifference of the user for the item ($d = 0$ in our tool), and a is a tunable parameter that controls the exponential decline [34].

A.2.5 Simulation Results

In the first part of the evaluation, we chose as similarity metric the Euclidean similarity and tuned the θ parameter for the hybrid recommendation scheme. The θ values used on this part of the experiment are:

- $\theta = 0$ for collaborative recommendation only,
- $\theta = 1$ for content-based recommendation only,
- $\theta = 0.5$ for the hybrid approach where both content and collaborative recommendations are equally taken into account.

Even though a similar evaluation was already performed in [21], in our evaluation, the collaborative recommendation part of the approach makes use of the "friends" concept where only a subset of the users is taken into consideration on the neighbour selection process.

In Fig. 14, one can see how the Profile Distance between the generated user profile and the expected one is affected with respect to theta. The smaller the distance, the more accurate the final representation of the user is, concerning his interests and preferences. As expected, the content-based only approach is the best performing one on this metric, while the hybrid approach's performance is close, since using only his own profile, the algorithm can easier tune it towards convergence. The least successful one is the collaborative approach only with significant distance from the other two, which is expected since the algorithm tries indirectly to deduce the user's profile through the profile of his friends. Even though the hybrid approach uses both content based and collaborative methods, its performance on the metric is more than satisfactory, while making use of the advantages provided by the collaborative method that we will discuss later on.

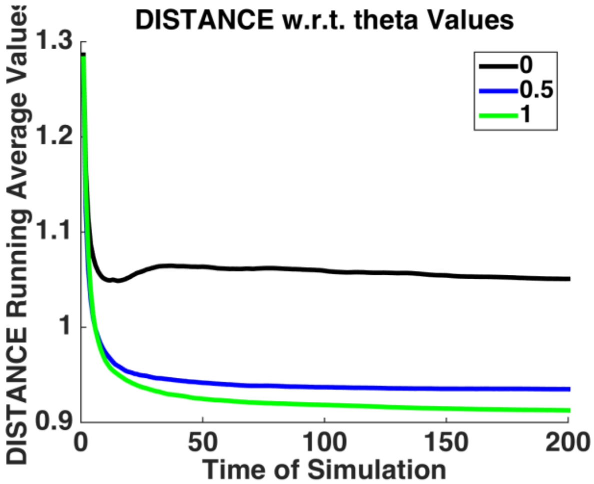


Figure 14. Average profile distance between the generated user profile and the expected user profile over simulation time for 3 different θ values.

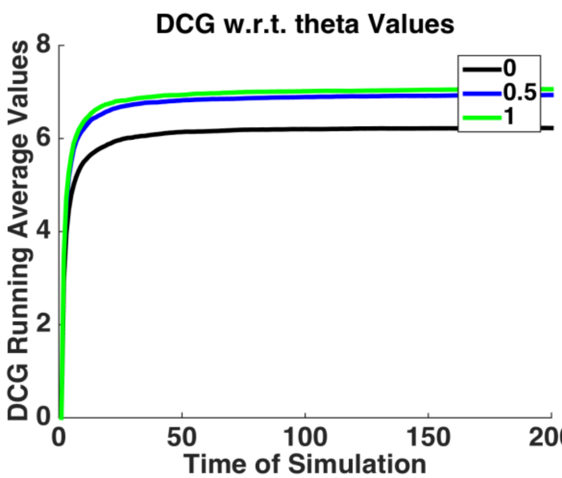


Figure 15. Average Discounted Cumulative Gain of the recommendations provided over simulation time for 3 different θ values.

Fig. 15 shows the Discounted Cumulative Gain of the recommendations provided over time. We can also see that the two best performing approaches are the content only and the hybrid approach, with the collaborative only following third. Again, the difference between the content only and the hybrid approach is not significant, validating once more the effectiveness of the hybrid approach.

Finally, in Fig. 16, we present the R-score of the recommendations list over time. The graphs follow the same pattern with the DCG, and so the hybrid approach succeeds in providing successful recommendations both on the total list and on the top recommended items.

The main disadvantage of using content-based only recommendations is the over-specialization of the algorithm on the user's choices. Collaborative filtering is important in introducing novelty and diversity in recommendations that allow the user to find interesting content that he would otherwise have missed. The element of surprise is important for a recommendation system and such diverse recommendations could lead a user in unexpected paths in his research as well as help him evolve his own taste and preferences. This fact cannot be easily captured in an offline experiment and requires online experimentation.

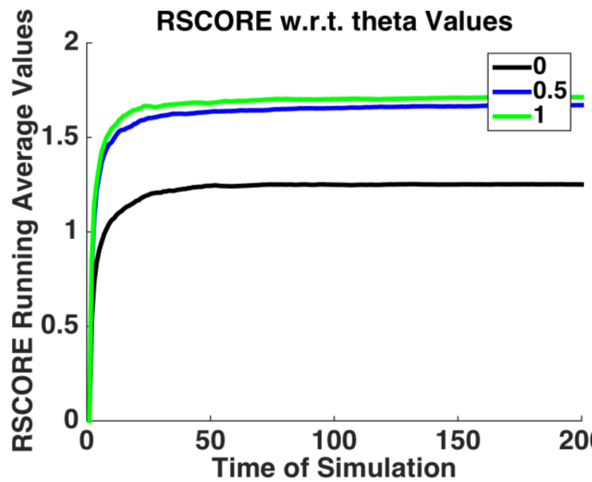


Figure 16. Average R-score of the recommendations list over simulation time for 3 different θ values.

Another problem the content-based only approach has to face is the cold start problem. When the system does not have enough information for a user, the system is basically unable to provide any meaningful recommendations. In this case, his friends network can be utilized to make use of information for users the system already has, and the recommendations provided are significantly more accurate. As a result, to overcome the problem, the collaborative approach seems effective.

From our analysis we can see that the hybrid recommendation scheme constantly achieves a smooth performance and thus successfully combines the advantages of both content and collaborative based filtering approaches.

For the next part of the evaluation, we compare the different similarity metrics used in our algorithms. In this experiment, we fix the theta parameter to $\theta = 0.5$ that corresponds to the hybrid recommendation scheme. An $input$ parameter is used in our simulation to specify the similarity measure used by our algorithms and corresponds to:

1. Inner product similarity
$$similarity = X \cdot Y$$
2. Cosine similarity
$$similarity = cos(\theta) = \frac{X \cdot Y}{\|X\|\|Y\|}$$
3. Euclidean similarity
$$similarity = \frac{1}{1 + d(X,Y)}$$
$$d(X,Y) = \sqrt{\sum_i (x_i - y_i)^2}$$

where $d(X,Y)$ is the Euclidean distance of the two vectors.

In Fig. 17, we can see that the Euclidean similarity is the best performing similarity measure, achieving a slightly better score than the cosine similarity, while the inner product similarity is the worst performing. What's more, the Euclidean similarity seems conceptually more appropriate in our use case, since each user and each item can be modeled as a point in the 14-dimensional metric space and the closer they are on the space, the more similar they are.

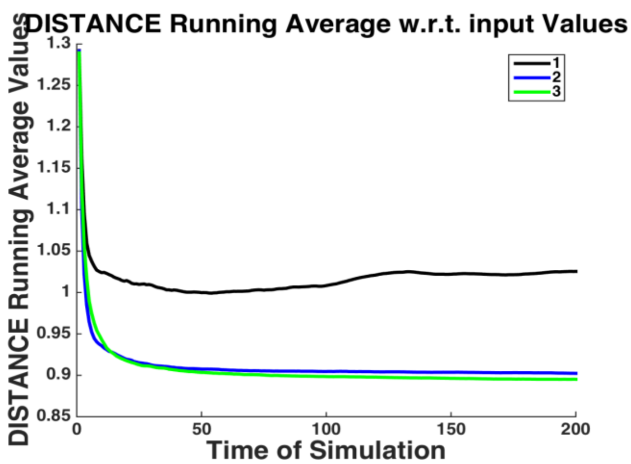


Figure 17. Average profile distance between the generated user profile and the expected user profile over simulation time for 3 different similarity metrics: 1) inner product similarity, 2) cosine similarity, 3) Euclidean similarity.

The Discounted Cumulative Gain is depicted in Fig. 18 and follows the same trend, showing that the Euclidean similarity outperforms the other two similarity measures by providing better overall recommendation lists to the user. The inner product, which is the simplest one, still performs worse than the rest.

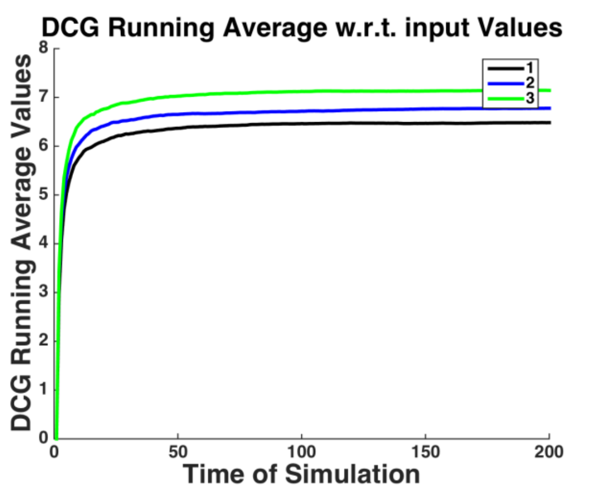


Figure 18. Average Discounted Cumulative Gain of the recommendations provided over simulation time for 3 different similarity metrics: 1) inner product similarity, 2) cosine similarity, 3) Euclidean similarity.

Finally, concerning the R-score (Fig. 19), the Euclidean and the cosine similarity achieve the highest score with minor differences, while the inner product achieves significantly lower score. The fact that the two first measures perform almost the same while in the DCG metric the Euclidean performs better, shows that the Euclidean similarity can better fine tune the lower scoring recommendations since even the lower scoring items, that the R-score ignores, are more likely to be more relevant to the user's preferences.

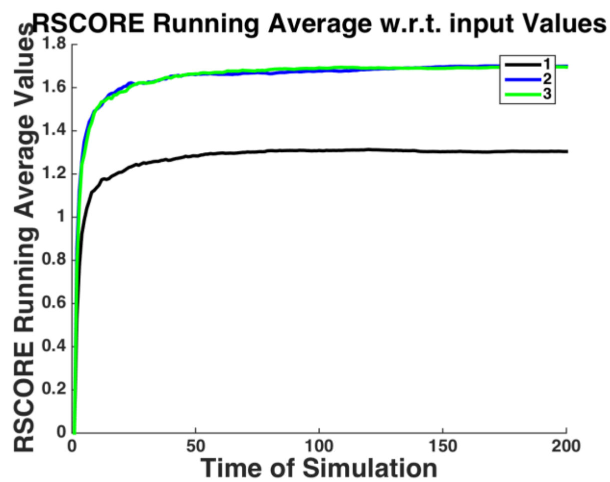


Figure 19. Average R-score of the recommendations list over simulation time for 3 different similarity metrics: 1) inner product similarity, 2) cosine similarity, 3) Euclidean similarity.

More simulations concerning the parameters used can be found in the work presented in [21].

B. Subjective benchmarking

1) Integrated Trends Discovery Tool

The Integrated Trends Discovery Tool was evaluated by numerous individuals that were mainly students from the National Technical University of Athens, which ICCS is affiliated with. The students were mainly coming from the Techno Economics Masters program[1], jointly offered by the Department of Industrial Management and Technology at the University of Piraeus and the National Technical University of Athens, which is a highly interdisciplinary graduate programme targeted at professionals with existing market/business/working experience. The evaluation process included the following steps:

a) A document describing the core concepts of the PRODUCER project and the core innovations of the ITD tool was initially shared with the testers.

b) After reading the document the testers watched a 10-minute video demonstrating the utilisation of the ITD tool. The video contained textual information about the internal mechanisms that contribute in generating the visualised outcome at the front end of the tool.

---

[1] http://mycourses.ntua.gr/course_description/index.php?cidReq=PSTGR1083

c) Finally, the testers answered an online Google Forms based questionnaire. The questionnaire is available under [36].

This process was completed by 157 individuals. In addition, another group of 20 individuals, after following steps a) and b), were requested to access a live version of the tool and to freely try the various functionalities. Then they proceeded on step c) and answered the same questionnaire as well. The results from the superset containing both user groups (177 individuals) are presented in the following figures. As depicted in Fig. 20, the ITD tool testers were mainly young persons (18-34 years old), and are in principle students and/or full-time employees. Their current occupations are mainly related to engineering, IT, and business/financial as presented in Fig. 21.

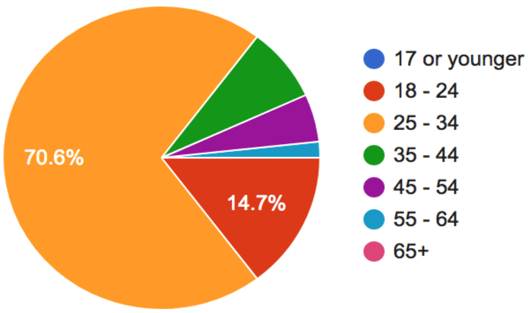


Figure 20. Ages of the user group that tested the Integrated Trends Discovery Tool.

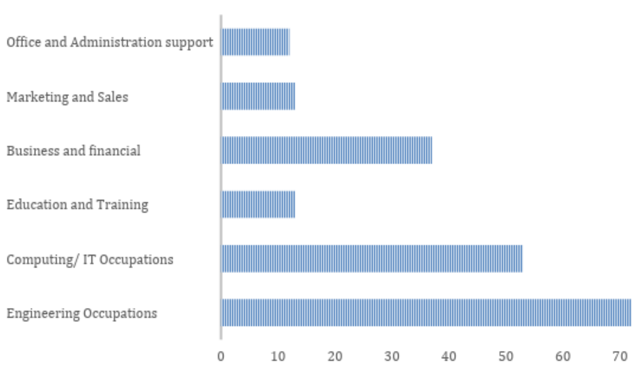


Figure 21. Occupation of the user group that tested the Integrated Trends Discovery Tool.

All testers are familiar with the concept of social media services as they utilize them for long time period (more than five years) and for 1 to 4 hours per day (Fig. 22, 23). In addition, most testers are highly interconnected with other users, having more than 100 connections (Fig. 24), and seem to prefer Facebook, LinkedIn, Google, Instagram and Twitter (Fig. 25).

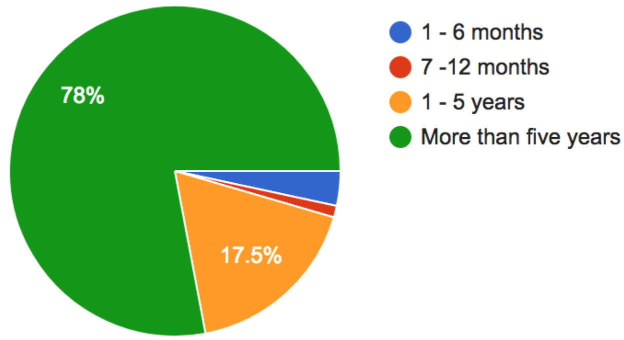


Figure 22. Time period of using Social Media Services.

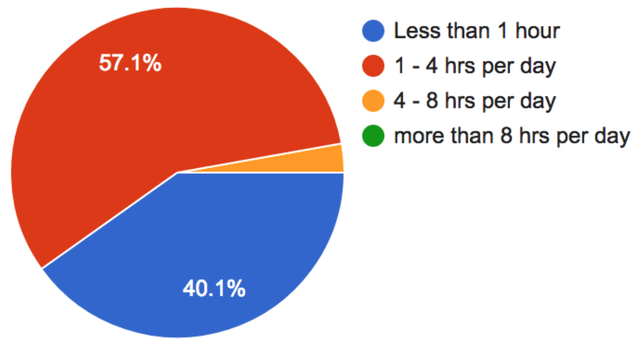


Figure 23. Time of usage per day of Social Media Services.

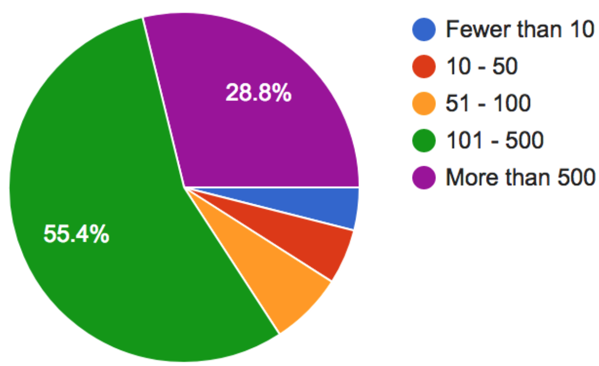


Figure 24. Number of connections each user has on his Social Media profiles.

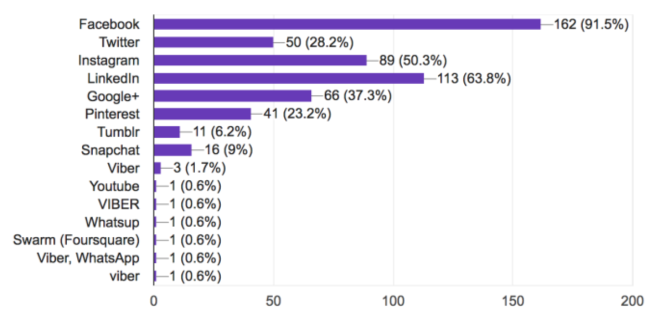


Figure 25. Social Networking Sites used by the user group.

Testers questioned about their purpose of Social media services utilization. Their replies are presented in Fig. 26. Replies such as: “To get opinions”, “To find information”, “To share your experience” are concentrating a significant amount of answers something, which is important because these views are in support of the core objectives of the ITD tool. The core concept of the ITD tool is based on the fact that it is possible to gain information about population opinions and interests through mining social media and search engines services.

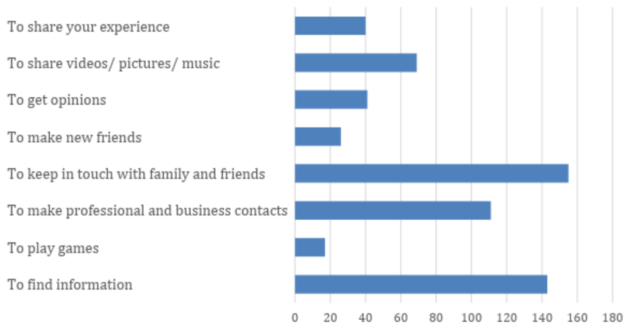


Figure 26. Purpose of using Social Media Services by the user group.

On the other hand, most testers consider that social media analytics can support the extraction of information regarding public opinion similar to the information extracted via opinion polls by survey companies (Fig. 27).

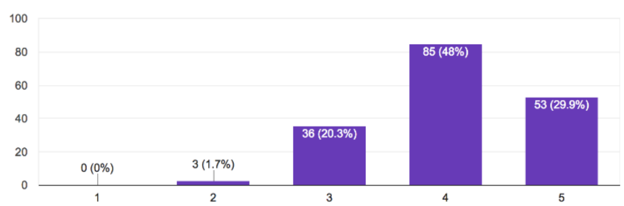


Figure 27. Do you think that Social Media analytics can support the extraction of information regarding public opinion (similar to the information extracted via opinion polls by survey companies)?.

The next question was about testers’ experience on using similar tools (Fig. 28), to which the users indicated they have limited or no experience in average.

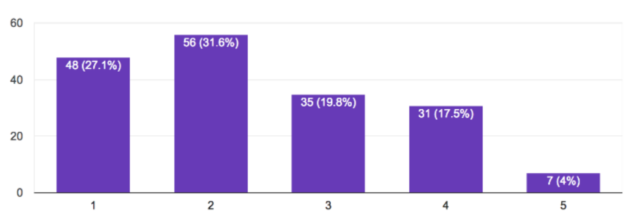


Figure 28. Evaluators’ level of experience in using tools that attempt to discover and process popularity/trends in Social Media and Search Engines.

The final question was about the ethical consequences on social media opinion mining. The actual question was: “The Integrated Trends Discovery Tool processes data that are freely available on the Internet but originate from users posts and searches. Do you consider that any ethical issues arise in this data aggregation process? Which of the following covers your opinion the most?”. Results illustrated in Fig. 29 show that most of the testers do not see any ethical issues, but a significant amount of replies considers that there are such issues. The ethical concerns of the users that appear to be significant introduce a major challenge that is further promoted by the General Data Protection Regulation (GDPR) (EU 2016/679) that took effect on May 2018 in Europe.

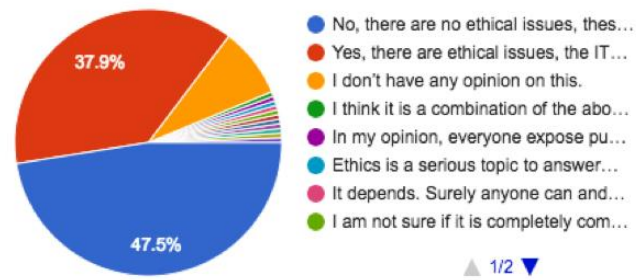


Figure 29. Ethical issues in the data aggregation process of the Integrated Trends Discovery Tool.

The next set of questions targeted directly on the tool utilization and underlying functionality. The first question was about how easy was for the testers to manage “Query Descriptions”. In order to create a new query process, users need to add the necessary information, e.g., textual description, targeted keyword, time range, targeted regions and provide parameters about inference of higher level information. Respective replies about ease of creating a new query process are presented in Fig. 30. Testers’ replies are based on a scale from 1 to 5 where 1 corresponds to “Very difficult” and 5 to “Very easy / intuitive”. Similar are the obtained findings concerning easiness with regards to managing existing Queries, as well as with respect to the generation of trend-related results.

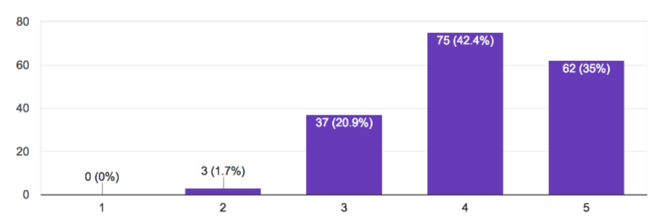


Figure 30. Ease of creating a new query at the "Add Query Parameters" page of the tool.

These findings indicate that the query configuration process was characterized as easy and/or very easy for the majority of the evaluators. The next question was about the ease of reading and understanding the results. Given that rendered results are the outcome of the integration of diverse statistical models derived from external APIs utilizing heterogeneous data models, this task was the one of the most challenging. Within the lifetime of the project we followed various iterations of design, evaluation and refinement of the way that the trend discovery results are presented to the end user. For this reason, various intuitive graphs (times series graphs, bar charts, pie chart, node graphs) are utilized in order to make the results comprehensible to users that are not demonstrating a background in statistics or in data engineering. The outcome of this evaluation is presented in Fig. 31 and most of the tool evaluators find the results reading process relatively easy.

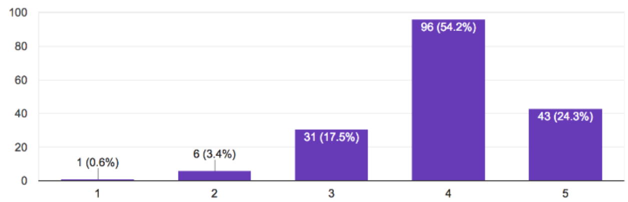


Figure 31. Ease of reading the results.

The last question related to the user interaction was "How user-friendly is the Integrated Trends Discovery Tool?" in general. The respective results are presented in Fig. 32.

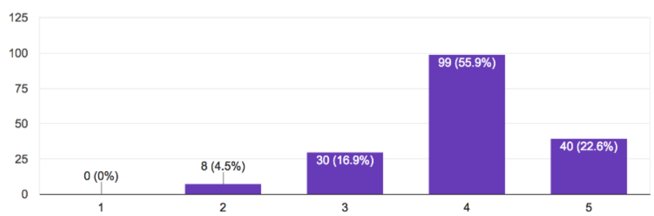


Figure 32. Overall user-friendliness of the Integrated Trends Discovery Tool.

As already described, evaluators at the first steps of the overall process had to read a textual description of the ITD tool objectives, which were also presented in the first minutes of the video describing the tool's utilization. Based on the presented list of innovations and after the demonstration and actual utilization of the tool, evaluators replied two different questions having the same target. The questions were: "*How successful is the Integrated Trends Discovery Tool in performing its intended tasks?*" and "*Meets expectations as these are defined in the innovations list presented upon video start*". Results are presented in Fig. 33 and Fig. 34.

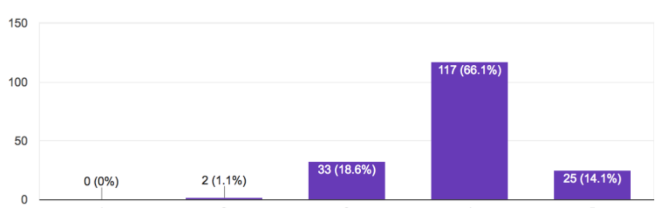


Figure 33. How successful is the Integrated Trends Discovery Tool in performing its intended tasks.

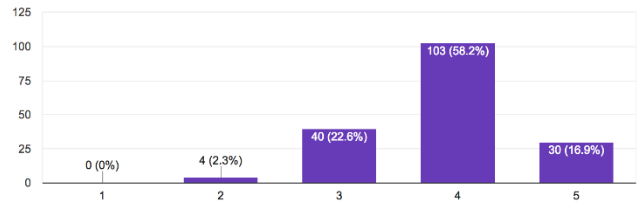


Figure 34. Meets expectations as these are defined in the innovations list presented upon video start.

The last question with regards to the actual evaluation of the tool was related to the overall software quality as this is disclosed through the execution of various tasks. Since this is a difficult question for evaluators with non-technical background, it was considered as optional and hence it was not replied by the whole set of testers. The respective results are illustrated in Fig. 35.

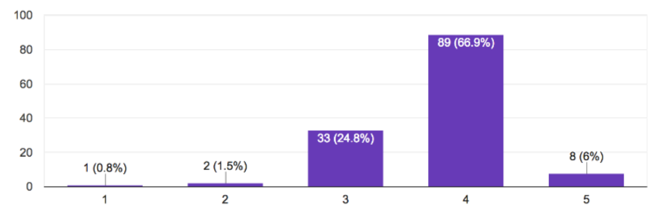


Figure 35. 65: Evaluate overall software quality.

ITD tool developers aim to continue the refinement of the service and to extend the provided functionalities. To this end, evaluators were questioned on which of the provided reports are the more useful. The responses are illustrated in Fig. 36.

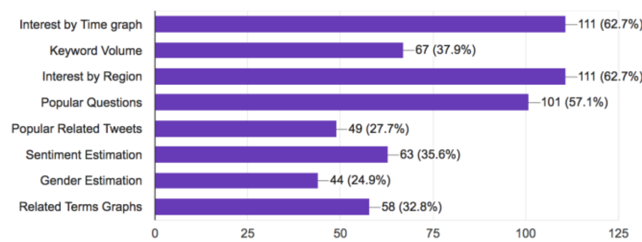


Figure 36. The Integrated Trends Discovery Tool provides various reports. Which are the more useful for you?.

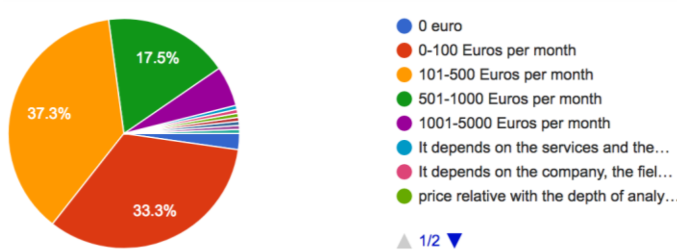


Figure 37. Estimation of cost in order to utilize ITD tool in business environment.

Finally, evaluators were questioned: “The Integrated Trends Discovery Tool currently utilizes mainly the free versions of public APIs (e.g., Google API, Twitter API, ...). Hence there are often delays and matters related to limited access to data. Do you believe that a company interested in the tool's results would be willing to purchase more advanced services (e.g., more detailed user demographics, data from larger user populations, data that span longer to the past) for an additional fee? If so, which of the following amounts do you consider as appropriate for the needs of a small company?”. The outcome of 177 responses is illustrated in Fig. 37.

2) Social Recommendation and Personalization Tool

For the evaluation of the SRP tool, 143 students from the same set of users used for the evaluation of the Integrated Trends Discovery Tool used the tool and answered the corresponding questionnaires [37]. The demographics of the aforementioned user base can be seen in Figs. 38, 39, 40.

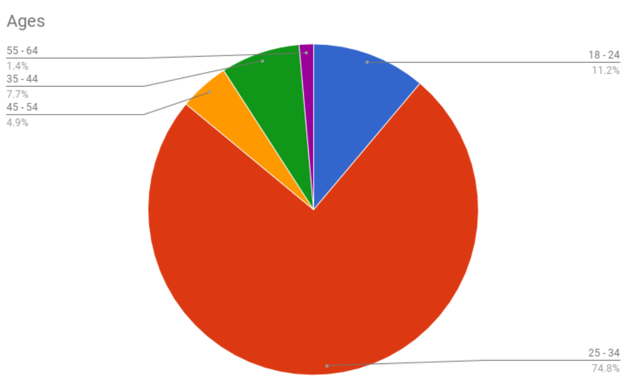
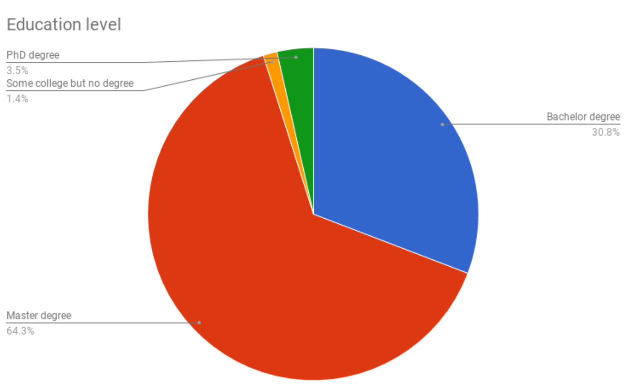


Figure 38. Ages of the user group that tested the Social Recommendation and Personalization Tool.



Education level of the user group that tested the Social Recommendation and Personalization Tool.A short video showing the functionalities of the tool and the expected interaction from the users was shown to the users and they were expected to use the tool on their own via its standalone GUI. After exposing themselves to the tool and using it until they were satisfied that they had formed an opinion on its capabilities, they were asked to respond to the corresponding questionnaire.

The experience of the users that participated in the process on recommender systems is shown in Fig. 41, confirming that a reasonable user diversity was well achieved.

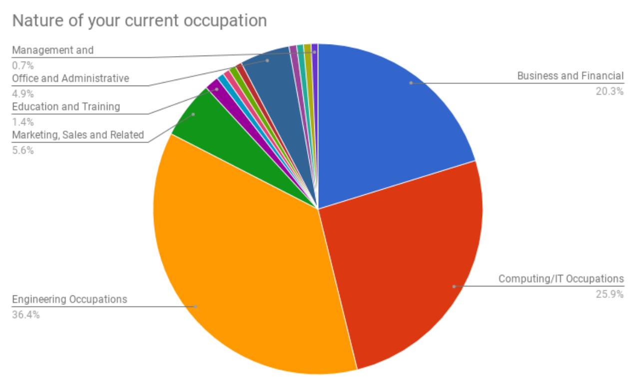


Figure 39. Occupation of the user group that tested the Social Recommendation and Personalization Tool

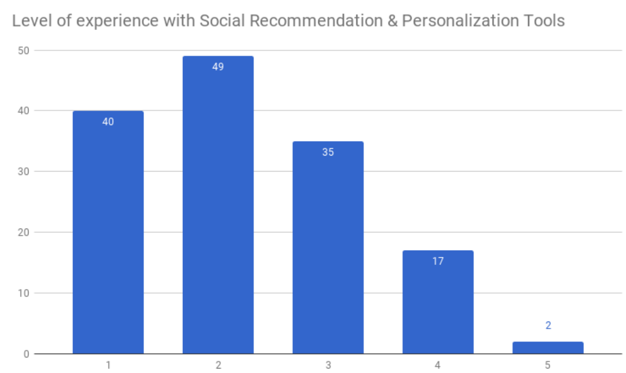


Figure 40. Level of experience with Social Recommendation and Personalization Tools (1: no experience, 5: much experience).

Users were asked to create an account on the tool inserting their information in order to create the basic profile. The information required are certain demographics (age, country etc.) and some personal information (name, email etc.) as well as a username and a password. The information required to be manually inserted by the users were limited, as can be confirmed by the responses of the users (Figs. 42, 43).

After creating his/her account, he/she continued to explore the actual functionalities of the tool. By clicking on the “Videos” tab, two options were available. On the one hand, the user could see the recommended videos that the tool suggests based on the profile the tool has created until now. In the beginning, the profile was created based on the demographics chosen by the user, so that content relevant to similar users was presented. On the other hand, a search functionality was available, where the user could search the database of the SRP tool of more than 2600 videos by providing text relevant to what he/she was searching for. The concept was to use the search functionality together with the recommended videos and based on the interaction the user had on the videos, the tool should be able to deduce the user's profile and suggest relevant videos to his/her interests.

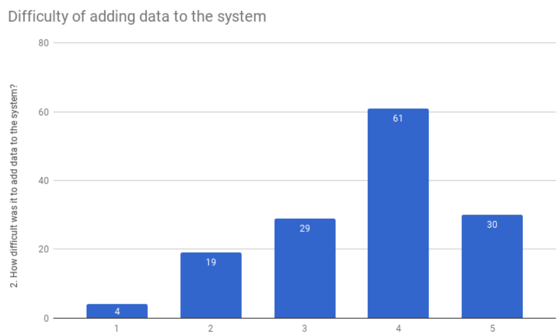


Figure 41. Difficulty of adding data to the system (1: very difficult, 5: very easy).

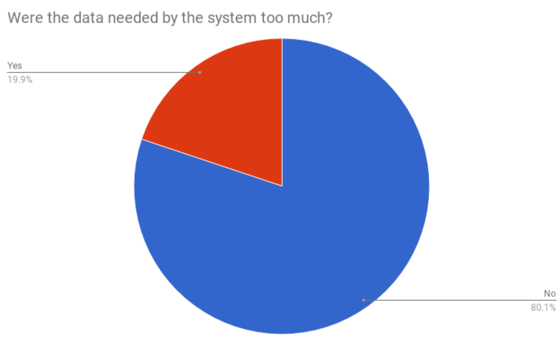
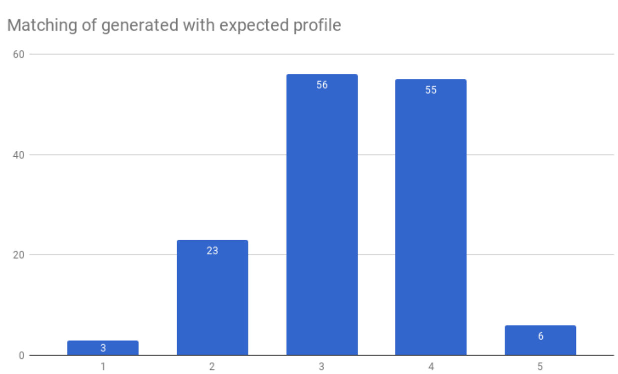


Figure 42. Were the data needed by the system too much?.

After some iterations of using the tool, the users had to rate the relevance of the recommended content and the user's interest in each of the 14 categories presented. The results of the procedure can be seen in Fig. 44 and Fig. 45.



Matching of the generated with the expected user's profile (1: unacceptable, 5: excellent).In both Fig. 44 and Fig. 45, we see that the majority of the users rate the tools performance as more than satisfactory. In Fig. 44, 39% of the users rated the profile matching generated by the tool and the one they had in mind while using the tool with 3 starts while 38% rated it with 4 stars. On the other hand, in Fig. 45, the matching of the recommended videos to the user's expectations shows again that the majority was satisfied, with a rating of 3 stars for the 39% and of 4 stars for the 36%. It is important to note that many times, the actual content of the video was rated by the users, something that is not important to the functionality of the tool, and so there could be some misinterpretation of the actual question. The limited availability of content could also play an important role in the results of the above questions.

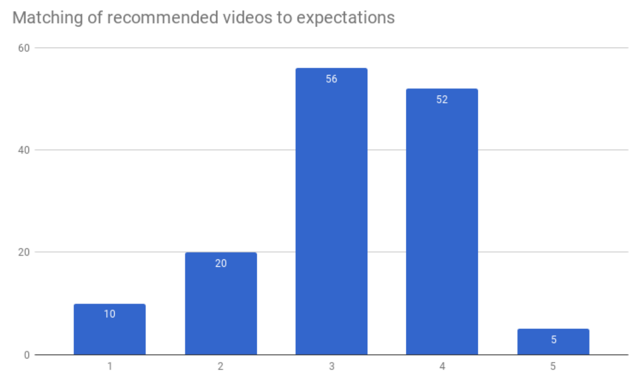


Figure 43. Matching of the recommended videos to the user's expectations (1: unacceptable, 5: excellent).

When asked about the overall Quality of Experience they had while using the tool, 49% of users rated the system with more than 4 stars (4 or 5 stars) stating that the Quality of Experience was more than satisfactory (Fig. 46).

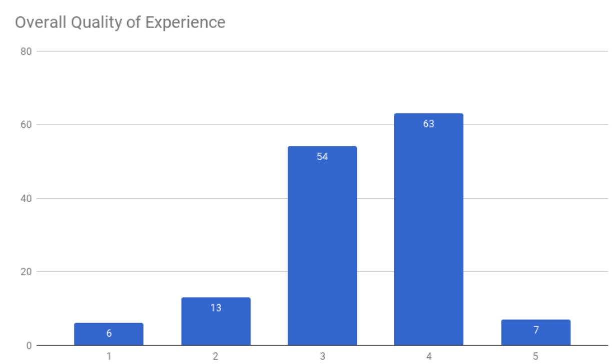


Figure 44. Overall Quality of Experience (1: unacceptable, 5: excellent).

One very interesting result coming from the questionnaires, is the importance the users give on such recommendation systems on a documentary content provider platform such as the PRODUCER platform (Fig. 47, Fig. 48). According to the graph, the Social Recommendation and Personalization tool provides a highly appreciated feature of the platform that definitely increases the Quality of Experience of the user, while helping him achieve tasks faster and more efficiently.

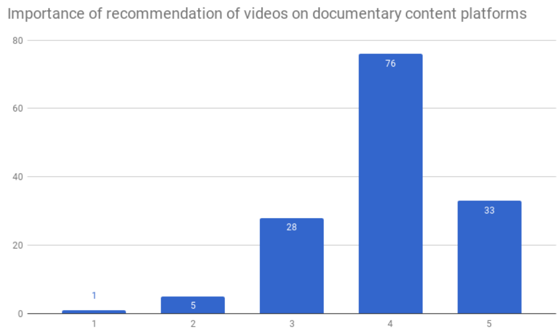


Figure 45. Importance of recommendations on videos (1: not essential, 5: absolutely essential).

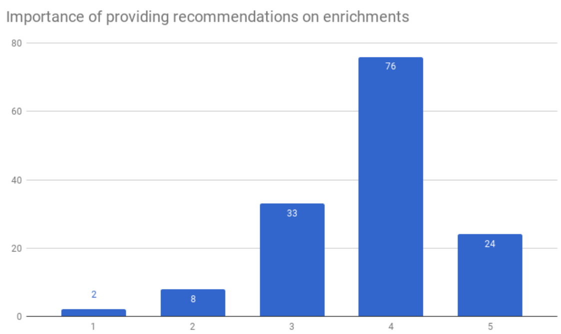


Figure 46. Importance of recommendations on enrichments (1: not essential, 5: absolutely essential).

Finally, users were asked about the relation that they expect between the video content and the enrichments that are recommended to the user by the tool. As we can see from Fig. 49, the majority has responded that they would like a balance between being relevant to the video content and the user profile, which shows that they are open to having recommendations that are more loosely tied to the content itself.

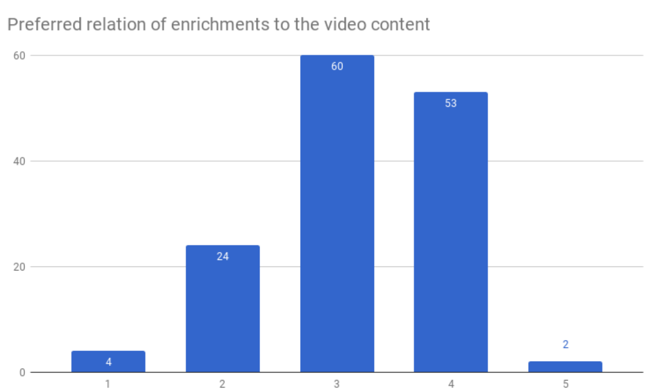


Figure 47. Preferred relation of enrichments to the video content (1: Tightly related to video content, 5: Tightly related to user profile).

Recommending something slightly out of context as far as it is of interest to the user seems to be an option opening some interesting research topics for future exploration. Adding the capability to tune that relation based on user's actions or the nature of the content could seem appropriate.

V. CONCLUSIONS

This paper analyses two software tools that aim to modernize the documentary creation methods. Initially the ITD tool is presented, which focuses on the targeted audience interests, identification and satisfaction. The ITD tool allows the identification of the most engaging topics to specified target audiences in order to facilitate professional users in the documentary preproduction phase. The SRP tool significantly improves the viewers' perceived experience via the provision of tailored enriched documentaries that address their personal interests, requirements and preferences. The core innovations of these tools and the delta from previously published work of the authors can be summarized as follows. First, both tools are used to reduce cost for the documentary production by filtering the content provided on both pre-production and post-production phases. Second, the ITD tool supports the reorientation of the documentary early on the preproduction phase based on the interests of potential audiences, thus targeting topics likely to attract larger audiences. Third, the ITD tool is designed to couple the knowledge extracted from several social media networks to investigate the audience's interests and identify the respective trends. This has already been tested over Twitter and Google Trends. Fourth, the ITD tool is also used to extract information regarding the user demographics, based on their interactions with social media. Evaluation results concerning the discovery of user gender have been presented. Fifth, the SRP tool exploits a different indexing method to classify the content on the 14 categories using NLP and the Word2Vec model instead of a naïve tf-idf algorithm. This has not been investigated before and it proves to be quite efficient in terms of performance. Sixth, the SRP tool supports collaborative filtering making use of the friends' network of the user instead of the entire user database, which enhances the performance of the proposed approach. Seventh, the evaluation of the SRP tool was performed based on different similarity metrics resulting in favouring the Euclidean similarity over the cosine similarity. Usage of this metric further enhanced the SRP tool's performance.

The prototype implementations of these two tools have been demonstrated and evaluated over a period of 3 months by end users of varying profiles. The evaluation process provided valuable feedback for further improving the overall functionality of the tools but also for the specification of reliable exploitation channels and the identification of related business opportunities

Future plans include the tools' integration with proprietary documentary production support services and infrastructures, as well as the extension of various stand-alone features that have been identified as more interesting and useful during the evaluation process. Moreover, the integration of additional social media networks and open data repositories to enhance the accuracy of the trends and interests identified by the two tools also lies among the authors' future plans. Finally, the authors plan to investigate the suitability of the tools for domains other than documentary production, adapt them and evaluate their performance in these domains.

ACKNOWLEDGMENT

This work has been partially supported by the European Commission, Horizon 2020 Framework Programme for research and innovation under grant agreement no 731893.

REFERENCES

[1] G. Mitsis, N. Kalatzis, I. Roussaki, E. E. Tsiropoulou, S. Papavassiliou, and S. Tonoli, "Social Media Analytics in Support of Documentary Production," 10th International Conference on Creative Content Technologies (CONTENT 2018) IARIA, Feb. 2018, pp. 7-13, ISSN: 2308-4162 ISBN: 978-1-61208-611-8.

[2] The PRODUCER project, http://www.producer-project.eu, 2017. [*Retrieved May 2019*]

[3] G. Mitsis, N. Kalatzis, I. Roussaki, E. E. Tsiropoulou, S. Papavassiliou, and S. Tonoli, "Emerging ICT tools in Support of Documentary Production," 14th European Conference on Visual Media Production, Dec. 2017, ISBN 978-1-4503-5329-8.

[4] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," Nature 457, pp. 1012-1014, 2009, doi:10.1038/nature07634.

[5] A. J. Ocampo, R. Chunara, and J. S. Brownstein, "Using search queries for malaria surveillance, Thailand," Malaria Journal, Vol. 12, pp. 390-396, 2013, doi:10.1186/1475-2875-12-390.

[6] S. Yang, M. Santillana, J.S. Brownstein, J. Gray, S. Richardson and S.C. Kou, "Using electronic health records and Internet search information for accurate influenza forecasting," BMC Infectious Diseases, Vol. 17, pp. 332-341, Dec. 2017, doi:10.1186/s12879-017-2424-7.

[7] F. Ahmed, R. Asif, S. Hina, and M. Muzammil, "Financial Market Prediction using Google Trends," International Journal of Advanced Computer Science and Applications, Vol. 8, No.7, pp. 388-391, 2017.

[8] N. Askitas and K. F. Zimmermann, "Google econometrics and unemployment forecasting," Applied Economics Quarterly, Vol. 55, pp. 107-120, Sep. 2009.

[9] S. Vosen and T. Schmidt, "Forecasting private consumption: survey-based indicators vs. Google trends," Journal of Forecasting, Vol. 30, No. 6, pp. 565–578, Jan. 2011, doi:10.1002/for.1213.

[10] S. Goel, J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts, "Predicting consumer behavior with Web search," National Academy of Sciences, Vol. 107, No. 41, Oct. 2010, pp. 17486–17490, doi:10.1073/pnas.1005962107.

[11] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," Fourth International AAAI Conference on Weblogs and Social Media, May 2010, pp. 122–129.

[12] M. X. Hoang, X. Dang, X. Wu, Z. Yan, and A. K. Singh, "GPOP: Scalable Group-level Popularity Prediction for Online Content in Social Networks," 26th International Conference on World Wide Web, Apr. 2017, pp. 725-733, ISBN: 978-1-4503-4914-7.

[13] A. Oghina, M. Breuss, M. Tsagkias, and M. de Rijke, "Predicting IMDB movie ratings using social media," 34th European conference on Advances in Information Retrieval Springer-Verlag, Apr. 2012, pp. 503-507, doi:10.1007/978-3-642-28997-2_51.

[14] B. Bhattacharjee, A. Sridhar, and A. Dutta, "Identifying the causal relationship between social media content of a Bollywood movie and its box-office success-a text mining approach," International Journal of Business Information Systems, Vol. 24, No. 3, pp. 344-368, 2017.

[15] Source code for Integrated Trends Discovery tool, https://github.com/nikoskal/itd_tool [*Retrieved May 2019*]

[16] J. Osofsky, "After f8: Personalized Social Plugins Now on 100,000+Sites," https://developers.facebook.com/blog/post/382, 2010. [*Retrieved May 2019*]

[17] Source code for Social Recommendation and Personalization tool, https://github.com/vinPopulaire/SRPtool [*Retrieved May 2019*]

[18] A. Micarelli and F. Sciarrone, "Anatomy and empirical evaluation of an adaptive web-based information filtering system," User Modeling and User-Adapted Interaction, Vol. 14, No. 2-3, pp 159–200, Jun. 2004, doi:10.1023/B:USER.0000028981.43614.94.

[19] G. Gentili, A. Micarelli, and F. Sciarrone, "Infoweb: An adaptive information filtering system for the cultural heritage domain," Applied Artificial Intelligence, Vol. 17, No. 8-9, pp. 715–744, Sep. 2003, doi:10.1080/713827256.

[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, Jan. 2013.

[21] E. Stai, S. Kafetzoglou, E. E. Tsiropoulou, and S. Papavassiliou, "A holistic approach for personalization, relevance feedback & recommendation in enriched multimedia content," Multimedia Tools and Applications, Vol. 77, No. 1, pp 283-326, Jan 2018, doi:10.1007/s11042-016-4209-1.

[22] J. MacQueen, "Some methods for classification and analysis of multivariate observations," 5th Berkeley symposium on mathematical statistics and probability, Jun. 1967, Vol. 1, No. 14, pp. 281–297.

[23] J. D. Burger, J. Henderson, G. Kim, and G. Zarrella. "Discriminating gender on Twitter," Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Jul. 2011, pp. 1301–1309, ISBN: 978-1-937284-11-4.

[24] A. Culotta, N. R. Kumar, and J. Cutler, "Predicting the Demographics of Twitter Users from Website Traffic Data," AAAI, Jan. 2015, pp. 72–78.

[25] Q. Fang, J. Sang, C. Xu, and M. S. Hossain, "Relational user attribute inference in social media," IEEE Transactions on Multimedia, Vol. 17, No. 7, pp. 1031–1044, Jul. 2015, doi:10.1109/TMM.2015.2430819.

[26] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," IEEE transactions on pattern analysis and machine intelligence, Vol. 32, No. 11, pp. 1955–1976, Nov. 2010, doi:10.1109/TPAMI.2010.36.

[27] O. Giannakopoulos, N. Kalatzis, I. Roussaki, and S. Papavassiliou, "Gender Recognition Based on Social Networks for Multimedia Production," 13th IEEE Image, Video, and Multidimensional Signal Processing Workshop (IVMSP 2018), IEEE Press, Jun. 2018, pp. 1-5, doi:10.1109/IVMSPW.2018.8448788

[28] N. Kalatzis, I. Roussaki, C. Matsoukas, M. Paraskevopoulos, S. Papavassiliou, and S. Tonoli, “Social Media and Google Trends in Support of Audience Analytics: Methodology and Architecture,” 7th International Conference on Data Analytics (DATA ANALYTICS 2018), Nov. 2018.

[29] Google trends engine, https://trends.google.com/ [*Retrieved May 2019*]

[30] Wikipedia pretrained glove models, http://nlp.stanford.edu/data/glove.6B.zip [*Retrieved May 2019*]

[31] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532-1543, doi:10.3115/v1/D14-1162.

[32] Question-words test, https://storage.googleapis.com/google-code-archive-source/v2/code.google.com/word2vec/source-archive.zip [*Retrieved May 2019*]

[33] Deliverable D4.3, the PRODUCER project, http://www.producer-project.eu/wp-content/uploads/2018/07/D4.3-Evaluation-Benchmarking.pdf [*Retrieved May 2019*]

[34] G, Shani and A. Gunawardana, "Evaluating recommendation systems," Recommender systems handbook, pp. 257-297. Springer, Boston, MA, 2011.

[35] C. L. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon, "Novelty and diversity in information retrieval evaluation," Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, ACM, Jul. 2008, pp. 659-666, doi:10.1145/1390334.1390446.

[36] ITD questionnaire, https://docs.google.com/forms/d/e/1FAIpQLSfkjQPiQbyOxI2iCj2wTzmT8V2Ilee-s_eLyg8h3n_696vWBg/viewform [*Retrieved May 2019*]

[37] SRPT questionnaire, https://docs.google.com/forms/d/1ihDQ5kM5joDHa848JNug8gORpECSGhtRrsFTjLxelus. [*Retrieved May 2019*]