MOVING

Deliverable 1.4: Final implementation of user studies and evaluation

Aitor Apaolaza, Ann Gledson, Angela Fessl, Irina Bienia, Alexandros Pournaras, Till Blume, Ahmed Saleh, Ilija Simic, Annalouise Maas, Franziska Günther, Peter Mutschke, Robert Lorenz, Chrysa Collyda, Markel Vigo

30/03/2019

Work Package 1: Requirements and use case development

TrainiG towards a society of data-saVvy inforMation prOfessionals to enable open leadership INnovation

Horizon 2020 - INSO-4-2015

Research and Innovation Programme

Grant Agreement Number 693092

Dissemination level	PU
Contractual date of delivery	31/03/2019
Actual date of delivery	30/03/2019
Deliverable number	1.4
Deliverable name	Final implementation of user studies and evaluation
File	`MOVING_d1.4_v1.0`
Nature	Report
Status & version	Final & v1.0
Number of pages	86
WP contributing to the deliverable	1
Task responsible	UMAN
Other contributors	KC, EY
Author(s)	Aitor Apaolaza, Ann Gledson, Angela Fessl, Irina Bienia, Alexandros Pournaras, Till Blume, Ahmed Saleh, Ilija Simic, Annalouise Maas, Franziska Günther, Peter Mutschke, Robert Lorenz, Chrysa Collyda, Markel Vigo
Quality Assessors	Iacopo Vagliano, ZBW
EC Project Officer	Hinano Spreafico
Keywords	User studies, final evaluation, requirements

Executive Summary

This deliverable "D1.4: Final implementation of user studies and evaluation" updates the requirement implementation presented in D1.2 (Barthold et al., 2017) and presents the results from the final user studies started in D1.3 (Apaolaza et al., 2018).

Section 2 details how the different MOVING modules addressed the identified requirements. Section 3 details how the MOVING interface has been annotated so the gathered interaction data could be used to carry out remote studies and to personalise the interface. Section 4 explains the methodology employed to analyse the data that produced the results presented in Section 5.

Table of contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Explanation
API	Application Programming Interface
ATS	Adaptive Training Support
DIS	Data Integration Service
FDC	Focused web-Domain Crawler
HTML	Hypertext Markup Language
LDAP	Lightweight Directory Access Protocol
LHTS	Learning-how-to-search widget
MOOC	Massive Online Open Course
NLP	Natural Language Processing
NER&L	Named entity recognition and linking
SSM	Social Stream Manager
SSOAR	Social Science Open Access repository
TAM	Technology Acceptance Model
TF-IDF	Term Frequency–Inverse Document Frequency
UI	User Interface

1 Introduction

1.1 History of the document

Table 1: Document history.

Date	Version
03/02/2019	v0.1: first ToC draft
17/02/2019	v0.2: ToC ready for QA
03/03/2019	v0.3: ToC comments addressed
10/03/2019	v0.4: content ready for QA
17/03/2019	v0.5: document back from QA
24/03/2019	v0.6: document ready for final QA
31/03/2019	v1.0: final document

1.2 Purpose of the document

This document presents the results of the remote user studies carried out for several aspects of the MOVING platform: the working environment, the use of the learning how to search widget, and the MOOCs. The methodology that made these studies possible are presented, detailing how the interaction data was retrieved, and the employed analysis pipelines. In addition to these results, this document also presents the latest updates to the list of requirements, providing details about the latest implementation stage of each of the MOVING components.

1.3 Structure of the document

Section 2 presents the final implementation status of the major functionalities of the MOVING platform. Section 3 details the methodology followed to annotate the MOVING user interface, explaining the criteria for the annotation, as well as how other MOVING modules could make use of it and specifying the nature of the interaction data they needed. Section 4 details the analysis pipelines employed to carry out remote analyses of MOVING visitors', and Section 5 presents the results of the application of these analyses.

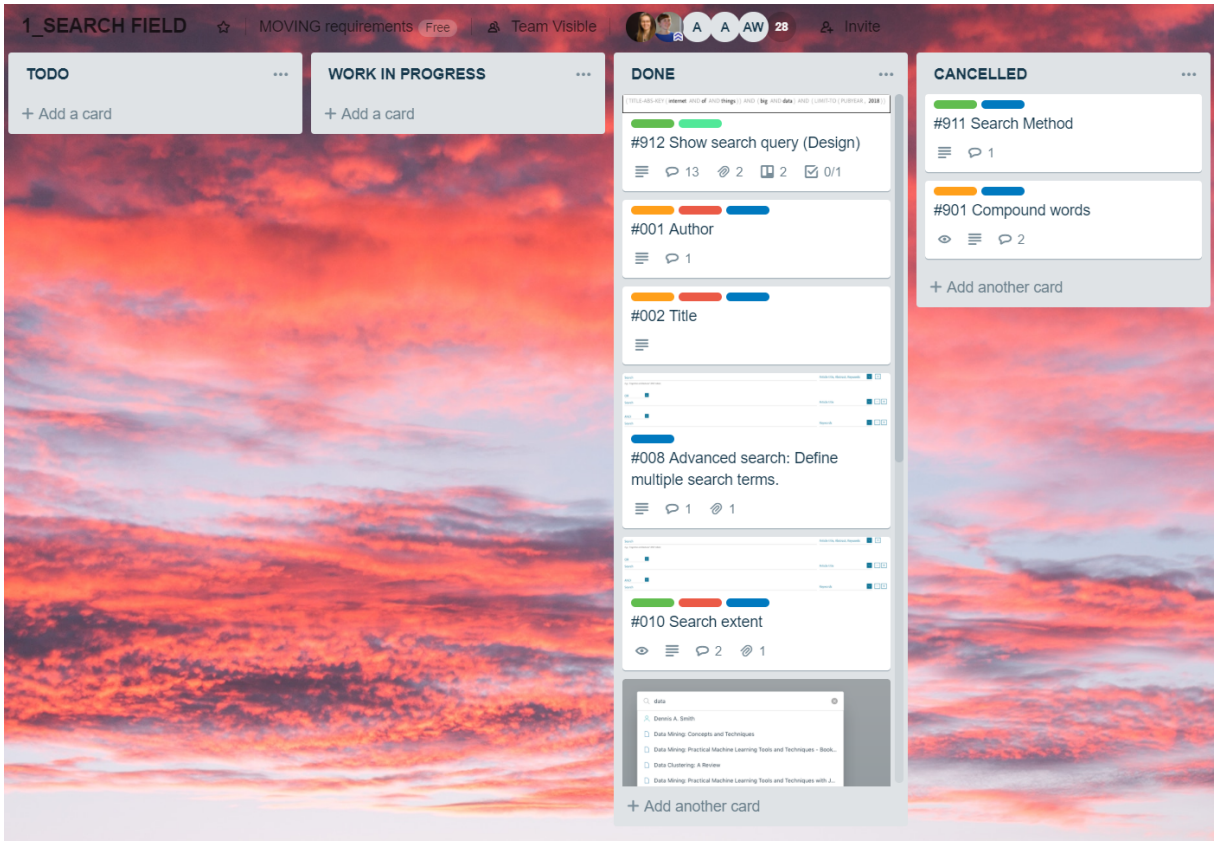
2 Final implementation status

In this section of the deliverable, we present an update of the implementation status as described in deliverable D1.2 (Barthold et al., 2017). The working process of the requirements implementation is presented in subsection 2.1 followed by subsection 2.2 on the requirements identified as a result of the user studies performed for D1.3 (Apaolaza et al., 2018). Additionally, subsection 2.3 presents an update on the status of implementation for responsiveness of the platform, the Adaptive Training Support (ATS), the community, crawling and processing videos, data sources, search functionalities and document analysis including disambiguation and deduplication, visualisations, user management and access, as well as the recommender system and entity extraction for both use cases supplemented with the visualisation of the entity extraction results.

2.1 Working process of requirements implementation

The working process of the requirements' implementation has not changed since it was described in detail in deliverable D1.2 (Barthold et al., 2017). We continued to use the project management software Trello in order to instrument a collaborative process. As an example, Figure 1 shows the Search Field board with all final requirements implemented on the platform. The initial requirements on Trello have also been listed with their implementation status in deliverable D1.3 (Apaolaza et al., 2018). In addition, new requirements have arisen since deliverable D1.3 (Apaolaza et al., 2018) due to user studies performed on the MOVING platform (see subsection 2.2 of the current deliverable). Being the result of direct user studies and impacting both use cases, these requirements have received high priority and focus from the technical partners. The high priority requirements have been integrated onto the platform or are in the final stage of integration. As Trello was used for implementation tracking since the initial requirements were outlined in deliverable D1.1 (Bienia et al.,

Figure 1: Search Field Trello Board



2017), some initial requirements have been cancelled or moved to the backlog due to changes in prioritisation or updated requirements. For each of these requirements, a detailed explanation has been added to Trello and was reported in deliverable D1.3 (Apaolaza et al., 2018).

2.2 Requirements identified in D1.3

As a result of the studies carried out in D1.3 (Apaolaza et al., 2018), a set of requirements to improve the usability of the platform were put together, categorised as presentation, features, and data (see Section B.12). These recommendations were included as additional requirements on the Trello platform, as described above, and were successfully implemented into the MOVING platform.

2.3 Status of implementation and outlook

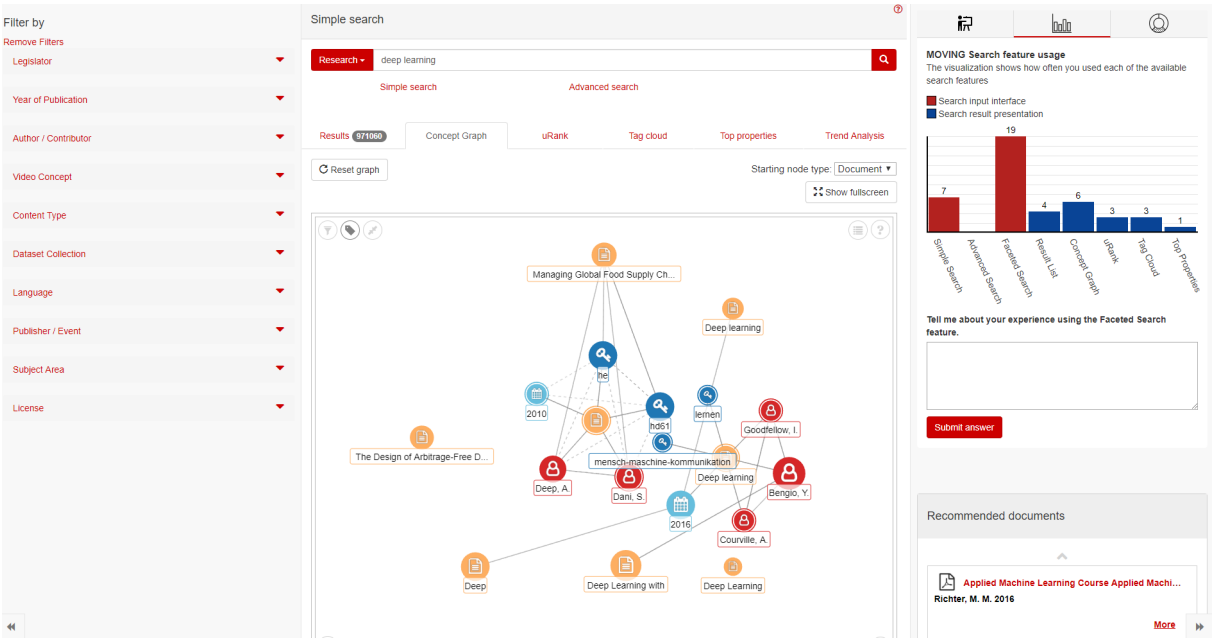
This subsection presents the latest status of implementation of the major functionalities consisting of initial and new requirements. In addition to the functionalities described in D1.2 (Barthold et al., 2017), the implementation status of several new functionalities has been here described.

2.3.1 Responsiveness

The updated technical aspects of the MOVING platform's responsive design are discussed in more detail in deliverable D4.3 (Lorenz et al., 2017) and are summarized here in short. As originally described in the previous WP4 deliverables (D4.1 (Gottfried, Grunewald, et al., 2017) and D4.2 (Gottfried, Pournaras, et al., 2017)) the Bootstrap framework[1] is used to implement the responsive design making. The bootstrap framework allows accessibility to the platform via multiple types of devices, ranging from desktop web clients to smartphones and tablets. The mock-ups of the MOVING platform, developed with the Balsamiq Software [2] and described in section 7 of deliverable D1.1 (Bienia et al., 2017), were used as a starting point for the implementation of the first version of the responsive design. Consequently, and by continuing the use of the Bootstrap framework,

[1] `http://getbootstrap.com`, last accessed at 22.09.2017.
[2] `https://balsamiq.com`, last accessed at 22.09.2017.

Figure 2: MOVING platform on a PC screen



we implemented the responsive design in all screens of the MOVING platform by using the implemented environment of the platform. More specifically, search, communities, learning, contacts, MOOC and My page screens are responsive.

Furthermore, the widgets in the Adaptive Training Support (ATS), which is located in the right sidebar of the MOVING Platform, always use all of the available width to display their content. Since the MOVING Platform itself is responsive, when resizing the browser window, the widgets of the ATS will adapt as well. The visualizations of the MOVING Platform act in a similar way. Given that they are integrated in the central part of the search result page, they also use the available width of their parent container. Therefore, when resizing the page, the container will resize too and the visualizations will adapt to it. An unresolved issue is still present in the Concept Graph, which has compatibility issues with browsers other than Google Chrome (Desktop & Mobile), due to recent updates of their viewing engines.

All views adapt to different screen sizes, where the layout is automatically changed according to the size and capabilities of the device (this also includes the visualization search results and the ATS and Curriculum widgets). For example, on a PC screen, the users see the content in a three-column view, as depicted in Figure 2, on a mobile phone, the content is presented in a single-column view, and on a tablet the same content is displayed with the menus on the top of the screen, as shown in Figure 3 (left) and (right).

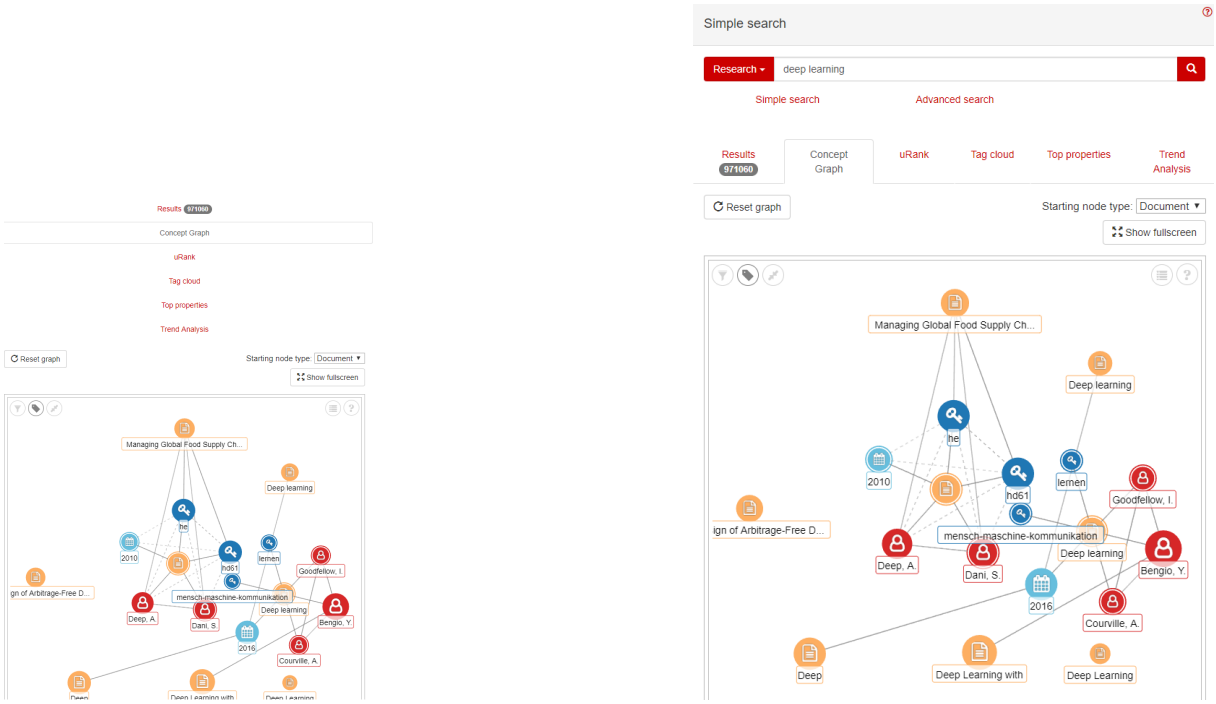
2.3.2 ATS

The ATS pursues the overarching learning goal "acquire digital information literacy skills and competencies". This goal can be divided into the following two sub-goals: the first goal is to support the training on how to use the MOVING platform; the second goal is to educate users to become information-savvy professionals by **providing automatic learning guidance to raise the learner's competence level for each competence in the curriculum to the expert level**. To be able to achieve both goals, the ATS consists of two different widgets integrated into the MOVING platform's user interface:

- "Learning-how-to-search" widget
- "Curriculum Reflection" widget

During the third year of the MOVING project, we finalized the development of the "**Learning-how-to-search**" widget by focusing on its integration and combination with the "Curriculum Reflection" widget, by adding responsive design and improving the visual appeal and stability of the widget. The basic idea of the widget is to mirror back the user's own search behaviour based on the user's activities with respect to the feature usage on the MOVING platform. During the third year, we updated the presented features accordingly, now consisting of three functionalities to initiate a search, namely "Simple Search", "Advanced

Figure 3: Example of the MOVING responsive design on a smartphone (left) and tablet (right)



Search", "Faceted Search", and five functionalities of presenting the results, namely "Result List", "Concept Graph", "uRank", "Tag Cloud" and "Top Properties". Besides visualising the feature usage per user, we also implemented a reflection guidance concept to show reflective prompts adapted to the user's needs. Figure 4 presents the final version of the "Learning-how-to-search" widget in the MOVING platform. Additionally, we conducted two different user studies to evaluate the widget. The results of the first "Learning-how-to-search" widget evaluation can be found in this deliverable in Section 5.3. The results of the second evaluation can be found in D2.3 (Gunther et al., 2019).

In the third year, we further developed and improved the concept for the "**Curriculum Reflection**" widget and implemented it in the MOVING platform. The Curriculum Reflection widget (Figure 5) consists of two parts, the *curriculum learning and reflection part* and the overall progress part. The curriculum learning and reflection part, is divided into two areas. The upper area contains either a learning prompt, suggesting to learn more about the next topic that would be the next in the current sub-module of their curriculum, and a button which opens the respective learning unit in a new tab, or it presents a reflective question that motivates the user to think about the currently learned topic. Below the prompts, a progress indicator shows the user the progress of the curriculum's current sub-module. The progress is defined by the amount of completed learning units in comparison to the available ones for this particular sub-module, and matches with the progress in the overall progress widget for this specific sub-module. Detailed information about both widgets can be found in D2.2 (Günther et al., 2019).

2.3.3 Community

Users of the platform can create communities (former projects) on the MOVING platform, in which they can collaborate with other users as described in more detail in D4.3 (Lorenz et al., 2017). For each community, the user can choose from a set of modules to enable specific functionalities, like wikis, forums and a news board (Figure 6).

Additionally, users can maintain a profile page with their research fields and competencies that are accessible to other users via the contact search (Figure 7).

Each user is assigned an individual profile page where a set of individual skills, research areas, and collaboration interests, information about institution affiliation, an email address, and the users ORCID[3] can be specified. Activities such as community membership or forum posts are also displayed on the dashboard (Figure 8).

[3]ORCID is a permanent digital identifier for researchers aimed at increasing the recognition of their work through the presentation of their professional activities (funding, publications or memberships) (`https://orcid.org/` last accessed on 06-03-19).

Figure 4: Final version of the "Learning-how-to-search" Widget in the MOVING platform

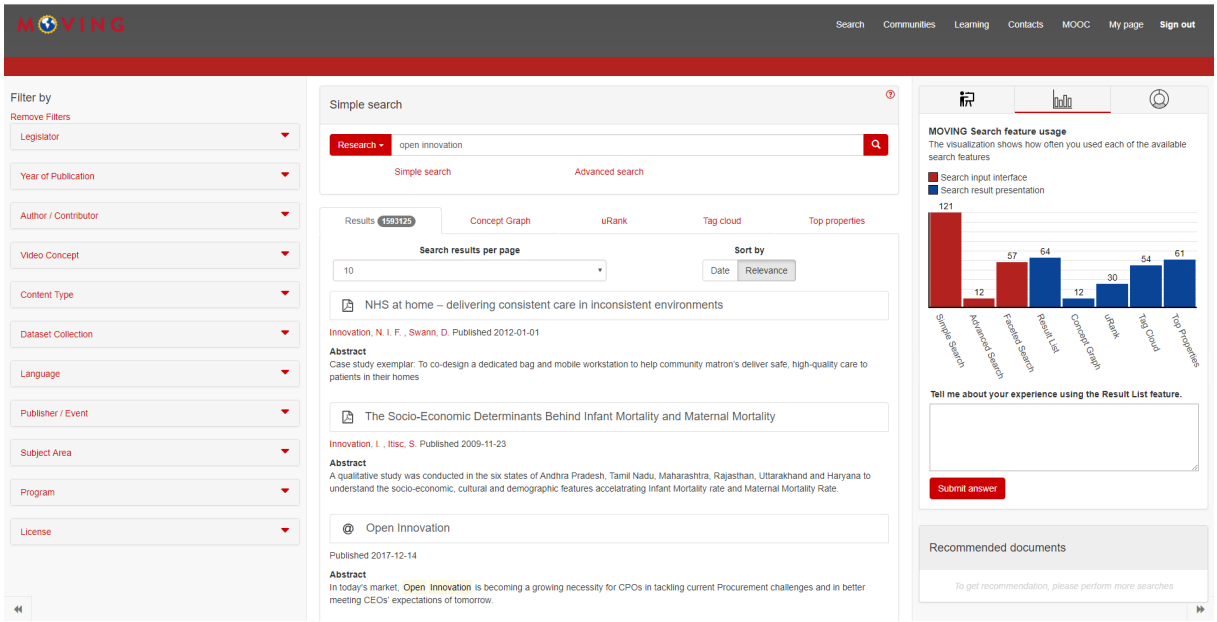
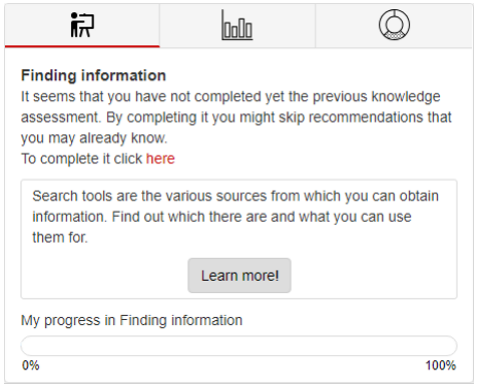
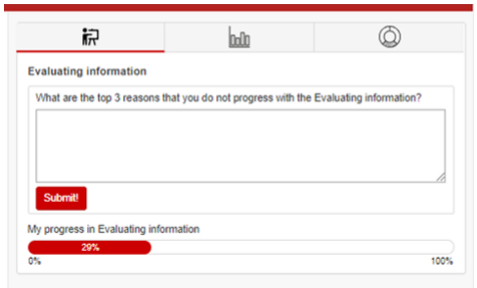


Figure 5: Final version of the "Curriculum Reflection" widget





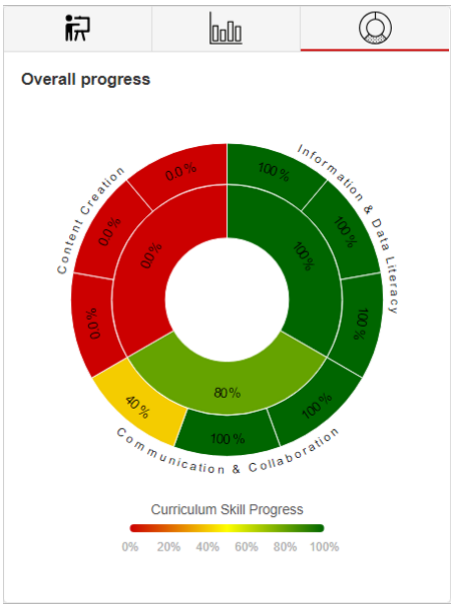


Figure 6: Example of the community structure with the navigation on the left (Overview, News, Wiki, Forums) and the main field in the middle (Summary, Description)

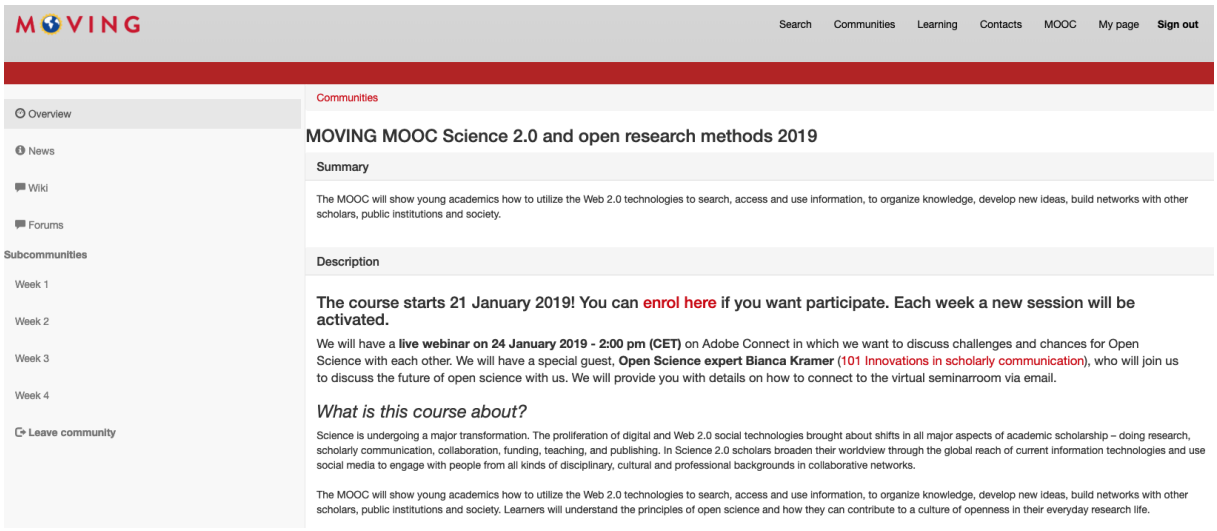


Figure 7: Contact search

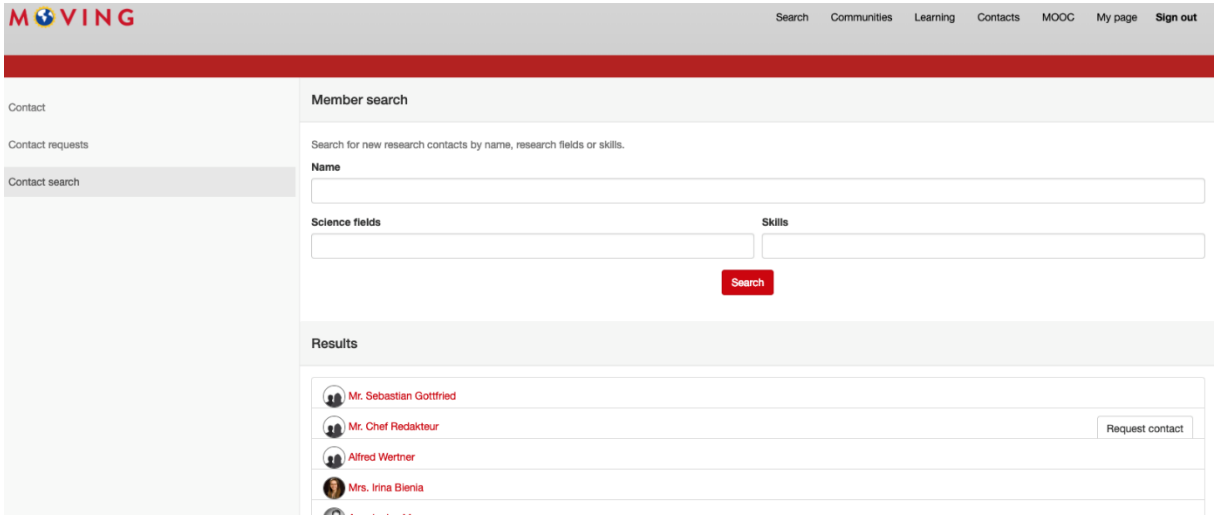
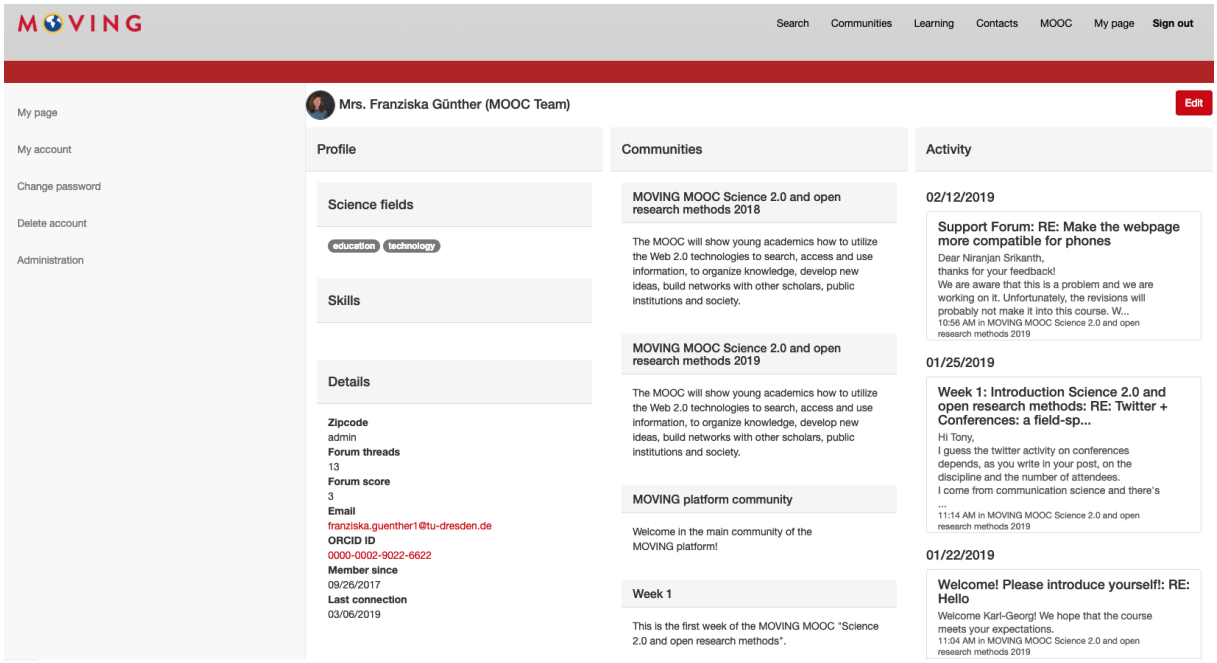


Figure 8: User Profile view



2.3.4 Crawling and video processing

There are 3 different crawlers running in the background in the MOVING platform. The Focused web-Domain Crawler (FDC) crawls specific websites inserted by the platform's administrator. The Social Stream Manager (SSM) crawls social media and the Search-Engine based web Crawler exploits the Google API to search the web, both for specific topics inserted by the administrator. All crawlers have built-in mechanisms to collect video URLs, which are later retrieved and analysed by the video processing external module. The crawlers are described in detail in D3.1 (Blume et al., 2018), D3.2 (Vagliano et al., 2018), and D3.3 (Vagliano et al., 2019).

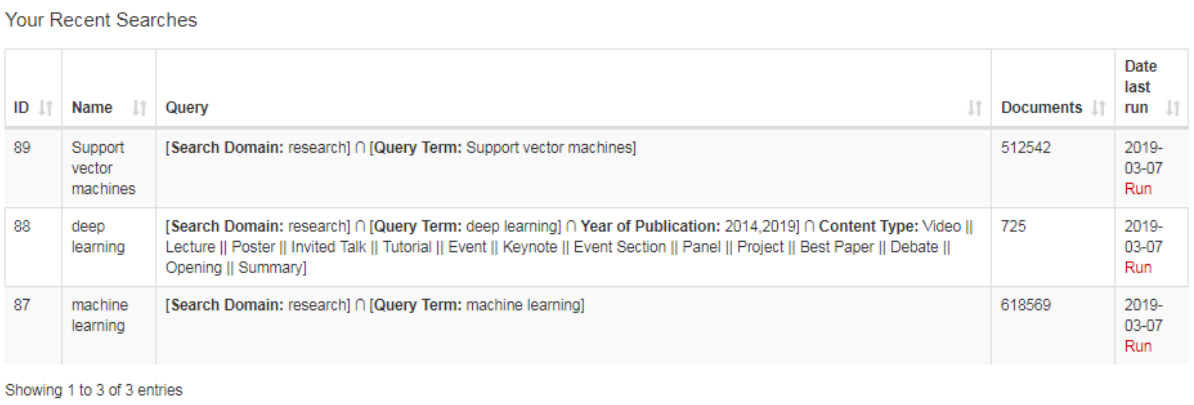
Video processing in MOVING consists of 2 different technologies for analysing videos. The first one performs visual analysis on the videos retrieved by the crawlers and the second one performs transcript analysis on the lecture videos. In both cases the analysis is temporal fragmentation and concept detection. Both video processing techniques are accessible via their REST services. For more technical details on the implementation and the underlying technologies please refer to D3.1 (Blume et al., 2018), D3.2 (Vagliano et al., 2018), and D3.3 (Vagliano et al., 2019).

2.3.5 Data sources

The MOVING platform provides access to a large variety of documents coming from different data sources. We distinguish between the document type (book, article, video, etc.) and the document source. For example, crawled websites or professionally created metadata from the ZBW economics corpus. All types of documents are stored in the search engine based on a common data model and are subject to quality tests before being indexed, as described in details in deliverables D3.3 (Vagliano et al., 2019) and D4.3 (Lorenz et al., 2017). The MOVING platform currently provides access to: (i) economics literature (ZBWEconomicsDataset[4]) ; (ii) social sciences literature (OAFulltexts, PublicationMetaData); (iii) laws and regulations from Wolters Kluwer; (iv) cross-domain literature crawled from the Linked Open Data cloud[5] (BTC2014); (v) various crawled websites specified by the use case partners containing learning material and funding opportunities; information about organisations; (vi) videos from VideoLectures.NET; (vii) information about projects (ProjectMetaData); as well as open access documents in various domain from the CORE repository[6] (CORE DB). Further information on the datasets included in the MOVING platform is available in deliverable D6.2 (Collyda et al., 2017), while a more quantitative summary of the data integrate is provided in deliverable D4.3 (Lorenz et al., 2017). About

[4]The name in parenthesis indicates the name of the corresponding dataset in D6.2 (Collyda et al., 2017)
[5]`https://lod-cloud.net/`
[6]`https://core.ac.uk/`

Figure 9: Search history


Your Recent Searches

ID	Name	Query	Documents	Date last run
89	Support vector machines	[**Search Domain:** research] ∩ [**Query Term:** Support vector machines]	512542	2019-03-07 Run
88	deep learning	[**Search Domain:** research] ∩ [**Query Term:** deep learning] ∩ **Year of Publication:** 2014,2019] ∩ **Content Type:** Video || Lecture || Poster || Invited Talk || Tutorial || Event || Keynote || Event Section || Panel || Project || Best Paper || Debate || Opening || Summary]	725	2019-03-07 Run
87	machine learning	[**Search Domain:** research] ∩ [**Query Term:** machine learning]	618569	2019-03-07 Run

Showing 1 to 3 of 3 entries

99% of the documents contained in the above-mentioned datasets passed our quality tests and can be accessed via the MOVING platform.

2.3.6 Search Functionalities

The search engine of the MOVING platform contains dozens of millions of documents. In addition to the search features presented in the previous deliverable D1.2, we have integrated more filters and functionalities in our MOVING's search page to increase its usability. In the following, we summarize the new functionalities:

- **Search history**. As shown in Figure 9, in the search page, the users of our MOVING platform can view a history of their recent searches, including all the filters they applied. In this way, they can easily repeat their searches. This is useful since search is an iterative process and it is rather common for users to search more times for the same terms in a given period, e.g. when approaching a new topic to learn or investigate. In order to activate this feature, the users should activate the option "Collect user interaction data about me for the Adaptive Training Support" from their "Privacy Settings" menu. In fact, the search history relies on the user logging module.

- **Exact match or phrase search**. Adding quotes around the search terms ensures that the retrieved search results will have this exact sequence of search terms.

- **Boolean operators**. Our MOVING platform allows the users to combine keywords with operators such as AND, NOT and OR to produce more relevant results. For example, the user can enter the search query "machine OR learning" which is equivalent to "machine || learning"; in this case the retrieved search results will match the term "machine" or the term "learning" or both. Furthermore, the user can exclude certain terms from the search results by adding a dash symbol ("-") or adding the word "NOT" in front of the terms that should not appear in the retrieved results. For example, the search query "machine –learning" will only retrieve the results that contains the term "machine" and does not contain the term "learning".

- **Prefix query**. This is a query in which the user only specifies a few characters from a term followed by the symbol "*". In this case the search engine will search for all the terms that start with these characters. For examples, the user can enter the search query "machin*", the search engine would then find all the terms that starts with the characters "machin", like "machine", and perform a search with these terms.

- **Trend analysis**. One desirable feature, notably for auditors, is analysing the content of the MOVING platform to detect what is (or was) increasing in popularity. The trend analysis module extracts the frequency of topics, such as data mining, and entities, like names, locations, organizations, etc. (see subsection 2.3.12), in the documents over time. This frequency over time corresponds to a time series, which is given to the trend detection model to determine whether the time series contains a trend, i.e. whether the given topic or entity is more popular in a given period of time than others. The model is described in deliverable D3.3 (Vagliano et al., 2019). A screenshot of the current trend analysis feature is depicted in Figure 10.

Figure 10: Trend detection module

2.3.7 Author name disambiguation

Author name disambiguation addresses the problem of mapping a single author name on a document, which is referred to just by a string of characters and not by an identifier, to the correct real-world author. For this problem, GESIS has provided a novel method which decides for a set of author mentions with the same name which of them belong to the same author and which do not. This is done by applying agglomerative clustering based on document features extracted from the document collection for the mention being studied (such as affiliation, co-authors, referenced authors, email addresses, keywords, publication years). The approach has been described in detail in Deliverable D3.1 (Blume et al., 2018), Section 3.7. The method has been fully integrated into the MOVING platform (for details see Deliverable D4.2 (Gottfried, Pournaras, et al., 2017), Section 7.5). As a result of the disambiguation process, each disambiguated author is assigned a unique internal authorID such that documents having the same author name but belonging to different real-world authors can be distinguished. In order to make the disambiguation results available to the user, a feature has been implemented into the MOVING platform interaction layer that allows the user to click on the name of an author given for a selected document. This click event then triggers a query for the author identifier assigned to the respective mention such that only documents authored by the author having the author identifier in question appear on the result page.

2.3.8 Document Deduplication

Document Duplication appears when two (or more) metadata entries refer to the same real-world document, but might describe it in a slightly different way. This problem becomes particularly relevant when documents originate from different sources, as in the MOVING case. GESIS has developed a method for identifying duplicates in a document collection and moreover have developed a method to keep the computational complexity low. The method has been described in detail in the Deliverable D3.2 (Vagliano et al., 2018), Section 3.4. The duplicate detection is implemented on the MOVING platform. The results of the duplicate detection process are inserted into the search index of the MOVING platform. On this basis, all documents within a cluster of duplicates are ranked according to a) the relevance of the source, b) the relevance of the document type, and c) the timestamp of the document. All duplicates except for the best ranked documents are finally removed from the index. This procedure is described more detailed in Deliverable D3.3 (Vagliano et al., 2019) (Section 2.1).

2.3.9 Visualisations

The following visualisations have been implemented in the final version of the MOVING Platform:

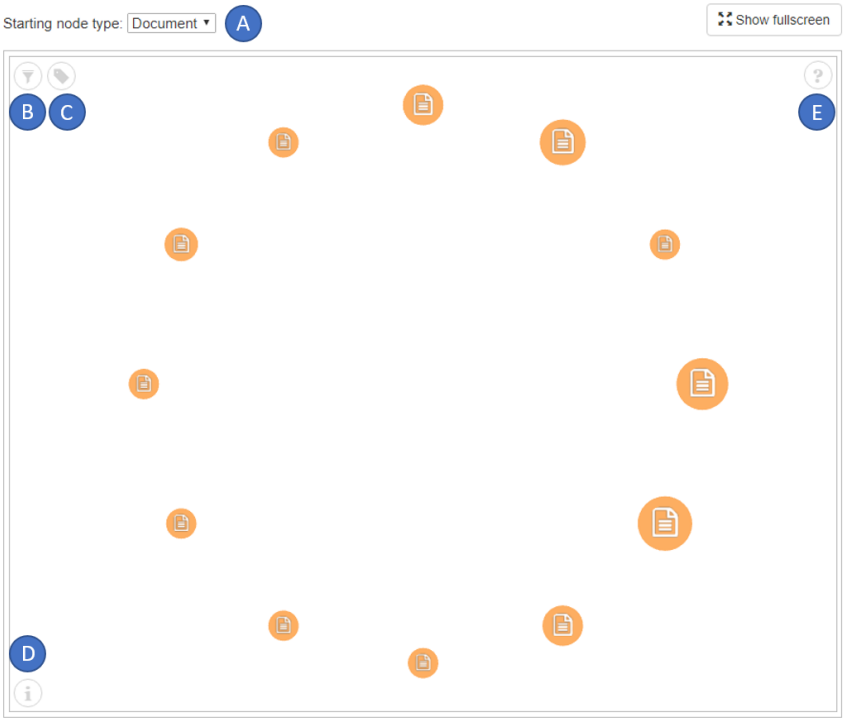
- **Concept Graph** - for the discovery and exploration of relationships between documents and their properties.
- **uRank** - a tool for the interest-driven exploration of search results.
- **Top Properties** - a bar chart displaying aggregated information about the properties of the retrieved documents.
- **Tag Cloud** - a visualisation for the analysis of keyword frequency in the retrieved documents.

All of the visualisations in their current version are explained in greater detail in D3.3 (Vagliano et al., 2019) Section 5. Only the Concept Graph had some additional improvements which are mentioned here and will be covered in detail in D4.3 (Lorenz et al., 2017).

2.3.9.1 Concept Graph

In the third year the Concept Graph has been further improved by adding various new functionalities and improving upon existing ones. Figure 11 shows the version of the Concept Graph as reported in D3.3 (Vagliano et al., 2019), Section 5.1. As a starting point, it is now possible to select the node type for the initial layout of the graph (A). Additionally, functionality to reduce the number of displayed nodes by means of filtering was added (B). The nodes can be filtered by node type, label, edge count, and for the nodes containing dates, also by year. The labels of all nodes can be simultaneously turned on and off with the help of a label toggle (C). Furthermore, additional statistical information about the number of nodes, edges, density and average degree can be inspected by clicking on the information button (D). The help text button was also moved to the upper right corner.

Figure 11: New user interface elements in the Concept Graph: A) starting node type drop-down list, B) node filter menu button, C) node labels toggle button, D) graph properties button, E) help text button



Furthermore, the following completed new functionalities will be reported in more detail in D4.3 (Lorenz et al., 2017): visualisation of extracted entities, co-occurrence analysis between entities, co-author edges, relevance edges between entity and document, and the functionality to aggregate nodes. Moreover, existing functionalities of the Concept Graph have been further improved. In the latest version of the Concept Graph, when using the ring-menu, it is possible to limit the node types which should be expanded. Additional minor improvements include: addition of a completion ring to entirely opened nodes, the existing help-text has been rewritten, the label positioning and displaying has been improved, the node interactions simplified, and an additional button for removing nodes has been added.

Figure 12 shows the improved ring-menu on the left, in which the different sections of each ring represent a different node type. An example of an aggregated subgraph can be seen on the right side. The outer ring shows the distribution of the aggregated nodes, while inside this ring it can be observed how interconnected the different nodes are with each other.

2.3.9.2 uRank

Figure 13 shows the final version of the visual interface uRank, fully integrated in the MOVING Platform. Based on the searched query, the top 100 retrieved results get displayed as a ranked list. On the right hand side in uRank, the extracted keywords get displayed in a tag cloud. By selecting keywords of interest, the results get re-ranked in such a way that the results containing the selected keyword move to the top. It is possible to select multiple keywords and even fine-tune the importance by using the slider under the selected ones. By clicking on a result, a dialog pops up which displays additional information about the retrieved document. Additionally, the user can also export the current view of uRank with the search configuration by clicking on the "Export visualisation" button.

2.3.9.3 Top properties

Figure 14 shows the current Top Properties bar chart visualisation. On top of the visualisation a drop-down menu can be used to select a property in the retrieved results set, for which the distribution of values should be displayed. It is currently possible to select between the following properties: Sources, Concepts, Keywords, Authors and Publication Year. Clicking on one of the bars reveals a small dialog that lists all the documents from the retrieved result set with this property. The ranking of the documents in this document selection dialog is based on the initial ranking provided by the search engine. Clicking on one of the document names

Figure 12: Left: The improved ring-menu - The added sections allow for more control in which way the graph should be expanded. Right: An example of an aggregated subgraph - the selected nodes are grouped into a single node, where it is possible to see how the different node types are connected to each other

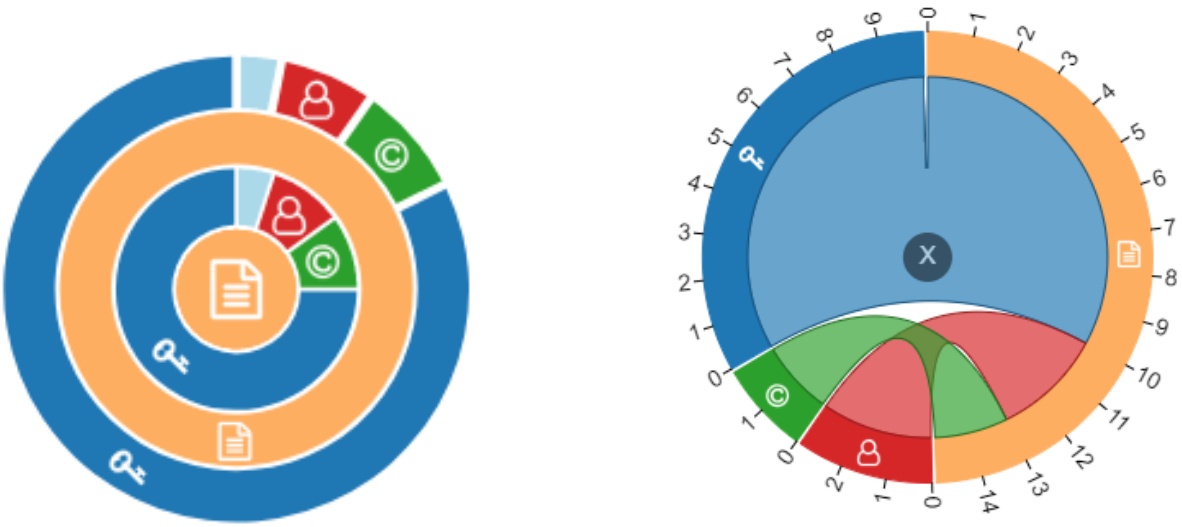


Figure 13: The uRank visual interface: Selecting keywords from the tag cloud on the right side re-ranks the result list, moving the results associated with the selected keywords to the top

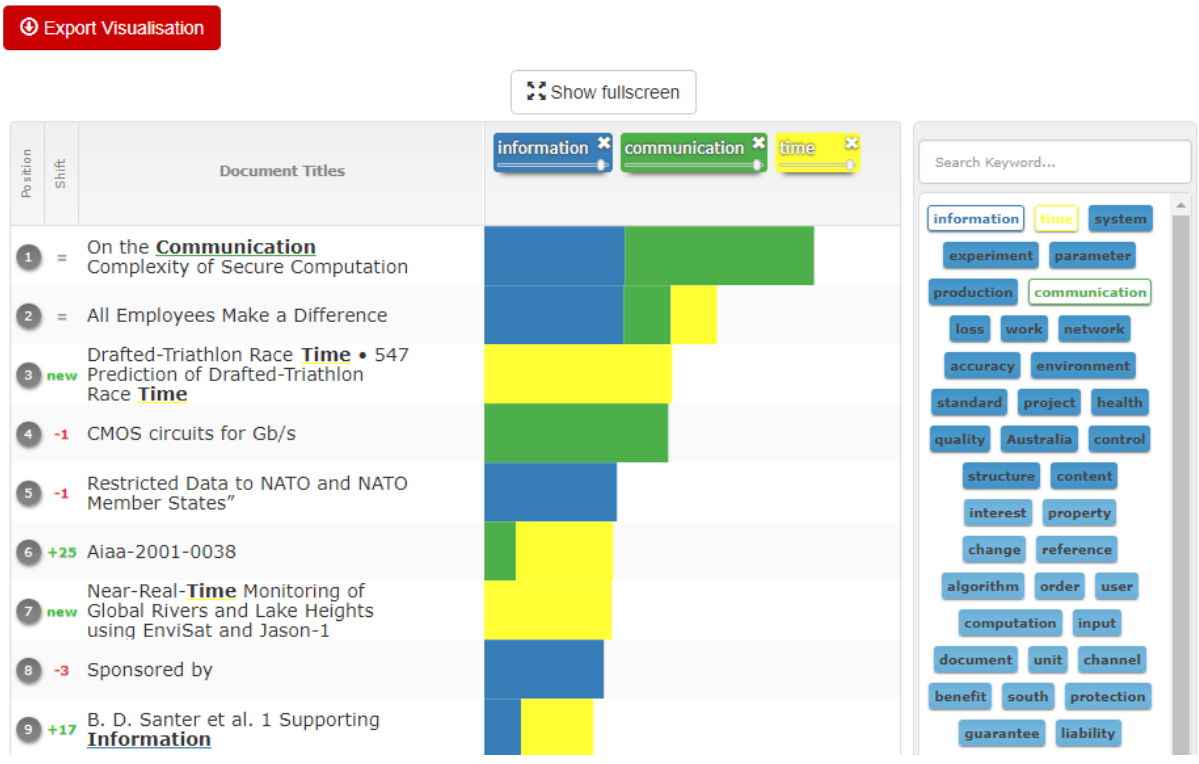
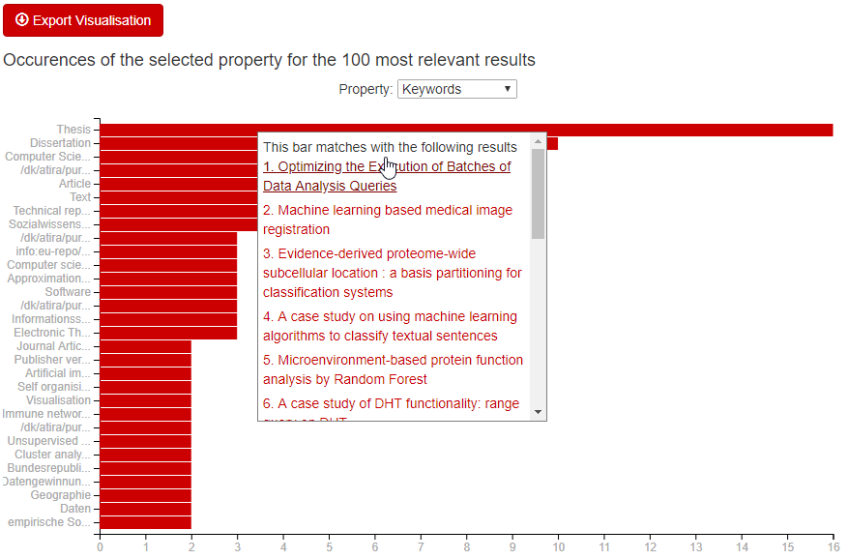


Figure 14: The "Top Properties" visualisation showing the results associated with a property. Clicking on one of the results opens the document in a new tab



opens the associated URL in a new tab. This visualisation also supports an export functionality, which triggers a download of the current state of the visualisation with the applied search configuration.

2.3.9.4 Tag Cloud

Even though uRank already contains a tag cloud, a separate tag cloud visualisation with additional filtering and exploration functionalities has been added to the MOVING Platform. This tag cloud visualisation can be seen in Figure 15. It shows the keywords, which were extracted from the top 100 retrieved results, ordered by frequency. In this tag cloud, it is possible to change the sorting order, search for keywords and to filter them by frequency (upper slider) or year (lower slider). Clicking on a keyword opens a small dialog with a listing of all the retrieved documents containing this particular keyword. Clicking on an item in this list will open the document in a new tab. Similar to the Top Properties visualisation and uRank, this visualisation also can be exported with the current state of the visualisation and the applied search configuration.

2.3.10 User Management and Access

2.3.10.1 User management systems

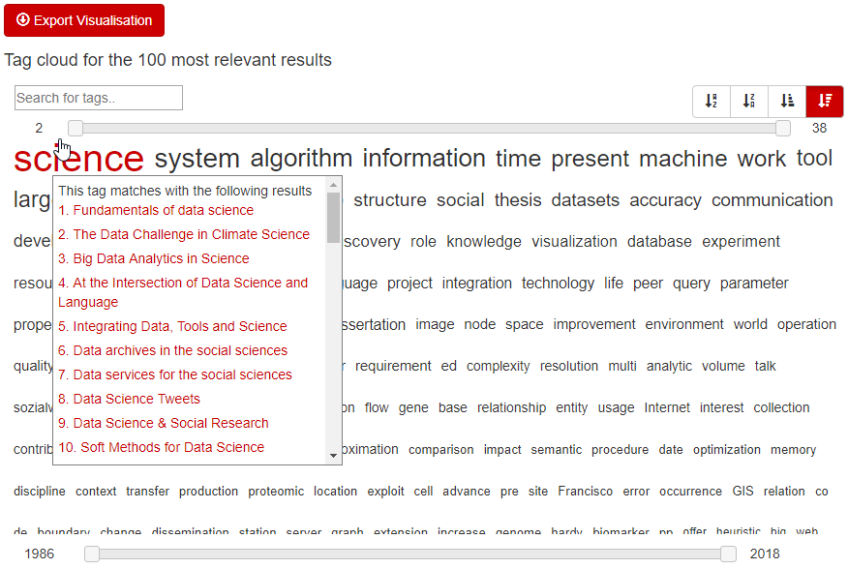
To integrate the platform into existing user management facilities of IT environments in research institutions the authentication via Shibboleth[7] was enabled. Shibboleth is a system for distributed authentication and authorization for web applications. Lightweight Directory Access Protocol (LDAP)[8] is implemented in the platform as another authentication system which is commonly used by private companies. If necessary, the appropriate settings have to be added and configured in the admin menu. In both cases, users are then allowed to authenticate using their institutional identity.

2.3.10.2 User Self-Registration

New users to the MOVING platform are asked to create a user account before they get access to the functionalities of the platform. Figure 16 shows the on-boarding procedure which is described in the following. By logging into the MOVING platform for the first time, one is asked to agree to the terms and conditions as well as the privacy agreement and allow the data collection process. To comply with the European data protection law, the user interaction tracking is turned off by default. Only when the user agrees with the collecting of interaction data, the tracking functionality is enabled. The user can enable or disable the user interaction tracking at any time via the profile settings. A profile page is created for every user by default. However, the user is able to decide whether to be found in the contact search and if the account should be private or visible to other users of the platform.

[7] `https://shibboleth.net/`, last accessed at 06-03-19
[8] `https://ldap.com/`, last accessed at 06-03-19

Figure 15: The "Tag Cloud" visualisation showing the results associated with a keyword. Clicking on one of the results opens the document in a new tab



2.3.11 Semantic profiling and recommender system

The MOVING platform hosts a vast amount of heterogeneous documents, such as publications, video lectures and tutorials, social media posts, etc. To help users to deal with this vast amount of information the recommender system suggests interesting documents. In this way, it enables users to discover documents without explicitly searching for them, since sometimes users do not exactly know what to search, or they may have found interesting documents not directly related to their current search. The MOVING recommender system exploits the HCF-IDF (Nishioka & Scherp, 2016) semantic profiling method to build users' and documents' profiles and then provides recommendations based on the match of document-user profiles. The user profiles are based on the search history, collected through WevQuery (Apaolaza & Vigo, 2017), the module of the platform which logs user-interaction data (see deliverables D3.1 (Blume et al., 2018), D3.2 (Vagliano et al., 2018), and D3.3 (Vagliano et al., 2019)).

The recommender system is integrated into the search page of the platform, through the widget shown in Figure 17. The widget is available before searching for documents (Figure 18), so that, alternatively to searching for content, users can click on one of the suggested documents if found to be interesting. The recommender system is extensively described in deliverables D2.1 (Fessl et al., 2019), D2.2 (Günther et al., 2019) and D2.3 (Gunther et al., 2019), as well as D4.3 (Lorenz et al., 2017).

2.3.12 Named Entity Recognition

2.3.12.1 Use case 1

Named Entity Recognition (NER) is a well-known Natural Language Processing (NLP) task that automatically recognises entities in a text and classifies them into a set of pre-defined classes such as person, organisation, and location. In the MOVING platform, two modules address this task, one focuses on the auditors' use cases and the other on the young researchers one. As motivated in detail in deliverable D3.3 (Vagliano et al., 2019), having this module is beneficial because of the variety of entities to be recognised and the type of data to be considered (e.g. scientific methods and tools in papers versus organizations in web sites). The module which focuses on the auditors' use case exploits the Stanford CoreNLP toolkit (Manning et al., 2014) with the pre-trained German and English language models. Furthermore, we tailored the solution to the auditors' use case by making use of DAX 20 company names[9] in the annotation process. Our NER method is implemented in the Data Integration Service (DIS), described in deliverable D4.3. For each indexed document, where there is the full-text and the language information available, entities are recognised and stored in the corresponding entities field of the document. Each recognised entity is stored with a default identifier, in order to enable disambiguation at a later stage. This approach is critical to the auditors' use case since it allows extracting entities from the full-text of documents on-the-fly. In this context, on-the-fly means without storing the

[9] https://www.bloomberg.com/quote/DAX:IND/members

Figure 16: Privacy settings, MOVING platform on-boarding

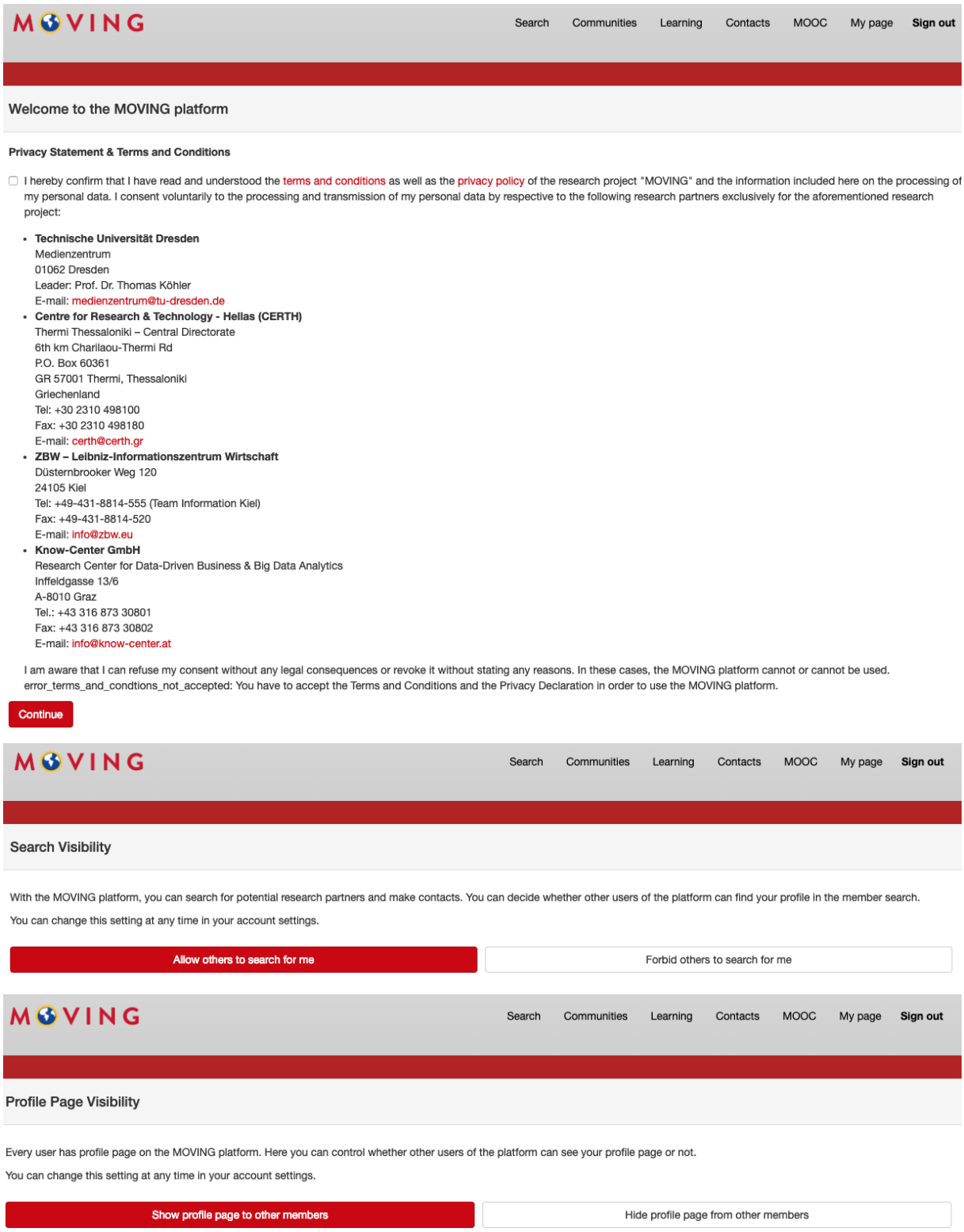


Figure 17: The recommender system widget

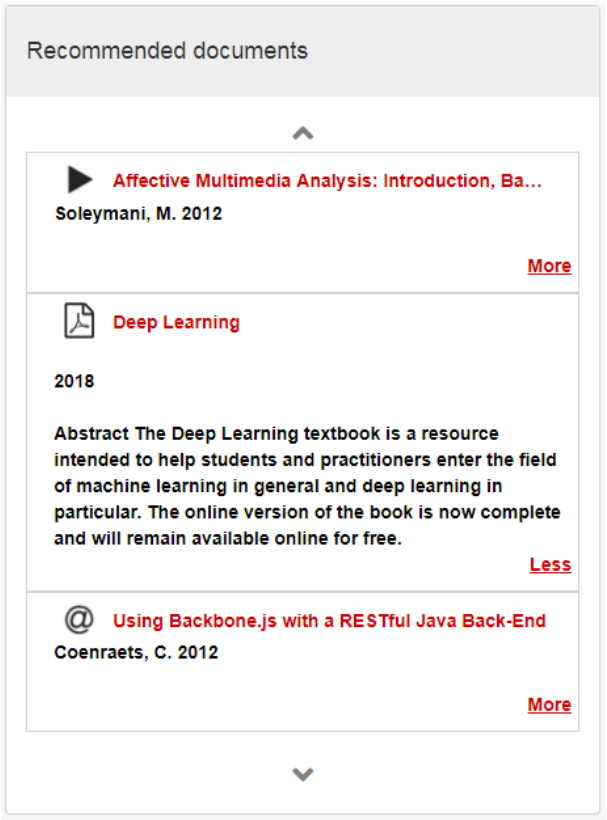


Figure 18: The recommender system widget in the search page

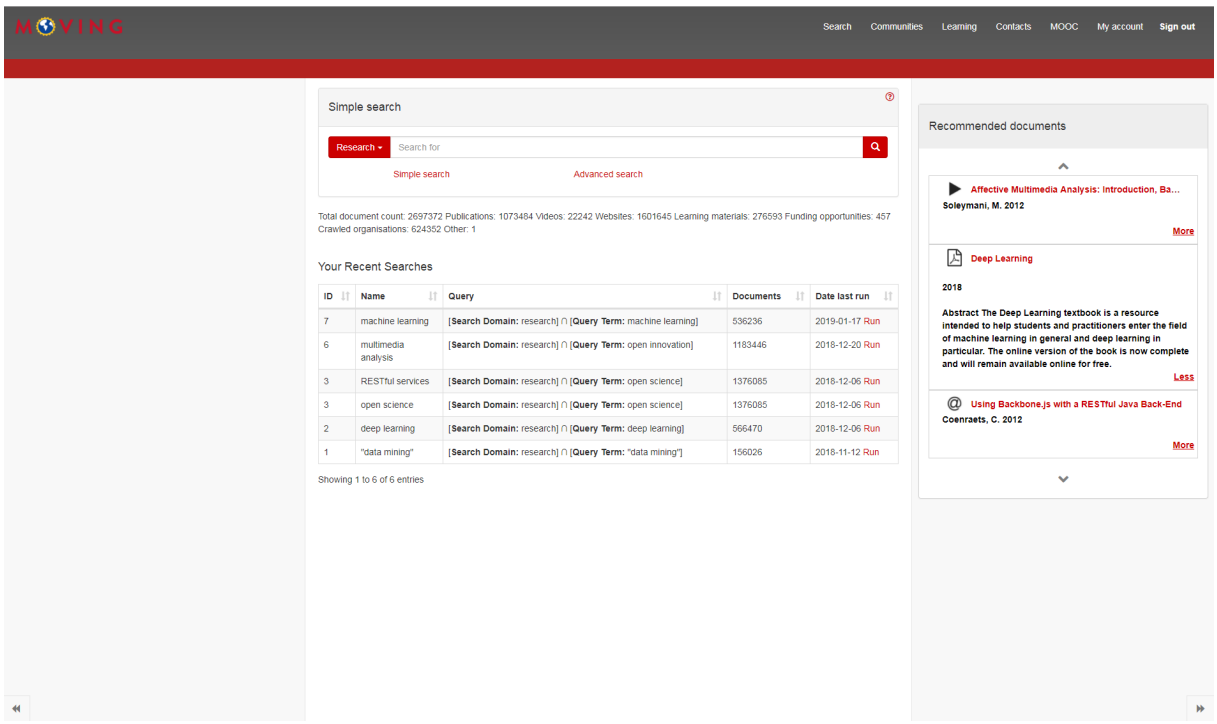
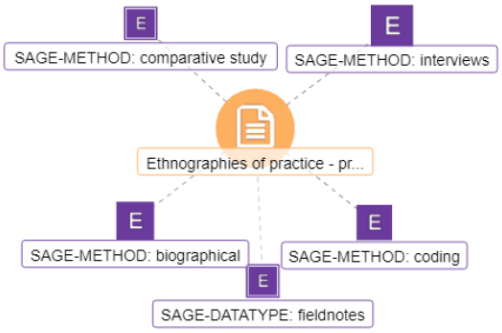


Figure 19: Concept Graph - entity document relevance



document in the index, i.e., without making it publicly available on the platform. Further information on the methods are available in deliverable D3.3 (Vagliano et al., 2019).

2.3.12.2 Use case 2

Named entity recognition and linking (NER&L) aims at semantically annotating unstructured textual sources by additional information extracted from the text, such as mentioned research methods, in order to improve search and retrieval. GESIS applied NER&L to the scientific use case of information retrieval in the Social Sciences. For this, we defined six basic entity types respectively, relevant for the Social Sciences, namely Research Method, Research Theory, Research Tool, Research Measurement, Research Dataset, and Research Field which we extracted from fulltexts provided by Social Science Open Access repository (SSOAR). To determine the relevance of terms identified TF-IDF scores are calculated. Extracted entities are linked to the SAGE Thesaurus[10] as an external knowledge base which we extended by automatically extracting further terms from SSOAR fulltexts, such as abbreviations, synonyms and related terms. The entire extraction pipeline is described more detailed in the Deliverable D3.3 (Vagliano et al., 2019) (Section 2.2). We inserted all extracted entities together with a relevance score in the search Index at the MOVING platform to be incorporated in the Concept Graphs visualisation, such that a user is able to explore the entities related to the document under study.

2.3.12.3 Visualisation of entities

The entities extracted from the documents of the results might contain information about the relevance of the entity to the document (TF-IDF score), or the positions, which are represented as an array of indexes, where the entities occur in the document of the result. The information about the relevance of the entity to the document is used to display a relevance edge between the document and the entity, which can be seen in Figure 20. Edges representing relevance are drawn with a dashed line, while the relevance is shown through the opacity of the line. The darker the dashed line, the more relevant the entity is in the document.

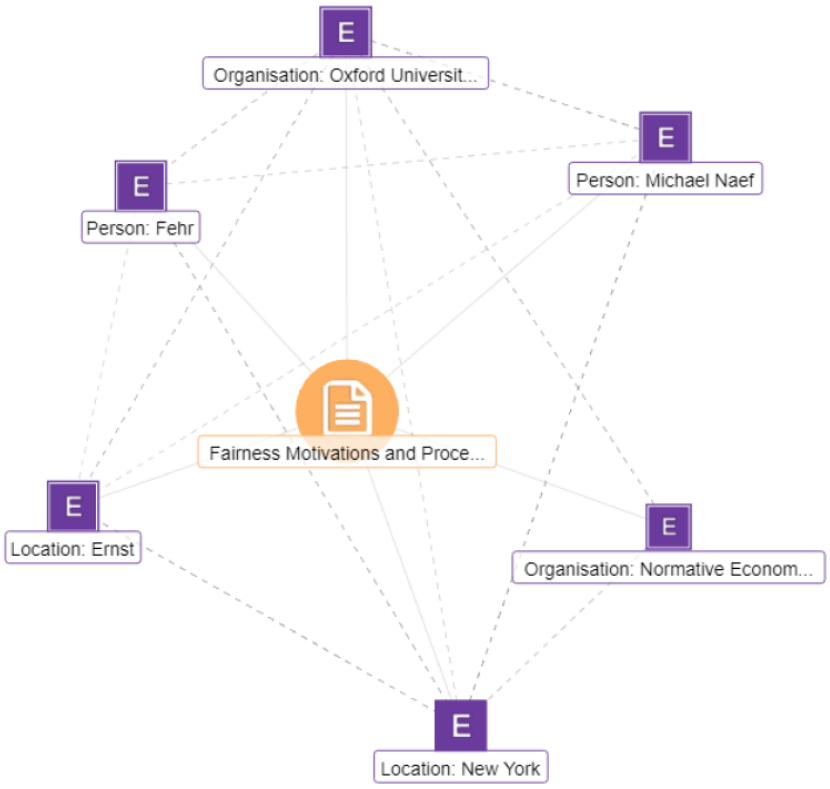
The positions of the entities in the document are on the other hand used to display the relevance between entities (figure 20). The closer two entities occur in a document, the darker the edge will be between them. If the distance exceeds a certain threshold, no relevance edge will be displayed. The visualisation of entities is integrated and is available on the MOVING platform.

3 Interface Annotations

Each interface element in the MOVING platform needed to be manually annotated, so the interaction data provided by WevQuery (see section 4.2.4 from D3.1 (Blume et al., 2018)) could be linked to particular user actions. HTML attributes (mainly `ID` and `class`) were modified depending on the nature of the interaction. The following subsections explain the criteria followed to modify these attributes, provide real examples from the MOVING platform, and details the requirements of the MOVING platform that initially motivated these annotations.

[10] `http://methods.sagepub.com/`

Figure 20: Concept Graph - entity nodes and relevance edges: entity to entity relevance



3.1 Methodology

Tables 2, 3, and 4 show some examples of the documented annotations. We have selected these examples as they can be generalisable to other web sites. Further examples can be found in Section A.3. Each table shows a description of the interface element, its location in the page, and three HTML attributes that can be used to locate interactions with this particular element (`type`, `ID`, and `Class`).

Depending on the nature of the interface element, a different attribute of the HTML attribute was selected to be changed. On the one hand, interface elements corresponding to uniquely identifiable user actions were annotated modifying their `ID` attribute. These actions could be pinpointed in single pages, as only that interface element allowed the user to carry out that action. Examples of these elements are the search action (see "Search button" in Table 3 and Figure 21) and the selection of particular visualisations (see "uRank" or "Tag cloud" in Table 4)). On the other hand, there were generic interface elements that shared similar actions. A common example is a search result item. Rather than looking for a particular result, all the clicked results can be retrieved looking for mouse interactions on elements of `class` "result-item" (see "Result items" description in Table 4). In this particular case, the document referred to by the result is stored as the `ID`, so the interaction can also be linked to a particular resource.

3.2 Documented annotations

A table documenting the annotated interface elements was created for each component of the MOVING platform. Table 2 documents the interface elements common to all MOVING web pages, such as the header and links to the various pages (e.g. "Search" or "Communities"). Table 3 documents the content specific to the simple and advanced search pages. Although similar, the content between these two options differed. The basic search would provide a single "Search text input" while advanced search contained various individual forms to further customise the search (e.g. "Person" field). Figure 21 shows an example of how the annotations shown in Table 3 relate to the various interface elements.

The search results page (see Table 4) also presents various filtering options referred to as "Faceted". Interaction with these elements can be retrieved using the `class`, or the `ID`, which allows to extract interaction with a filter of a particular category (the filter category name is included in the `ID`, replacing "CATEGORYNAME").

Table 2: Annotation of common page elements

Description	Location	Type	ID	Class
Navigation toggle (Mobile)	Top	button	navbar_toggle_button	
MOVING Logo	Top	link	home_link	
Search	Top	link	search_link	
Communities	Top	link	projects_link	
Learning	Top	link	learning_link	
Contacts	Top	link	community_link	
My account	Top	link	account_link	
Sign in	Top	link	signin_link	
Contact	Bottom	link	contact_link	
Terms of Service	Bottom	link	terms_link	
Privacy Policy	Bottom	link	privacy_link	
Imprint	Bottom	link	impressum_link	
Toggle left side	Bottom	link	toggle_leftside_link	
Toggle right side	Bottom	link	toggle_rightside_link	

Table 3: Search page annotation

	Description	Location	Type	ID	Class
Simple Search	Simple Search	Middle	link	search_simple_link	
	Advanced Search	Middle	link	search_advanced_link	
	Container for search items	Middle	form	search_form	
	Research/Learning/Funding button	Middle	button	search_domain_button	
	Search button	Middle	button	search-button	
	Search text input	Middle	input	q	
Advanced Search	Simple Search	Middle	link	search_simple_link	
	Advanced Search	Middle	link	search_advanced_link	
	Container for search items	Middle	link	advanced_search_Form	
	Search button	Middle	button	search-button	
	Research/Learning/Funding dropdown	Middle	input	search_domain	
	Title	Middle	input	advanced_query_title	
	Abstract	Middle	input	advanced_query_abstract	
	Fulltext	Middle	input	advanced_query_fulltext	
	Person	Middle	input	advanced_query_person	

Figure 21: Simple search interface annotation example

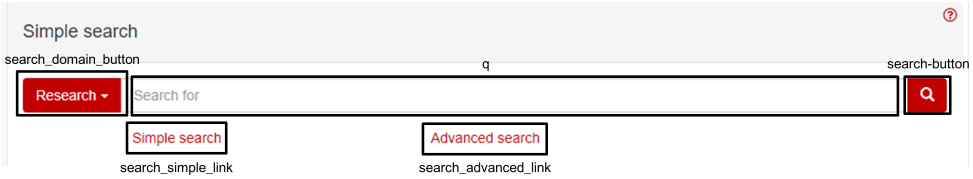


Table 4: Search result annotation

Description	Location	Type	ID	Class
Results	Above result list	link	search-tab-results	
Concept Graph	Above result list	link	search-tab-concept-graph	
uRank	Above result list	link	search-tab-urank	
Tag cloud	Above result list	link	search-tab-tag-cloud	
Top concepts	Above result list	link	search-tab-top-concepts	
Top sources	Above result list	link	search-tab-top-sources	
Date mentions	Above result list	link	search-tab-date-mentions	
Result items	Middle, list of results	link	Database document ID	result-item
Result items authors	Middle, list of results	link		result-item-author
Result items concepts	Middle, list of results	link		result-item-concept
Pagination elements	Below result list	link	pagination	
Sort by date	Above result list (Right)	link	sort_by_date	
Sort by relevance	Above result list (Right)	link	sort_by_relevance	
Filter container	Left side	form	facets-form	
Remove Filters	Left side	button	remove_facet_filters_button	
Faceted category collapse	Left side	link	collapse_facet_CATEGORYNAME	collapse-facet
Faceted checkboxes label	Left side	div	faceted_checkbox_CATEGORYNAME	checkbox
Faceted checkboxes	Left side	input	filters_CATEGORYNAME_	filters-checkbox

3.3 Conclusion

Extraction of user interaction data using these annotations has been employed for user behaviour analyses (see Sections 4 and 5), as well as for other components of the MOVING platform.

The **learning how to search widget** has used the number of times each of the search features has been used to provide a barplot that makes the user reflect on their use of the platform. Details on the use of the interaction data by this widget can be found in Section 5 (Adaptive training Support) from D4.2 (Gottfried, Pournaras, et al., 2017) We have carried out a study to determine the effect of this widget and the results can be found in Section 5.3.

The **recommender system** has used the searches carried out to provide suitable recommendations adapted to each user's search behaviour. The use of the interface annotations allows the widget to discern if the search was carried out via the use of the simple search (`search_form`), advanced search (`advanced_search_Form`), or an update to the filter options (`facets-form`).

The **search history** used a custom event that provided the list of queries run along with the number of results for each one.

The **crawler** extracted the document IDs of the result items the users clicked on and used them to update crawling criteria. For example, if documents from a particular source are often used, the crawling frequency for that source would be increased. Details of this adaptive crawling can be found in Section 3.1.4 of D3.3 (Vagliano et al., 2019).

4 Analysis pipeline

In this section the analysis pipelines employed to produce the data for the various studies reported in Section 5 are explained. First, the list of engagement metrics extracted from the interaction data are detailed in Section 4.1. These metrics have been employed in the Learning-how-to-search (LHTS) and Massive Online Open Course (MOOC) studies. Then, the methodology employed to carry out a data-driven analysis of emerging low-level behaviours is presented in Section 4.2. Some of the design decisions were dependent on the nature of the data from the MOOC and Working Environment studies.

4.1 Extraction of engagement metrics

An initial list of engagement metrics have been identified from research (Lalmas, O'Brien, & Yom-Tov, 2014). From that list, a final list of metrics was composed, based on the possibilities of the gathered interaction data, and the nature of the MOVING platform. The following are the engagement metrics identified from previous research:

- **Dwell time** is a common metric describing the amount of time spent on a web page. In general this metric can be used directly as a measurement of engagement. However, depending on the purpose of the Web site the expectations for this metric can differ. If the Web site is a search engine, it might be desirable to keep dwell time low, as it can be considered a metric of efficiency.
- The **number of pages visited** can be used as an indicator of the amount of content explored by a user. The higher the engagement, the more pages a user might visit on a web site.
- **Scroll distance** is a relevant indicator of users engaging with the content in the page. The amount of scrolling can also be used to determine which parts of the web page have been visualised by the user (Lagun & Lalmas, 2016).
- The **number of clicks** can be used as a measurement of the amount of interaction with the web site.
- A high **number of sessions** might indicate that the users are willing to come back to the web site frequently.

The following list details the engagement metrics extracted from the interaction data, to be employed in the studies. These metrics have been computed for each user. If the analysis consisted on a comparison between user groups, these metrics were averaged within user groups to then look for statistically significant differences (e.g. analysis carried out in Section 5.3). Alternatively, plots showing the distribution of these metrics for each group have been compared to discern engagement differences between groups (see Section 5.2).

- **Active time** is the equivalent of dwell time. It has been computed measuring the time spent interacting with the page, excluding any inactive period longer than 50 seconds. Excluding inactive periods, which could have been caused by the visitors doing something away from the computer or mobile device, allow us to compute a precise measurement of the time spent by users interacting with the page.
- **Inactive time** is a measurement of the inactive periods discarded by the previous metric.
- **Number of episodes** is the equivalent of *number of sessions* previously mentioned. We use the term episode instead of session, as we are using a custom timeout (40 minutes) to split the episodes, instead of the common 30 minutes used in research and commercial Web analytics tools. This timeout has been derived from the analysis of real interaction.
- **Number of keypresses**
- **Number of pages visited**
- **Number of clicks**
- **Number of learning resources disclosed**: this metric is specific to the design of MOVING. In the MOOCs the various learning steps are initially hidden, and the users can show them one at a time by clicking on an accordion item. Extracting the number of clicks on these elements allow us to determine how many "learning resources" each user has visualised.
- **Scroll distance**: the distance scrolled using mousewheel interaction has been computed.
- **Video interaction**: MOOCs make use of external videos to provide information to the users. Interaction with these videos could be extracted using APIs [11] specific to the video source. Only user interaction events were considered (play/pause, and skip video).
- **Number of times each feature was used**: inspired by the "Learning how to search widget", that shows users how many times they have used each feature in the MOVING platform, the number of times each feature was used was included as another engagement metric.
- **Number of Selected Results**: this metric counts the number of times a user clicked on a result.

4.2 Data-driven behavioural clustering of users

N-grams are a useful method for capturing low level patterns of user behaviour without the need to parse and translate every user event into a specific GUI tasks. Such complex parsing is likely to be manually time consuming and/or computationally expensive and may lead to overfitting on current user interaction datasets. N-grams on the other hand can be run on a variety of GUI environments and are easily scalable to be extracted from large interaction event datasets.

We define an event n-gram as consisting of a time ordered sequence of *n* consecutive events by a single user that is fully contained within a single user *episode* (see 'number of episodes' above in Section: 4.1). Before creating these n-grams, events are pre-processed to remove noisy and repeating data.

4.2.1 Pre-processing

1. Select URLs: For each experiment dataset (LHTS, MOOC and Search Environment) only the URLs corresponding to the appropriate domain were considered, using a list of acceptable URL prefixes for each dataset.
2. Before carrying out the analysis, the data was processed to improve the readability of the results. Some events typically follow each other, like a mouseup following a mousedown, introducing noise in the data, so we combined them into a single event, mousepress. Other events correspond to similar actions, and were given a more comprehensive name. For example, the action of using the scrollwheel of the mouse (mousewheel) and the change of scroll state in a window (scroll). On the other hand, sometimes a single event can be mapped to various actions. Keyboard interaction can indicate that the user is typing, or that the user executed a command (such as pressing enter). We mapped those actions as separate events, to allow us to further understand them. Finally, some known combinations of events were mapped as new events, to provide concise useful information in single events (e.g. blurfocus_leavepage) The following is a summary of this processing:

[11] `https://developers.google.com/youtube/iframe_api_reference`, `https://developer.vimeo.com/`

(a) mousedown/mouseup => `mousepress`

(b) mouseover/mouseout => `mouseinorout`

(c) mousewheel/scroll => `scrollorwheel`

(d) 'keydown' events were categorised into ' `keydown_write`' (any writing event such as letters, numbers and punctuation), ' `keydown_command`' ('ctrl', 'shift','tab','alt','enter') or `keydown_other` (any other keydown events).

(e) A `blurfocus_leavepage` event was created whenever there was a sequential blur/focus events on same page).

(f) A `windowfocus_switchtab` event was created whenever there was a window focus event where the use went onto new page (new url)).

3. The data was then sorted chronologically for each user and episode. A new field was generated for each event '`event-node-url`', indicating the event name, the UI target of the event, and the URL (trimmed to the higher level domain) where it took place, respectively.

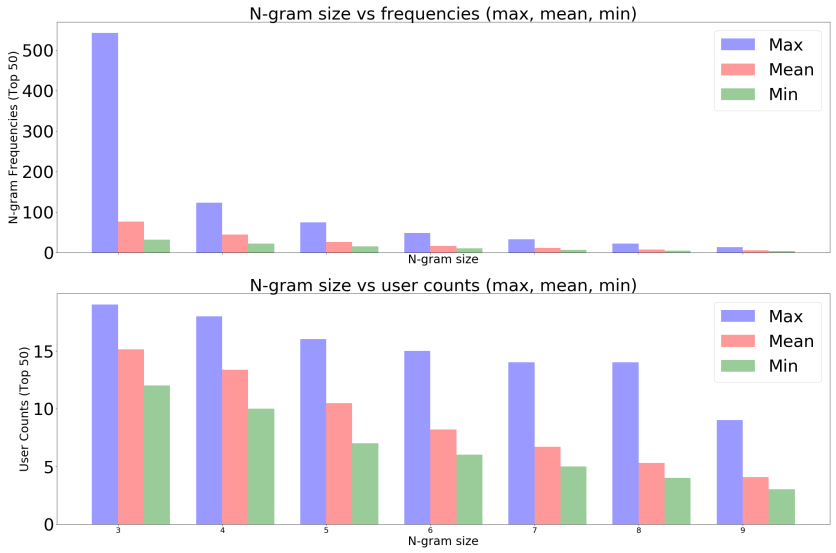
4. Finally, to avoid having n-grams with repeated events in them, all directly sequential, identical `event-node-url` events were merged (and marked as multiple). The case of `scrollorwheel` was special, as these events were merged even when the `node` was different. This was done as events from the same scrolling action often were classified into different `nodes` as the mouse moved around the page, and the UI target has not been considered relevant for this particular event.

4.2.2 N-gram creation

To concatenate series of user episode events into n-grams, a sliding window approach was used. This approach considers all possible subsequences of a given sequence as candidates to be extracted as n-grams. This results in overlap and therefore repetition of events between different n-grams but provides full coverage of all sequences used.

Firstly a suitable value for n (number of consecutive events) is required. The selection of this number introduces a tradeoff. Higher values for n produce longer series of events and more detailed descriptions. However, the frequency of the n-grams is reduced, hindering the discovery of interesting higher level patterns. Some preliminary visualisations were produced to look at the effects of this trade off. We collected n-gram count and unique user coverage for n-gram collections where n is in the range [3:9]. The counts are shown in Figure 22 (LHTS), Figure 23 (MOOC) and Figure 24 (working environment). For each value of n (x-axis), the frequency for each n-gram occurrence (upper bar chart) is shown, along with the number of users who used each n-gram (lower bar chart). Due to the long tailed distribution in n-gram occurence counts, the bar charts display only the top 50 n-grams in each respect: occurrence counts and user counts.

Figure 22: LHTS: N-gram size with n-gram counts and user coverage counts



Due to the smaller size of the LHTS study (19 subjects) it was decided to select the n-gram size where the top n-gram sequences were performed by at least half of subjects. n=4 was therefore selected and during the analysis this size was found to be large enough to allow patterns to be extracted for a large number of frequent n-grams, across all users who were fully engaged in this study. Testing was also performed using n=5 and the n-gram frequency data became noticeably more sparse for each user. It was also felt that 4-grams are long enough to represent meaningful sequences in user interactions

Figure 23: MOOC: N-gram size with n-gram counts and user coverage counts

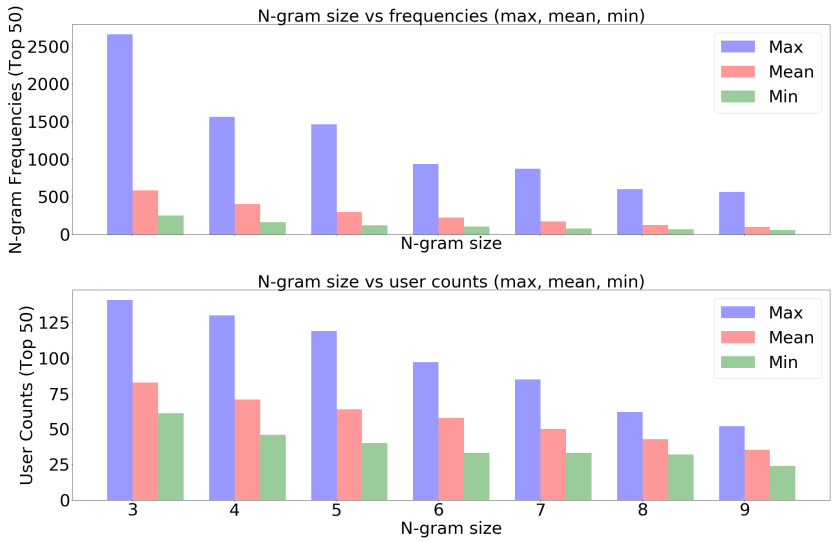
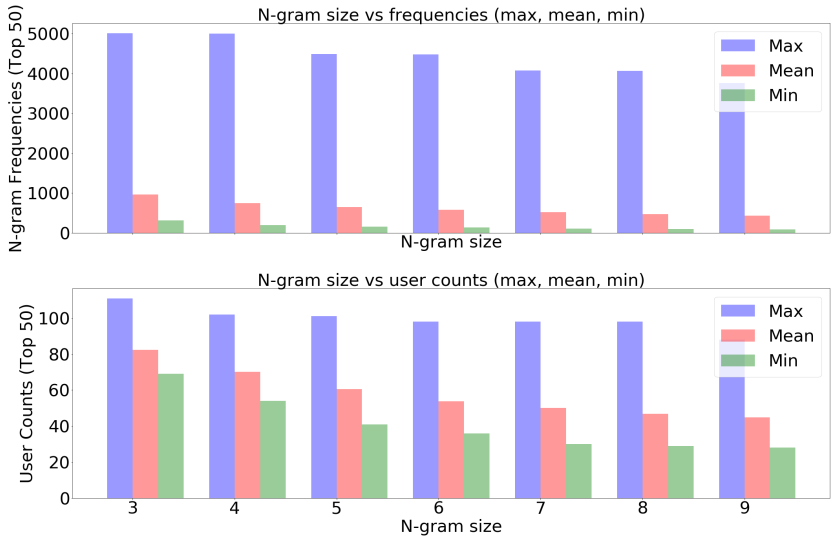


Figure 24: Working environment: N-gram size with n-gram counts and user coverage counts



For the two studies of live platform users (MOOC users and working environment users) the number of users were much larger (193 and 152 respectively) which allowed more flexibility to choose larger n-gram sizes. Evidence of this can be seen in Figures 23 and 24 and this was also shown to be the case when the higher n-gram sizes were analyed. For consistencey with the LHTS analysis, we again chose n=4.

Once the 4-grams were compiled for each user/episode, further analysis was done to check the distribution of these 4-grams occurring across episodes. Figures 25, 26 and 27 show that the vast majority of episodes contain the lowest number of n-grams but with a long-tailed distribution of low numbers of episodes with very high 4-gram counts.

The MOVING working environment also followed this pattern but had more spikes of episodes with very high 4-gram counts (Figure 27). Of the 3 studies, the use of the search facility could be said to be most

Figure 25: LHTS: frequency of episodes with 4-gram counts

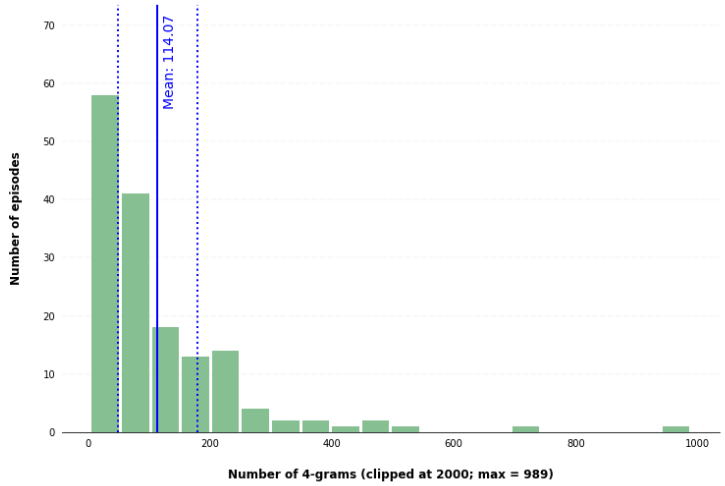
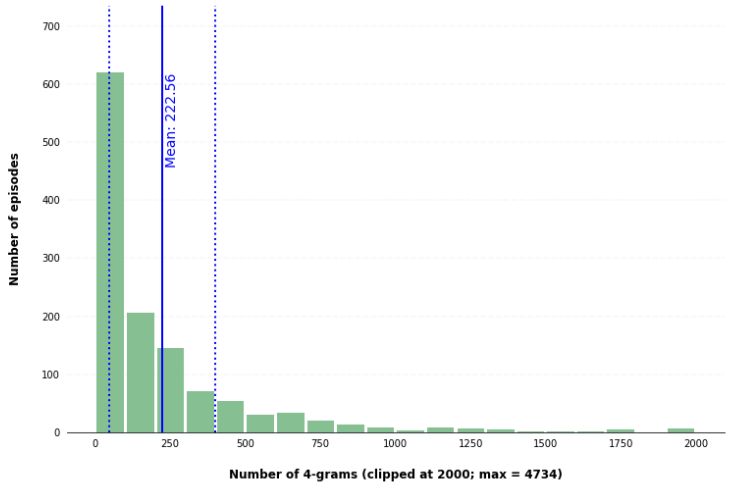


Figure 26: MOOC: frequency of episodes with 4-gram counts



representative of the usual 'free style' web usage, where instructions and guidelines are not being followed and this is likely to explain the more random high n-gram count results and longer episodes with more events.

4.2.3 Output

Bar-chart In order to observe the 4-gram frequencies and their distribution across the users, a bar chart was created for each study (Figures 28, 29 and 30), with the frequency bars for the top 50 4-grams broken down into colours for each user. Note that the legends are not shown for the MOOC and working environment bar chart due to the higher number of participants.)

Figures 28, 29 and 30 show that the 4-gram frequencies follow a long-tailed distribution and the frequent 4-grams are perfomed by a varied set of participants. The proportional drop off is more gradual for the LHTS study followed by the MOOC and then the working environment studies. This could be due to the more prescribed nature of the LHTS study (and slightly less so the MOOC), where the users followed specific tasks, so there is less evidence of ideosyncratic/unique user behaviours.

Figure 27: Working environment: frequency of episodes with 4-gram counts

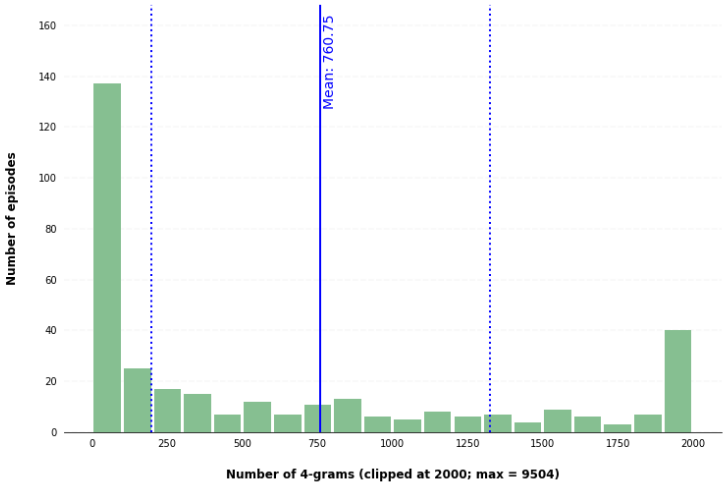


Figure 28: LHTS: Top 50 4-gram frequencies by user

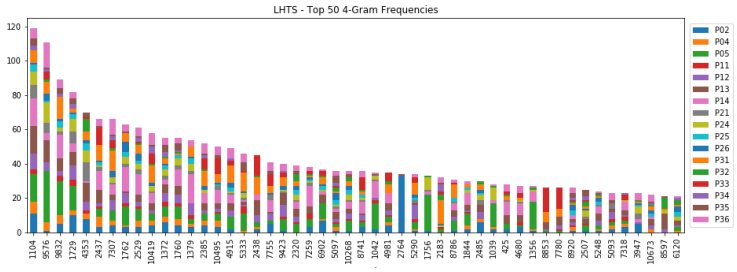


Figure 29: MOOC: Top 50 4-gram frequencies by user

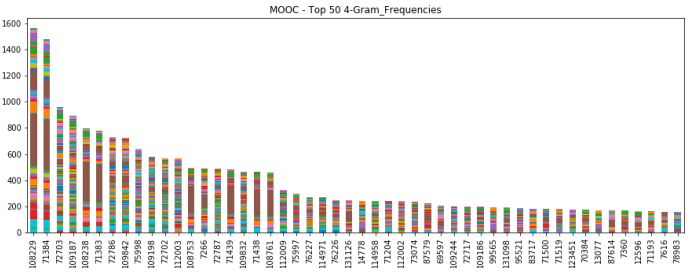
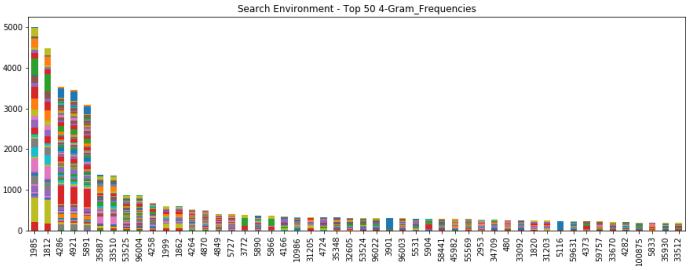


Figure 30: Working environment: Top 50 4-gram frequencies by user



Create Cluster-map We use the cluster-map visualisation as it performs hierarchical clustering on both the users and the 4-grams and plots the results as a heatmap. As the the n-grams as well as the users are clustered, similar n-grams are merged (by usage frequencies). So it groups very similar n-grams together.

The parameters used when creating the cluster-map are:

- Clustering Distance Metric: euclidean. This is the most simple distance metric and is used frequently by clustering algorithms. It calculates distance based on the straight-line length between 2 data points in the euclidean space. Various common distance metrics were tested, but none were found to produce more distinct clusters.
- Clustering Method: The 'Complete-linkage' clustering technique is a commonly used agglomerative hierarchical clustering technique that begins by assigning all elements to their own cluster. The clusters that are closest are iteratively combined into larger clusters. The 'complete-linkage' method uses the distance between two elements (one from each cluster) that are greatest distance apart (also known as 'Farthest neighbour clustering'). Complete linkage is good for finding compact clusters with roughly equal dimensions, but can be affected by noise. We mitigate this potential disadvantage by introducing a maximum cell count parameter, as described below, to remove noise. Other techniques were also attempted ('single', 'average' and 'centroid'), but none were found to outperform the selected method.
- N-gram count parameters: We introduced three further parameters to improve the clustering results and remove potentially noisy data. These parameters were selected depending on the study data being used and the number of users and 4-grams available. The first is *minimum n-gram count* which only allows the inclusion of n-grams with a minimum count in the data. The second is *minimum user count* which limits the n-grams included to just those which were performed by a minimum number of users. Finally, the *maximum cell count* is the upper limit of the count of 4-grams by each user, with any higher values being set to the maximum. This prevents very high counts from skewing the data, and 20 was found to be the best number across all 3 study datasets.

Obtain cluster labels The above cluster-map visualisation (dendrogram) was used to derive the final user clusters. This was done by using the dendrogram tree view of users (for example the horizontal tree-view along the top of Figure 31). A maximum of three of the tree-view's binary splits were considered and in the case of clusters that contained less than 5 users, the cluster is combined with it's sibling. For both the MOOC and Working Environment studies where this heuristic was followed, it resulted in 4 clusters. Users were then labelled with their cluster number and these were compared with the engagement metrics results.

5 Studies

We have compared the engagement of the MOVING platform with two comparable web sites. EconBiz[12] is the ZBW – Leibniz Information Centre for Economics's subject portal for Economics and Business studies. It offers a literature search across important German and international databases. VideoLectures.NET [13] is an award-winning free and open access educational video lectures repository maintained at the Center for Knowledge Transfer in ICT at the Jožef Stefan Institute. Both these sites allow users to search and access learning resources, in the form of scientific papers or videos. We extracted the average active time for visitors to these sites during the year 2018. In the case of EconBiz, where visitors go to find external resources, the average visit duration is 2 minutes 32 seconds. This site is comparable to the working environment of the MOVING platform, where users can also search for external resources, with an average visit duration of 15 minutes. In the case of VideoLectures, users stay in the page to watch videos, and the visit duration is 8 minutes 18 seconds. This site is comparable to the MOOCs hosted in MOVING, as they both provide content for the users to consume. The average visit duration for the visitors in the MOOC is 8 minutes 30 seconds. In summary, the engagement in the MOVING platform has been found to be higher for search tasks, and comparable in the case of platforms providing learning content.

The detailed behaviour analysis we carried out for the various studies in MOVING gave us precise insight into the percentage of users who were active in the platform. This information refers only to the users who actively visited the MOVING site and agreed to the interaction capture, as no interaction data would be available otherwise. Therefore it represents the percentage of users actively engaged with the platform, from all visitors to the site. This metric can be comparable to the bounce metric from commercial Web analytic tools. In the case of the working environment, ~45% of users were found to be less active than the rest. This percentage also corresponds to the attrition rate shown in the MOOC (see Figure 32), where around ~50% of users dropped out after the first module. In comparison, VideoLectures had a higher bounce rate of ~57% during the year 2018.

[12] https://www.econbiz.de/
[13] http://videolectures.net/

5.1 Working environment

Presents a data-driven analysis of interaction data captured from the search environment, clustering users into various profiles according to their behaviour.

Following the clustering method described in Section 4.2.3 we grouped the users of the MOOC into clusters based on their interactive behaviour.

The following n-gram count parameters were set (see Section 4.2.3):

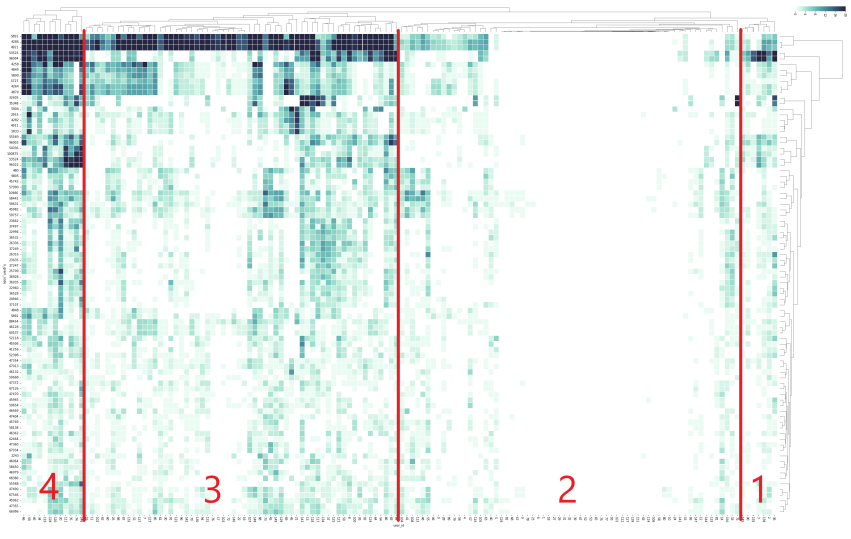
1. *Minimum n-gram count* (only allows the inclusion of n-grams with a minimum count of): 21
2. *Minimum user count* (minimum number of users who exhibited each 4-gram): 45
3. *Maximum cell count* (upper limit of the count of 4-grams by each user): 20

This resulted in the clusters displayed in Figure 31. We end up with four clusters if we group together the users under each subtree when spliting the hierarchy three times: the distribution of users per cluster is 7 (Cluster 1), 65 (Cluster 2), 60 (Cluster 3) and 12 (Cluster 4). A qualitative analysis of the behaviour of users when they interact with the working environment suggests that the clusters are indicators of the amount of activity. Darker cells mean that those specific n-grams were exhibited more frequently. Hence, those users in Cluster 4 are more active than those in Cluster 3, 2 and 1. Cluster 2 stands out as the cluster with individuals who are less active.

The top lines of the dendrogram in Figure 31 are search behaviours. Understandably these are the behaviours that are exhibited more often in the MOVING platform in that its working environment is conceived as an information retrieval system.

- Cluster 1. These users are more information explorers than searchers. While they exhibit search behaviours they are more active in exploring the results of the MOVING platform.
- Cluster 2. In this cluster we have the users who are not very active and those who search and consume information a few times. This are visitors who came to MOVING but did not revisit the platform in a continuous fashion.
- Cluster 3. This is the group of users who are active in the MOVING platform. The main feature of these users is that they are searchers of information and make use of the functionalities provided by the search environment including faceted search.
- Cluster 4. This is the group of users who are more active in the MOVING platform. They seek for information and consume it, and they are engaged. Unlike the users of Cluster 3, they make use faceted search functionalities in addition to the visualisations. Scrolling behaviour on the search result pages indicates that these users are not only searchers, but also information explorers and consumers.

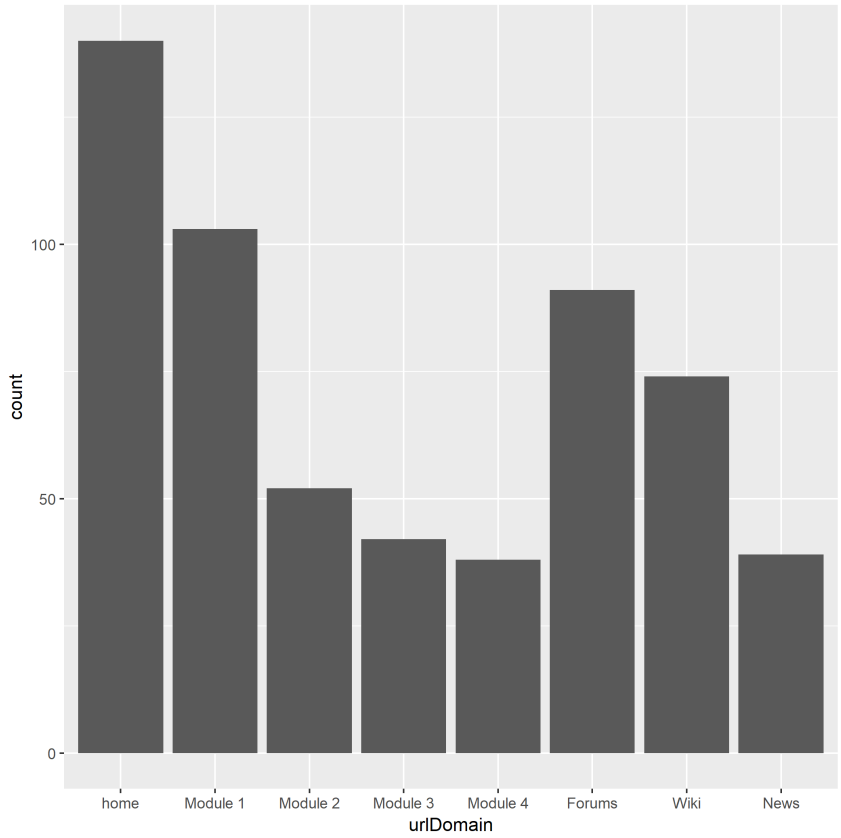
Figure 31: Hierarchical clustering of interactive behaviour on the Working environment (columns=users; rows=4-grams).



5.2 MOVING MOOC engagement analysis

We collected user interface events from the MOVING MOOCs about science 2.0 and open research methods [14]. The MOVING MOOC took place in two waves: 12 November – 16 December, 2018 and 21 January – 17 February, 2019. The MOOC consists of four modules, each of which was made available every week over a period of four weeks. After that, all the materials were available to students and they could access them in any order at any time. The MOOC contained static materials (i.e. Web content), pointers to external resources, videos, a forum and wiki where students could write down their reflections on the course. Since the MOOC follows the connectivist principles there is a strong emphasis on collaboration and creation, and there are no tests. Apart from the four main modules, the analysis has also considered: the homepage, where users were given the option to enroll in the MOOC; the forums, used by users to carry out various tasks from the various modules; the wiki, which provides information about the MOOC; and the news, with updates informing users of when new modules are available.

Figure 32: Number of unique users per MOOC resource.



For the analysis we used interaction data captured from participants of the MOOC who gave their consent to the interaction tracking (a total of 193 participants). Figure 32, which shows the number of users that accessed the above-mentioned resources, confirms a typical engagement distribution of MOOCs: i.e. a decline in participation over the weeks in which the MOOC takes place. Interestingly, beyond the course materials, users were active in the forums, wiki and news sections of the MOOC. This is mostly due to the topic of the MOOC about open science and the encouragement by the instructors to use collaborative tools.

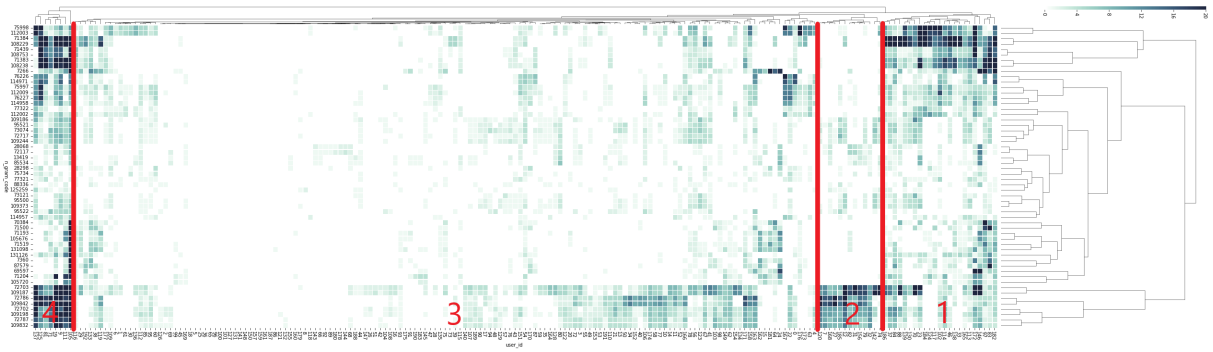
Following the clustering method described in Section 4.2.3 we grouped the users of the MOOC into clusters based on their interactive behaviour. The following n-gram count parameters were set (see Section 4.2.3):

1. *Minimum n-gram count* (only allows the inclusion of n-grams with a minimum count of): 21
2. *Minimum user count* (minimum number of users who exhibited each 4-gram): 43
3. *Maximum cell count* (upper limit of the count of 4-grams by each user): 20

[14] `https://moving.mz.tu-dresden.de/mooc`

This resulted in the clusters displayed in Figure 33. We end up with four clusters if we group together the users under each subtree when spliting the hierarchy three times. The distribution of users per cluster is 23 (cluster 1), 13 (cluster 2), 149 (cluster 3) and 8 (cluster 4).

Figure 33: Hierarchical clustering of interactive behaviour on the MOVING MOOC (columns=users; rows=4-grams).



The engagement metrics listed in section 4.1 were computed for each of the clusters. For the sake of brevity, we will display only those metrics that are relevant to characterise different clusters. Figure 34 shows active time (left) and number of episodes (right) per cluster. In terms of **active time** Cluster 1 shows the typical decline in participation from Module 1 to Module 4 while being active on the forums. Cluster 3 can be considered an attenuated projection of cluster 1 active time in Module 1–4 except for the participation in the forum which is minimal. Cluster 2 consists of those users whose active time is low compared to other clusters and they even withdraw from the MOOC after checking the third module.

We observe a correspondence between the **number of episodes** and active time, which suggests that the more often users accessed the MOOC the more overall time they engaged with it. Yet, there is an exception for Cluster 4 in that while the number of episodes follows the expected attrition pattern, the active time does not. The boxplots suggest that these users accessed Module 2 and Module 4 less often but for longer periods of time. Still Cluster 1 and Cluster 4 exhibit very similar engagement metrics.

When it comes to engagement metrics (see Figure 35) that derive from interaction events, the **number of visited pages** and **number of resources explored** provide a similar picture as the time-based metrics discussed above: Cluster 1 and Cluster 4 show similar enagagement behaviours whereby the number of visits to different entries of the forum is the feature that stands out more clearly. The number of resources explored suggests that users of Cluster 3 show a balanced engagement across resources. Interestingly, the **number of clicks** and **scroll distance** mirror time-based metrics for Cluster 1 and Cluster 3. This mirroring effect is not observed in other clusters as in the particular case of Cluster 4, the number of clicks and scroll distance is higher than in the remaining clusters.

Taking into consideration the engagement metrics, the clusters characterise their users as follows:

- Cluster 1: high engagement with course materials and forum, typical attrition rate.
- Cluster 2: low engagement and eventual withdrawal.
- Cluster 3: medium engagement overall, with minimal attrition. Users in this cluster exhibit fewer episodes than other users.
- Cluster 4: high engagement with course materials and forum, typical attrition rate. Their behaviour is similar to Cluster 1 but engagement with the Forum is higher. This group is extremely active in terms of interactions with the MOOC including mouse clicks and scrolling distance.

5.3 Learning-how-to-search widget study

As a preliminary remote usability study, we evaluated the effect that the Learning-how-to-search (*LHTS*) widget has on visitors to the MOVING platform. This study needed to be longitudinal, so the Learning-how-to-search widget would gather enough data to adapt to each visitor.

5.3.1 Research Questions

Main Research Question

Figure 34: From left to right: Active time and number of episodes for the identified clusters. From top to bottom: group 1, group 2, group 3 and group 4

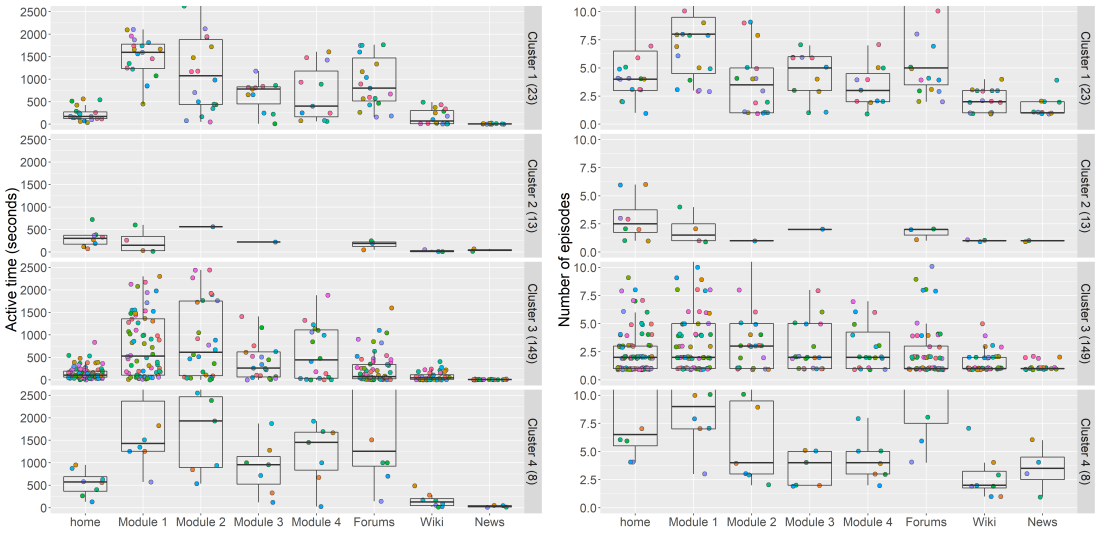
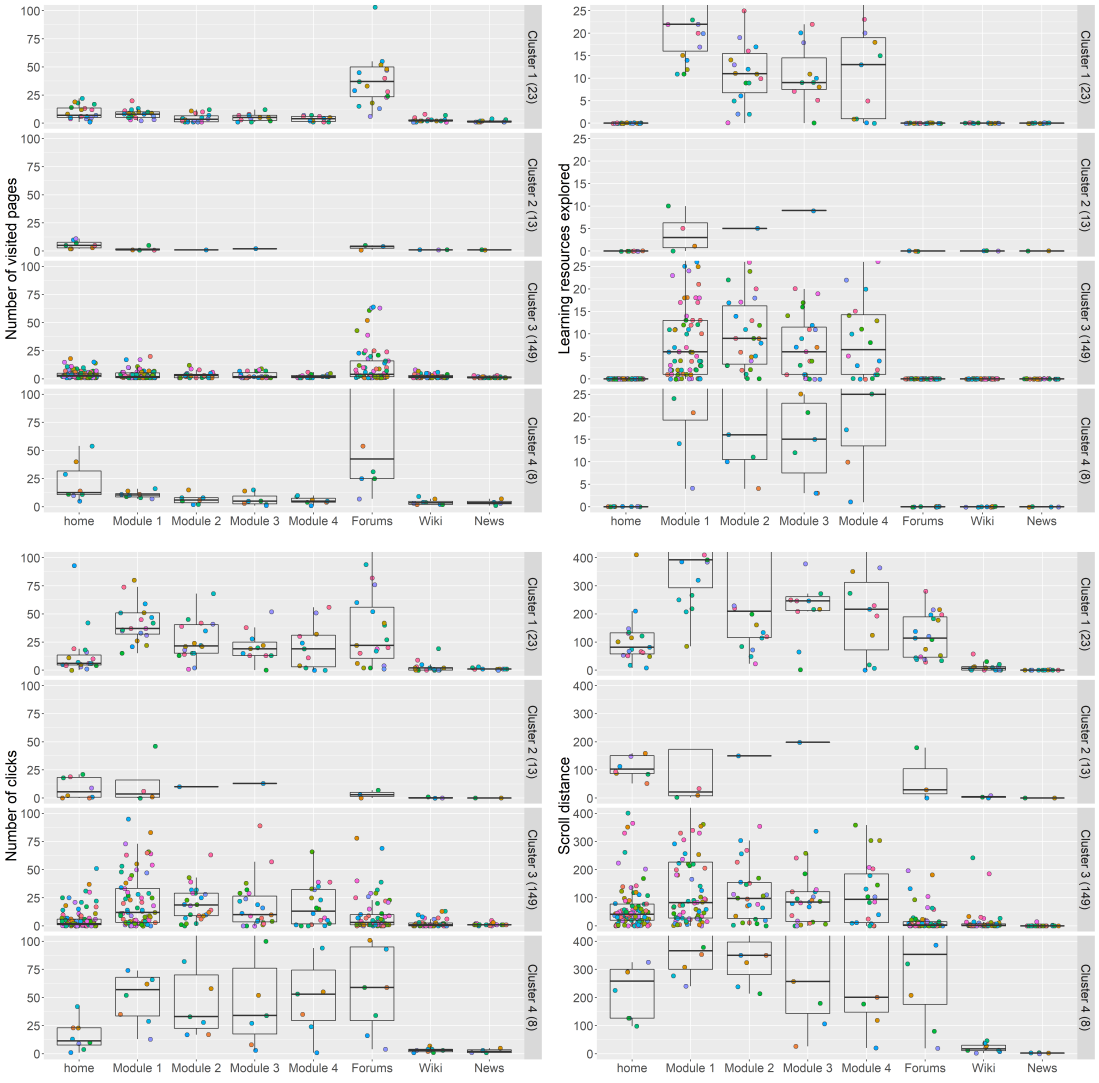


Figure 35: From top-left to bottom-right: number of visited pages, learning resources accessed, number of clicks and scroll distance.



Main RQ: Does the LHTS influence the behaviour of the users of the platform?

The design of the study was guided by this main research question, which builds on top of the motivation for the study. The assumption that guides the study is that the inclusion of the LHTS in the interface will have a positive effect on the interaction, encouraging visitors to try different features, keeping their attention (thus reducing bouncing rate) and reducing attrition. However, this main research question is too vague, and needs refinement in order to be able to effectively guide subsequent analyses.

Analysis Research Questions The following analysis research questions have been formulated to narrow down the design of the study. Each of the research questions focus on a different aspect of the interaction that could be affected by the inclusion of the Learning-how-to-search widget in the interface.

RQ1: Engagement metrics Did the widget lead to significant changes in the engagement of users?

RQ2: Navigational patterns Did the widget lead to different navigational behaviours?

RQ3: Use of functionalities Did the widget lead to different use of functionalities?

5.3.2 Design

As a way to evaluate the effect of the widget on the visitors, a comparison between the use of the platform with and without the widget was planned. The MOVING platform was modified so the Learning-how-to-search widget would only be shown under particular conditions.

The study was split into two periods of one week each. During the first period the participants would use the platform without the Learning-how-to-search widget, which would be activated for the second period. This approach allowed us to compare the use of the platform with and without the widget, as well as allowing the widget to use interaction data captured during the first period to provide recommendations.

For each period participants were asked to carry out one search task per working day. These tasks follow a strict order, so if a participant missed one, they would have to carry it out the following day before they were given the next one. Therefore, participants would carry out up to five tasks per period.

We tried to keep tasks from both periods as similar as possible by using the same instructions, but changing the topic to search for. The following are the tasks given to participants, where TOPIC was replaced with "Big data" or "Global warming" depending on the period:

1. What is TOPIC? Provide a general description of the topic
2. Identify key relevant topics related to TOPIC. Select one and provide details on their relation.
3. Prominent authors about TOPIC and their domains. Just select a few of them, and describe what their publications are like.
4. Select one relevant paper related to what you have found during your research this week.
5. For the author of the content you sent yesterday, find their most prolific year.

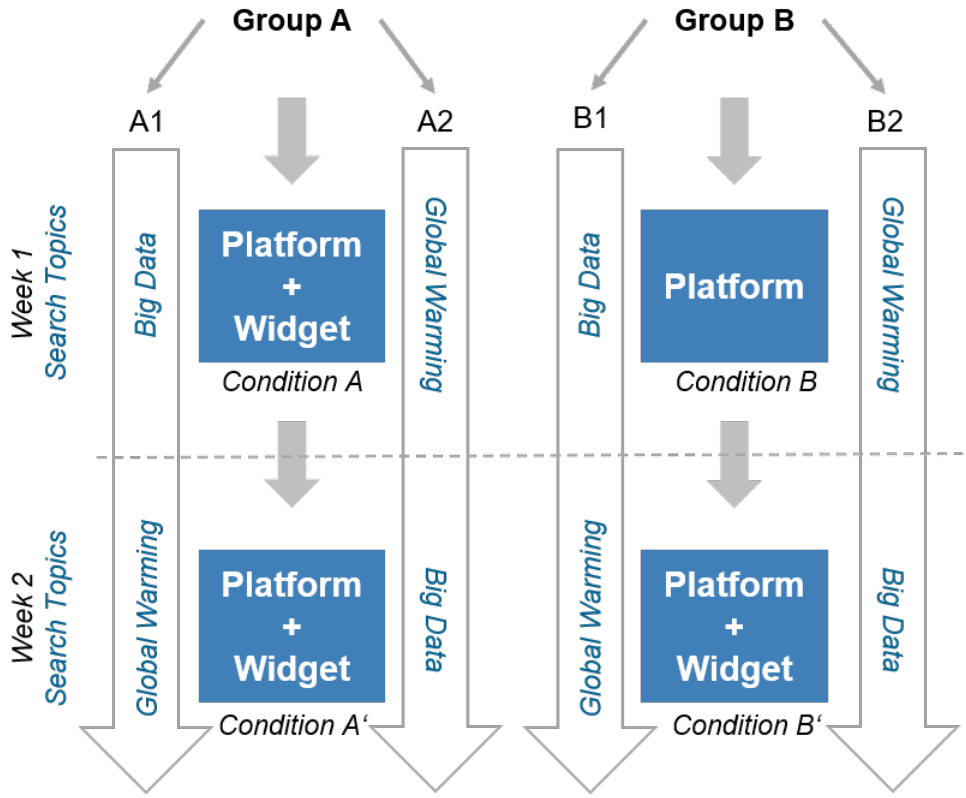
Although we tried to keep tasks from both periods as similar as possible, we randomised the order of the assigned topics to determine if starting with a particular topic had an effect on participants' behaviour. Additionally, we also included a control group in the study to account for the learning effect, inherent to the use of a new platform over a prolonged period of time. This control group carried out the same tasks, but in their case the Learning-how-to-search widget was visible for both periods.

Figure 36 summarises the design of the study. Group A is the control group where the LHTS widget is always visible and is formed by two subgroups where the topics for the search tasks are randomised. Group B is the study group where the LHTS widget is only visible during the second period (*Week 2*) and is also formed by two subgroups with randomised search topics.

5.3.3 Interaction data

Interaction data from the participants in the study was captured via UCIVIT (Apaolaza, Harper, & Jay, 2013) (see section 4.1 from D3.1 (Blume et al., 2018)). This data was then processed to answer the formulated analysis research questions. Features describing engagement metrics, the use of functionalities, and usage patterns needed to be extracted, so comparisons between the various groups could be made.

Figure 36: Diagram of the study design.



Engagement metrics As a way to determine if the Learning-how-to-search widget had an effect on users' engagement with the platform (see Research Question 1 above) we needed to extract features related to engagement from users' interaction. We combined common engagement metrics from previous research with metrics specific to the current study. The final list of engagement metrics are the following:

- **Active time**: the time elapsed carrying out the task. In order to ensure this metric reflects only the time spent by the participant interacting with the page, inactive periods longer than 50 seconds were not accounted for.
- **Inactive time**: the MOVING platform indexes external resources, so the results shown for search queries take visitors to external links. Therefore, visitors might spend much of their time with the platform visiting external pages, where no interaction can be captured for analysis. Users might even carry on searching for information from these external pages instead of returning to the MOVING platform, indicating a lack of engagement.
- **Number of Selected Results**: the number of results explored for each task can be an indicator of search engine efficiency, but also of engagement. Selecting more than one result per task may not indicate inefficiency, but rather curiosity to explore multiple resources.
- **Number of times each feature was used**: the main purpose of the Learning-how-to-search widget is for users to reflect upon their own search behaviour and to encourage visitors to interact with existing yet less used features from the platform. Therefore the number of times each of the features is used can be an indicator of the Learning-how-to-search fulfilling its purpose.
- **Number of episodes per task**: a timeout of 40 minutes is used to split interaction into different episodes (similar to the commonly used 30 minutes timeout to split interaction into sessions).
- **Number of searches**: this metric is similar to the number of selected results.
- **Amount of scroll**: measuring the scroll interaction from users is a common metric to measure engagement with a site.

Usage patterns Investigating the changes in navigational behaviour described in Research Question 2 arise non-trivial issues. Instead of extracting precise metrics to be compared between the various participant groups, pattern mining and n-gram analysis has been employed so navigational patterns for each of the groups can be extracted and compared.

We used the pattern mining module (see Section 4.2 in D3.2 (Vagliano et al., 2018)) in WevQuery to extract common usage patterns from the various comparison groups. Frequent itemsets have been extracted using the Apriori algorithm (Agrawal, Srikant, & others, 1994), and frequent sequences using the PrefixSpan algorithm (Han et al., 2001). The results allowed us to manually identify emerging behaviours for each group, so they could be compared.

Use of functionalities Research question 3 focuses on the use of functionalities, so we identified the main functionalities in the MOVING platform, and extracted the times each user interacted with them. We grouped these fuctionalities into three main categories:

- **Search**: Users interacting with the search button and users switching between simple and advanced search.
- **Learning-how-to-search**: Interaction with the Learning-how-to-search widget.
- **Visualisations**: Selection of one of the multiple visualisation tools (e.g. concept graph, tag cloud).

5.3.4 Participant feedback

Apart from remotely captured interaction data, participants were requested to answer several questionnaires during the study. The aim of these questionnaires was to captured the subjective experience of the participants with respect to the Learning-how-to-search widget. Although not explicitly requested, participants were also free to provide feedback about the platform in general.

Questionnaires Altogether, we distributed three different questionnaires, the pre-questionnaire at the beginning of the study, the in-between questionnaire after the first week of the study and the post-questionnaire at the end of the study. Pre-, in-between and post-questionnaire measured agreement with the questions on a 5-point Likert scale (from 1 = strongly disagree to 5 strongly agree). Additionally, qualitative data was collected with open questions.

The development of the questionnaire follows the model of Kirkpatrick' (Kirkpatrick & Kirkpatrick, 2006) for assessing training effectiveness in organisations. Kirkpatrick argues that learning should be evaluated at four levels that build up on each other in the sense of one level needing to be evaluated "positively" before success can be achieved at the next level. The levels are: reaction (how users react to technology), learning (whether and what learning occurs), behaviour (how behaviour of people is changed) and results (work performance). Second, we used the Technology Acceptance Model (TAM) proposed by (Davis, 1989) and extended by (Venkatesh & Davis, 2000), called TAM2, which is an information systems theory that helps to explain and predict user behaviour of information technology. While the original TAM focused on the perceived usefulness and perceived ease-of-use as factors for technology of being accepted by users, the TAM2 was extended and applied by (Park et al., 2009) and also includes social influences (subjective norm, voluntariness, and image), cognitive instrumental processes (job relevance, output quality and result demonstrability) and experience. By using the model of Kirckpatrick and the TAM as a basline for the development of the questionnaires and here especially with the focus on the last questionnaire, the purpose of each of the questionnaires is as follows:

- **Pre-questionnaire**: The purpose of this quesitonnaire was to get the participants' consent to participate in the study, to get some demogrpaphic information as well as to get some insights into their current web and search behaviour. Therefore, it consists of a very detailed participant information sheet and a consent form, a demographic questionnaire gathering demographic information like the age, profession etc. (11 items) as well as questions about the participants' background with respect to their computer and Web experience (6 items) based on Aula (Aula & Nordhausen, 2006).
- **In-between questionnaire**: With this questionnaire (8 items) we were aiming to illicit the first experiences of our participants with respect to the platform and the widget.
- **Post-questionnaire**: This questionnaire consists of questions about the feedback on the overall experience based on TAM, asking questions about usability (6 items), usefulness (8 items), attitude towards

the widget (2 items), widget specific questions (6 items), learning outcome (3 items), search behaviour (8 items) and technological self-efficacy (5 items).

The participant information sheet, the consent form as well as all questionnaires can be found in Appendix A.2.

5.3.5 Results

Demographics Altogether 15 (10 male, 5 female) people aged between 17 - 46 (Mean $(M) = 28.8$) participated in the study. Six participants were employees working in a research centre for data-driven business and big-data analytics in Austria, four were students studying computer science or software development. One participant had just finished school and another was attending the last year in school. The years of computer experience was on average $M = 15$ (Standard Deviation $(SD) = 7.18$) and the average year of web experience was $M = 14$ $(SD = 4.4)$. 73% of the participants stated that they use search engines and the web on a daily bases and 27% almost daily, 66.6% of them use a computer daily and 33.3% almost daily. With regard to a self-estimated evaluation of their own search skills, 20% of them rated themselves to be very skilled, 60% saw themselves as skilled and only 20% rated their search skills as neutral.

Procedure Before carrying out the analyses, we determined which groups (see Figure 36) should be compared, to answer each of the given research questions.

- Task order: to determine that the order in which the tasks are given had no effect in the interaction.
  - A1 vs A2
  - B1 vs B2
- Period comparison: to determine if there was a difference between the first and second period, in both the control group (A), and the study group (B).
  - A period 1 vs A period 2
  - B period 1 vs B period 2
- Period comparison outside tasks: as we found there was some interaction outside the time scheduled for carrying out the tasks, we also included this comparison to determine if the LHTS had any effect outside the time allocated for the study. Essentially, this comparison refers to the use of the MOVING platform by participants in their free time (i.e. unrequested in the study).
  - A period 1 vs A period 2
  - B period 1 vs B period 2

Engagement and use of functionalities We employed the Wilcoxon text (Wilcoxon, 1945) on the metrics extracted for engagement and the use of functionalities for each of the comparisons mentioned above. After computing the p-values, we used Bonferroni correction (Benjamini & Hochberg, 1995) to prevent Type 1 errors (false positives) arising from carrying out multiple comparisons. The full list of results can be seen in Table 8.

The task order comparisons show that there is no effect caused by the randomisation of the search topic. Therefore, the subgroups can be merged, and comparisons between group A and group B are possible (i.e. A1 and A2 are considered as a single group A). Unfortunately, none of the engagement nor use of functionalities metrics indicated any effect from the LHTS widget.

N-gram analysis. We compared the effect of introducing the LHTS on the interactive behaviour of those users who (i) had already been exposed to the MOVING platform (A1 vs **B2**); (ii) had no exposure to the LHTS nor to the MOVING platform before (**A1** vs B1); (iii) were already familiar with the MOVING platform (B1 vs **B2**); (iv) used it for an extended period of time (A1 vs **A2**); (v) had exposure to the MOVING platform but not the LHTS (A2 vs **B2**). Next we provide a guide to interpret Table 9, where coefficients around 0.4 and above are considered to be moderate correlations, and those above 0.6 are strong correlations:

- A high Kendall τ (KENDALL, 1938) and Spearman ρ (Spearman, 1904) correlation indicates that the rankings of two vectors are similar. The former is considered more strict and will typically produce a lower correlation coefficient. When in doubt, the p-value of Kendall's test is known to be more reliable.

– A high Pearson (Pearson, 1895) r suggests that the frequencies of the n-grams are associated (despite their ranking in their respective vectors).

Based on the above guide and the results obtained (see Table 9 and also Figure 37 for the corresponding scatter charts) our findings suggest that:

– **A1 vs B1**: a high Pearson correlation and low Spearman suggest that behaviours are exhibited a proportionately similar number of times but their rankings are not the same (i.e. the frequency based order changes).

– **A1 vs B2**: low correlations tending toward moderate correlations indicate slightly different behaviours on first exposure to the LHTS, which suggests that experience with the MOVING platform makes a difference.

– **B1 vs B2**: high correlations that are consistent across rankings and frequencies suggest that there was not a behaviour change when the LHTS was introduced.

– **A1 vs A2**: low correlations across the tests we run indicate that behaviours changed over time probably due to the learning effect and higher exposure to the platform.

– **A2 vs B2**: low correlations suggest different behaviours between those who have been exposed equally to the platform but get the LHTS incorporated later.

The conclusion derived from these findings suggests that the LHTS does not make users to exhibit *new* behaviours, but makes users prioritise other behaviours that are already in their repertoire (A1 vs B1). The effect of the LHTS is particularly significant for those who interact with the MOVING platform for the first time as once users get familiar with MOVING (B1 vs B2), the posterior incorporation of the LHTS makes no effect in their use of the platform. This means that users have become confident with the way in which they use the platform, are reluctant to change and that support for training is adopted when the learning gap is perceived to be large. For those who have the LHTS from the beginning, a change of behaviour is observed over time, probably due to the acquisition of compentences facilitated initially by the LHTS (A1 vs A2), that boosted the learning effect (in terms of the number of behaviours exhibited) at the outset (A1 vs B2), although this learning would happen anyhow if the period of exposure to the platform was longer. We do not long how long it would take to make the two groups simialar as one week does not seem to be enough time (A2 vs B2).

Table 5: LHTS user group vs period: correlations of top-100 n-grams, where N=4.

	Kendall τ	p value	Pearson r	p value	Spearman ρ	p value
A1 vs B1	0.16	0.02	0.62	0.00	0.27	0.007
A1 vs A2	0.08	0.28	0.23	0.02	0.11	0.27
A1 vs B2	0.21	0.005	0.38	0.00	0.27	0.006
B1 vs A2[15]	0.11	0.15	0.17	0.08	0.14	0.17
B1 vs B2	0.39	0.00	0.58	0.00	0.50	0.00
A2 vs B2	0.15	0.06	0.10	0.32	0.12	0.07

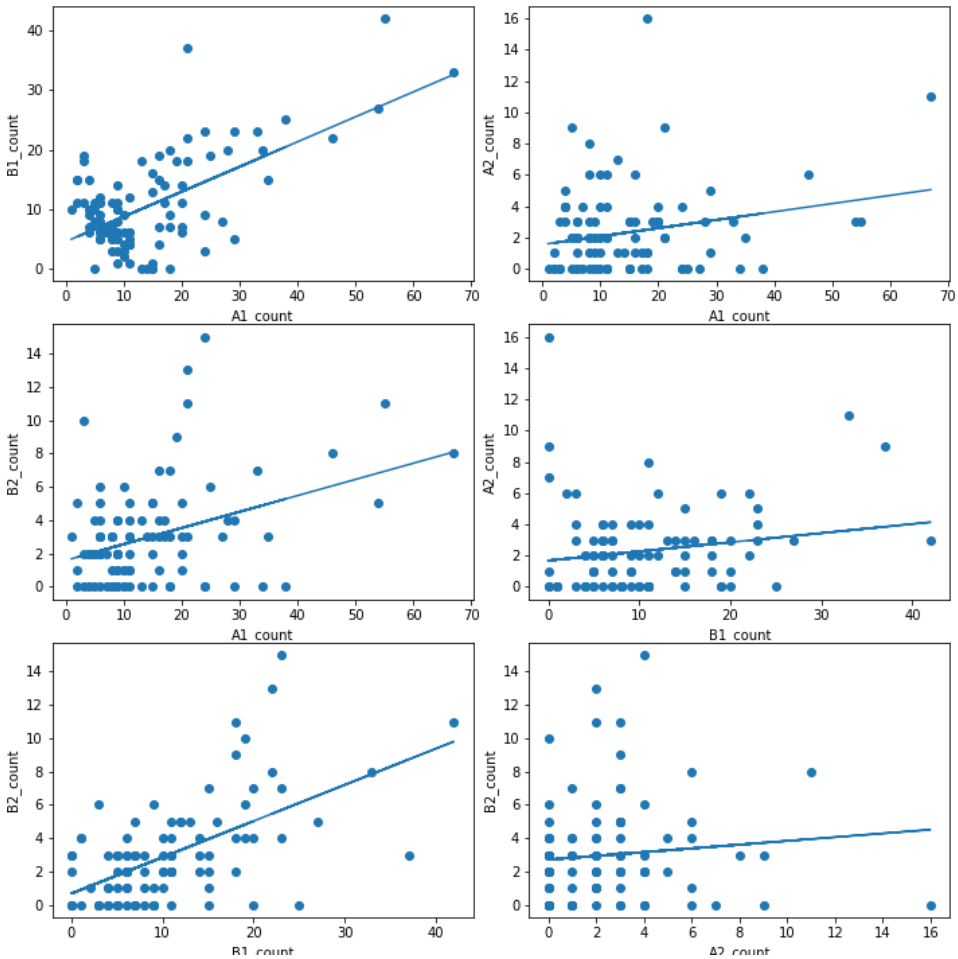
Usage patterns The manual observation of these emerging behaviours allowed us to determine the most common interactions and find differences between the various groups. Regarding **task order**, slight differences regarding the use of features can be seen between the groups. In the case of group A, participants that started with the task 'Big Data' used the concept graph and the top-concepts, while participants who started with task 'Global Warming' preferred tag cloud and pagination. Participants in group B showed similar behaviour, favouring concept-graph and urank for task 'Big Data' and tag cloud and concept-graph for task 'Global Warming'. Task 'Big Data' participants also preferred to use the date filter instead of subjects.

These results are likely to be biased due to the differing amount of interaction between the periods. As period 1 contained more interaction, behaviours exhibited during this period are accentuated.

When comparing the **first and second period**, in the case of the group A, participants stopped using the visualisations and filters they did not like. They also carried out fewer searches, as they probably became more efficient, and wanted to spend less time on the page. Most of the interaction focused on the visualisations.

[15] B1 vs. A2 is added for completeness reasons but the comparison may not be meaningful.

Figure 37: Scatter Charts.



Similar behaviour could be seen from group B participants, using fewer features and filters. During the first period, there was some interaction with the right sidebar, which was empty, and stopped altogether when the LHTS was shown for the second period. Similarly to group A, there were fewer iterative searches and less interaction.

Questionnaires After the first week, we asked our participants to complete an **in-between questionnaire**, consisting of questions regarding the motivation to use the search platform. Additionally we posed two open questions about first experiences and impressions with the platform.

The participants slightly agreed that the platform visualises relevant information to reconstruct the search behaviour ($M = 3.53$, $SD = 0.74$), while the platform motivates participant to think about their own search behaviour ($M = 3.40$, $SD = 0.74$) or to change their typical search behaviour ($M = 3.20$, $SD = 0.86$) were both rated as rather neutral. In contrast, the participants agreed that the platform motivates them to try out different search functionalities and visualisations ($M = 4.33$, $SD = 0.72$) and other functionalities like advanced search or the available filters ($M = 4.40$, $SD = 0.51$). The participants slightly agreed that the platform raised the engagement on the usage of the platform's functionalities, interfaces and visualisations ($M = 3.67$, $SD = 0.72$). Table 6 summarizes the mean values and standard deviation for all participants as well as the four individual groups. A conducted Kruskal-Wallis Test (Kruskal & Wallis, 1952) did not show any significant differences between the groups.

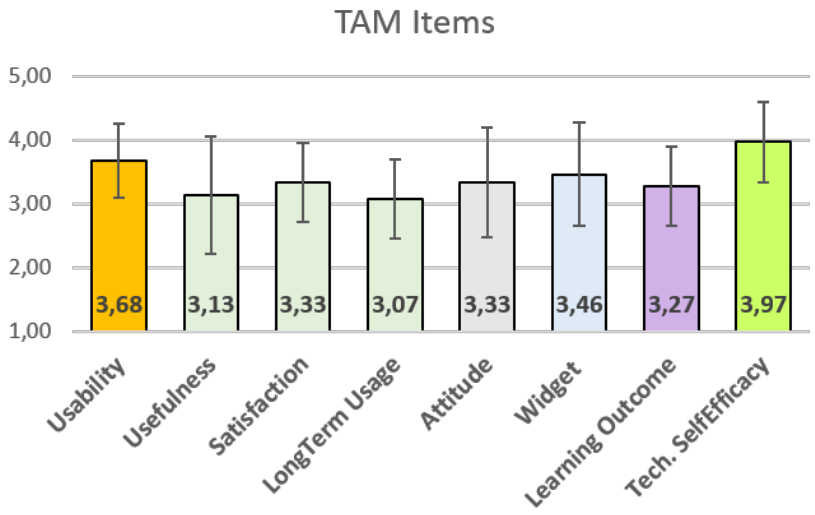
Table 6: Results of the in-between questionnaire for all users and per user group

Item	All	A1	A2	B1	B2
The platform visualizes relevant information to reconstruct my search behaviour.	$M = 3.53$ $(SD = 0.74)$	$M = 2.67$ $(SD = 0.58)$	$M = 4$ $(SD = 0)$	$M = 3.5$ $(SD = 1)$	$M = 3.75$ $(SD = 0.5)$
The platform motivates me to think about my own search behaviour.	$M = 3.40$ $(SD = 0.74)$	$M = 3.33$ $(SD = 1.15)$	$M = 3.5$ $(SD = 0.58)$	$M = 3$ $(SD = 0.82)$	$M = 3.75$ $(SD = 0.5)$
The platform motivates me to change my typical search behaviour.	$M = 3.20$ $(SD = 0.86)$	$M = 3.33$ $(SD = 1.53)$	$M = 3.75$ $(SD = 0.5)$	$M = 2.5$ $(SD = 0.58)$	$M = 3.25$ $(SD = 0.5)$
The platform motivates me to try out different search interfaces and visualisations.	$M = 4.33$ $(SD = 0.72)$	$M = 4$ $(SD = 0)$	$M = 4.25$ $(SD = 0.96)$	$M = 4.5$ $(SD = 0.58)$	$M = 4.5$ $(SD = 1)$
The platform motivates me to try out other functionalities like the advanced search or the available filters.	$M = 4.40$ $(SD = 0.51)$	$M = 4.33$ $(SD = 0.58)$	$M = 4.25$ $(SD = 0.5)$	$M = 4.5$ $(SD = 0.58)$	$M = 4.5$ $(SD = 0.58)$
The platform raises my engagement on the usage of the platform's functionalities, interfaces and visualisations.	$M = 3.67$ $(SD = 0.72)$	$M = 3.67$ $(SD = 0.58)$	$M = 3.5$ $(SD = 0.58)$	$M = 3.25$ $(SD = 0.96)$	$M = 4.25$ $(SD = 0.5)$

After the two weeks, the participants filled in a **post-questionnaire**. We structured the post-questionnaire along the Technology Acceptance Model (Park et al., 2009) using items on self-efficacy, subjective norm, system accessibility, perceived usefulness, perceived ease of use, attitude and behavioural intention to use such a widget. All items of the questionnaire were rated on a 5-point Likert Scale (1 - strongly disagree, 5 - strongly agree). Fig. 38 presents the average ratings of the different items. At a first glance, one can see that all items are rated as neutral or as agreed to.

Regarding the usability of the widget, the participants agreed that the widget was easy and fun to use, easy to understand and that they liked its look and feel with $M = 3.68$ $(SD = 0.58)$ (see bar "Usability" in Fig. 38). The answer to the questions given with respect to the usefulness of the widget, namely that the widget would be useful to explore different search functionalities and increase the search performance and productivity, this was slightly agreed to with $M = 3.13$ $(SD = 0.92)$ (see bar "Usefulness" in Fig. 38). Also the participants slightly agreed to being satisfied with the widget with $M = 3.33$ $(SD = 0.62)$ (see bar "Satisfaction" inFig. 38). Again the participants slightly agreed that they would like to use the widget in the future $M = 3.07$ $(SD = 0.62)$ (see bar "LongTerm Usage" in Fig. 38).

Figure 38: TAM Items: Overview of the average ratings per item



In addition, we posed the participants with the following question: "Do you have further comments for us regarding the usability, usefulness or satisfaction of the widget?". Two of the participants in Group A1 did not find the widget useful as stated by *"[...] For me using the widget didn't make much of a difference. The system's bunch of functions is easy enough to overlook, so you rather quickly find what helps you search best and what not with or without the widget."* and *"I didn't understand how the value for Simple Search is calculated, as I started almost every search with simple search, but the value 1 was always shown for it in the widget.[...]"*. In contrast, one participant of group A2 found the widget quite useful *"The widget is quite useful. I like the design and that it helps me to use the search engine more efficiently."* and two of the same group gave some suggestions for improvement, like for example*"Those are search widgets, so the quality of the underlying knowledge base directly affects usage satisfaction. Since the quality of search results was not optimal in some use cases, I really think that boosing the knowledge base quality would directly boost widget usage and satisfaction."*. Again in group B2 one participant liked the widget *"I liked the widget as I found it was really helpful to be able to visualise my search behaviour.[...]"* while another participant stated that *"[...] the widget, although intuitive and usable, did not significantly change my interaction behaviour with the MOVING platform."*. No participant of group B1 answered this question.

Furthermore, the participants slightly agreed that the widget is a good idea and that they are in favour of using the widget for their own search activities with $M = 3.33 (SD = 0.86)$ (see bar "Attitude" in Fig. 38).

Also the questions about whether the widget motivates users to think about the typical search behaviour, to change their own search behaviour or to try out different search interfaces or visualisations was slightly agreed to with $M = 3.46$ $(SD = 0.81)$ (see bar "Widget" in Fig. 38). Questions on the learning outcomes namely if the participants made a decision of how to perform further searches or if they got a deeper understanding of their search behaviour was also slightly agreed to with $M = 3.27$ $(SD = 0.62)$ (see bar "Learning Outcomes" in Fig. 38).
Afterwards we asked the participants the following open question regarding the learning outcome: *"Do you have any further comments for us regarding your learning outcome with help of the widget?"*. Only one participant from group A1 answered this question by stating *"I didn't learn from using the widget - it just made me more aware of how I'm usually doing my search without wanting to change that behaviour.*

Regarding the self-estimation about the participants' own technological self-efficacy, the participants agreed with $M = 3.97$ $(SD = 0.63)$ that they felt confident in using a software like the MOVING platform and the widget and that they have the necessary skills to use it (see bar "Tech. SeflEfficacy" in Fig. 38).

Additionally we asked the question: *What further features would you like to have in the widget?*". We get two very sophisticated answers, one from a participant in group A1 stating that *"[...]more in-depth information, which you probably can only gain with more complex machine learning or analytics algorithms, about search*

behaviour would be more interesting than just a superficial overview of what features you use the most. I mean real information about the whole process of how you search presented in a way everyone can understand." and one from group A2 *"a kind of recommendation which search feature fits best for me and my search request"*. Again a request to enhance the widget with the faceted search was stated by a participant from group B2 *"As mentioned previously, I think it would be good to have a feature to say how many times you have applied a filter on the search results."*

While we summarized the items of the overall evaluation topics, it is worth having a closer look on the single items of the search behaviour as depicted in Table 7.

While group A2 slightly agreed to use the widget in the future ($M = 3.25,\ SD = 0.5$), the participants of all other groups slightly disagreed to user the widget in the future (A1: $M = 2.67$ $(SD = 0.58)$, B1: $M = 2.75$ $(SD = 0.96)$, B2: $M = 2.5$ $(SD = 1)$. Interestingly the standard deviation for group B1 and B2 is around 1 which is rather high. Participants in groups A2 ($M = 3.5,\ SD = 0.58$), B1 and B2 (both $M = 3.25,\ SD = 96$) slightly agreed to have made a conscious decision to change their search behaviour, while the participants of group A1 slightly disagreed ($M = 2.67,\ SD = 0.58$). These answers given are also in line with the next question where group A2 ($M = 2.25,\ SD = 0.5$), B1 ($M = 2.5,\ SD = 1$) and B2 ($M = 2.75,\ SD = 0.96$) slightly disagreed with stay with their former search behaviour, while group A1 ($M = 3,\ SD = 0$) rated this as neutral. Again, group A2 ($M = 3.75,\ SD = 0.5$), B1 and B2 both with ($M = 3.5,\ SD = 0.58$) agreed that they are confident to have taken the right decision for themselves while again group A1 ($M = 3,\ SD = 0$) rated this question neutral. Regarding the motivation to change their search behaviour, group A2 ($M = 3.5,\ SD = 0.58$) and B2 ($M = 3.25,\ SD = 1.5$) slightly agreed to be motivated to change it, while group A1 ($M = 3.07,\ SD = 088$) rated their motivation neutral which corresponds with the answers given before. In contrast, B1 slightly disagreed ($M = 2.75,\ SD = 0.5$) to be motivated to change their search behaviour, which is not in line with answers given before. While Group A1 ($M = 3.67,\ SD = 0.58$), A2 and B1 ($M = 3.25,\ SD = 0.96$) slightly agreed to be confident to change their search behaviour, B2 ($M = 3,\ SD = 0.82$) rated it as neutral. To the last question asking if the participants would recommend the widget to friends and colleagues, group A2 ($M = 3.75,\ SD = 0.5$) and B1 ($M = 3.5,\ SD = 1$) agreed to it, while group B2 ($M = 2.75,\ SD = 0.96$) slightly disagreed to it and group A1 ($MD = 2,\ SD = 1$) disagreed to recommend it. Again, the standard deviations of 1 (except group A1) are rather high. However, a conducted Kruskal-Wallis Test (Kruskal & Wallis, 1952) did not show any significant differences between the groups.

We also added an open question asking "Could you quickly describe how you plan to change your search behaviour?", however, we received no answers.

Table 7: Results of the single items related to the search behaviour

Item	All	A1	A2	B1	B2
I intend to use the widget in the future.	$M = 2.80$ $SD = 0.77$	$M = 2.67$ $SD = 0.58$	$M = 3.25$ $SD = 0.5$	$M = 2.75$ $SD = 0.96$	$M = 2.5$ $SD = 1$
I made a conscious decision to change my search behaviour.	$M = 3.20$ $SD = 0.94$	$M = 2.67$ $SD = 1.53$	$M = 3.5$ $SD = 0.58$	$M = 3.25$ $SD = 0.96$	$M = 3.25$ $SD = 0.96$
I made a conscious decision to stay with my former search behaviour.	$M = 2.60$ $SD = 0.74$	$M = 3$ $SD = 0$	$M = 2.25$ $SD = 0.5$	$M = 2.5$ $SD = 1$	$M = 2.75$ $SD = 0.96$
I am confident that the decision I made on my search behaviour is the right one.	$M = 3.47$ $SD = 0.52$	$M = 3$ $SD = 0$	$M = 3.75$ $SD = 0.5$	$M = 3.5$ $SD = 0.58$	$M = 3.5$ $SD = 0.58$
I am motivated to change my search behaviour.	$M = 3.07$ $SD = 0.88$	$M = 2.67$ $SD = 0.58$	$M = 3.50$ $SD = 0.58$	$M = 2.75$ $SD = 0.5$	$M = 3.25$ $SD = 1.5$
I am confident that I can change my search behaviour.	$M = 3.27$ $SD = 0.8$	$M = 3.67$ $SD = 0.58$	$M = 3.25$ $SD = 0.96$	$M = 3.25$ $SD = 0.96$	$M = 3$ $SD = 0.82$
I will recommend the widget to my friends and colleagues.	$M = 3.07$ $SD = 1.03$	$M = 2$ $SD = 1$	$M = 3.75$ $SD = 0.5$	$M = 3.5$ $SD = 1$	$M = 2.75$ $SD = 0.96$

Finally, we gave our participants the opportunity to provide further general feedback by asking *"If you have any further comments, please write them below."* We again received only three answers from three different groups. A participant in group A1 stated *"[...]For the widget alone, it might be interesting to novices, for me the different filters & visualizations as such would be of much greater interest.[...]"* and suggestions for improvement, like a *"better search"*, *"interactive tutorials for uncommon visualisations* and *"visualisations should work"*. The other two questions were general with respect to to the MOVING platform and are reported

in the paragraph about the feedback below.

Although we did several further analyses to find significant differences between the different groups nothing was detected so far. Therefore, we decided to look for overall correlations over all topics using Spearman ρ (Spearman, 1904). The results show that there is a significant correlation between the usefulness and the attitude towards the widget ($r_s = .546,\ p < .05$) as well as strong significant correlation between the usefulness and the learning outcome ($r_s = .695,\ p < .01$). Strong significant correlations were found between the usefulness and longterm usage ($r_s = .704,\ p < .01$), the attitude towards the widget ($r_s = .814,\ p < .01$) as well as the widget specific questions ($r_s = .684,\ p < .01$) and the search behaviour ($r_s = .603,\ p < .05$). Correlations were also found between the longterm usage and the attitude towards the widget ($r_s = .808,\ p < .01$), the widget specific questions ($r_s = .630,\ p < .05$) and the search behaviour ($r_s = .718,\ p < .01$). The attitude towards the widget significantly correlated with the widget specific questions ($r_s = .810,\ p < .01$) and the search behaviour ($r_s = .535,\ p < .05$). Finally the widget specific questions correlated also with the search behaviour ($r_s = .556,\ p < .05$)

Feedback The open questions that we posed in the questionnaires aimed at getting deeper insights about the widget as well as the MOVING platform.

In the in-between questionnaire, we posed two open questions: On the first open question *"What are your first impressions about the MOVING platform?"* we got positive answers like for example *"My first impressions in the MOVING platform are positive, as I found the platform's interface easy to use and intuitive"* or *"It visualizes a great amount of data and gives a great choice of different searching mechanisms (like Graph etc.)"*. Furthermore, we also received feedback providing us with suggestions for improvement, like for example *"Nice layout, lovely visualizations, although looking for something specific (like in your tasks) was quite tricky as you mostly just get the abstracts and not a lot of information."* or *"Visually appealing, but the underlying database makes efficient research really difficult."* or *"Seems like a good idea, but the search engine is bad."*

Also for the second open question *"How would you describe your first experiences on the MOVING platform?"* we received some mixed results. On the one hand, the participants gave us positive feedback like *"After a first search however, browsing search results through the platform's features is very intuitive even without experience with the platform."* or *"Mostly good, the interface is easy to navigate, and most of the filters are intuitive to use.[..]"*. On the other hand, we also received some feedback which helps us to improve the platform like for example *"At the beginning it is not really user friendly, but after a couple of times I understood how to use it"* or *"Very frustrating, because I could not find a decent article describing big data"*.

To the open question *"If you have any further comments, please write them below."* that we posed at the end of the post-questionnaire, we got the following two more general answers. One participant from group B1 complained *"The search engine is not very good, it displays dead links. Also, it displays the same link multiple times."*. In contrast, one participant from group B2 was happy with the platform stating *"The widget and MOVING itself are great to use and like I mentioned in a previous Mail, MOVING helps a lof if you do an academic work or a publication for example. It visualises a great amount of data and supports you by giving relevant data and different search options."*.

5.3.6 Discussion & Research Questions

We have evaluated the widget according the technology acceptance model (TAM) (Park et al., 2009). In general we showed that the widget is easy to use and is understandable, that it is somehow useful, that they liked it and that the widget is a good idea. Furthermore - as this is the aim of the widget - the participants started to think about their search behaviour and it made the participants aware of other functionalities that are available on the platfrom. Analysing the results according the model of Kirkpatrick (Kirkpatrick & Kirkpatrick, 2006), we found that the participants would accept the widget (Level 1: Reaction) and did at least become aware of their search behaviour (Level 2: Learning). Whether we could achieve a behaviour change, which corresponds to the third level of Kirkpatrick (Level 3: Behaviour), will next be discussed by answering our research questions below.

RQ1: Engagement metrics: Did the widget lead to significant changes in the engagement of users? From the in-between questionniare, we have evidence that the users of groups A1, A2 and B2 at least started to think about their search behaviour and that the platform motivates changes to the search behaviour. Furthermore, all groups slightly confirmed that the widget raises the engagement on the usage of the platform's functionalities, interfaces and visualisations. This was also confirmed by the results of the post-questionnaire.

On the one hand, these results are promising as this is the aim of the widget. On the other hand these results have to be put in question as group B1 and B2 have not seen the widget in the first week. From the post-questionnaire, we could show that in groups A2, B1, and B2 the widget led to a conscious decision to change the search behaviour and that especially group A2 and B2 are at the same time motivated to change their search behaviour.

Unfortunately, none of the engagement metrics employed for the analysis of the interaction data indicated any effect from the LHTS widget.

RQ2: Navigational patterns: Did the widget lead to different navigational behaviours? The **n-gram analysis** suggests that the LHTS has a significant effect on participants who interact with the platform for the first time. However, incorporating the widget once the participant has become accustomed to the platform had no effect. This effect facilitated the acquisition of competences and boosted the learning effect (in terms of the number of behaviours exhibited). A comparison of the **usage patterns** also showed that participants narrowed down which features to use over time (visualisations and filters), carried out fewer searches, possibly due to increased efficiency, and spent less time on the page.

RQ3: Use of functionalities: Did the widget lead to different use of functionalities? The results of the in-between questionnaire showed the participants of all groups are motivated to try out different search interfaces and other functionalities. Again these results are very promising as they indicate that not only the widget but also the MOVING platform have the potential to motivate people to try out other functionalities.

Unfortunately these results could not be supported with the analysis of the interaction data, where no significant effect was found for the use of functionalities.

5.4 Conclusions

Carrying out studies for each main aspect of the MOVING platform allowed us to adapt our approach to their nature. In the case of the **Working environment**, a large pool of visitors interacted freely with platform. We categorised these users according to their behaviours. A large percentage of users ($\sim$45%) were not active, and did not revisit the web site frequently. However, the rest of the visitors engaged with the platform, and can be classified according to their repeated use of advanced search features. Some of these active users just carried out searches ($\sim$5%), most of them included the use of the filters in their searches ($\sim$42%). Finally, some particularly active users ($\sim$8%) combined the use of filters with visualisations, and scrolling actions on the results page, indicating a higher level of involvement with the platform.

The **MOOC** allowed us to focus the analysis on each of the modules from the course. Users who enrolled in the course were expected to come back every week for a new module. We compared their engagement with the various timely released modules, and looked at engagement trends between the behaviour-based clusters of users. In general the engagement decreased as the MOOC progressed, showing a typical attrition rate. Similar to the results of the Working Environment, we identified groups of users who engaged more with the platform, and in particular with the forum (which was needed to carry out MOOC tasks).

As the Working Environment and the MOOC visitors were not instructed to take part in this study, the results of these studies have been promising. Similarly to the usually high "bounce rate" in web analytics reports, a high percentage of inactive users were expected. However, the majority of the users kept revisiting and interacting with the platform.

The **Learning-how-to-search widget** study was different to the previous two studies, as users were instructed to come back to the MOVING site, and were given tasks to carry out. Three research questions guided the study, to test if the LHTS had an effect in participants' engagement, navigational patterns, or use of functionalities. Questionnaire data indicated a clear effect on engagement and use of functionalities. The LHTS encouraged the users to reflect upon their use of the functionalities, and motivated them to try different ones. However, the metrics employed to analyse this effect in the interaction data failed to support this effect. Regarding the navigational patterns, the LHTS had a significant effect when the widget was available the first time the participants accessed the platform, increasing the acquisition of competences and boosting the learning effect. However, such effect was not significant if users were already familiar with the platform.

6 Conclusions

This deliverable starts by presenting the final status of the implementation of the main components of the MOVING platform. The final list of requirements, including the recommendations provided in D1.3 (Apaolaza et al., 2018) is also provided, indicating the status of their implementation.

This document also presents the methodology followed to annotate the UI elements of the MOVING platform. This annotation made the extraction of relevant interaction data possible, supporting remote user studies and the personalisation of widgets. Examples of how this information has been documented and its application are given.

Before presenting the results of the studies, the analysis pipelines employed to process the interaction data are presented. These pipelines allowed us to extract engagement metrics, as well as to classify MOVING users according to their interaction behaviour.

We carried out three different studies, each of them focused on a different aspect of the MOVING platform. The working environment study focuses on the search pages and presents a data-driven analysis of visitors' behaviour. Four different groups were identified, who showed different search strategies, from just using search to combining it with filters, visualisations and scroll interaction. In the case of the MOOC study, a more content-oriented approach could be taken. Each week a new module is released, so we explored how users' engagement evolved as they progressed in the course. Most users exhibited a typical attrition rate, where they would be less engaged with the content over time. The Learning-how-to-search widget study explored the effect the widget had on users. The results were positive, as we discovered that users felt encouraged to try new visualisations and to reflect on their behaviour. The effect the widget has on users was found to be stronger if it was shown the first time the users interacted with the platform.

References

Agrawal, R., Srikant, R., & others. (1994). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487–499).

Apaolaza, A., Bienia, I., Maas, A., Wiese, M., Günther, F., Barthold, S., & Vigo, M. (2018). *Deliverable 1.3: Initial evaluation, updated requirements and specifications* (Tech. Rep.). MOVING.

Apaolaza, A., Harper, S., & Jay, C. (2013). Understanding users in the wild. In *Proceedings of the 10th international cross-disciplinary conference on web accessibility* (pp. 13:1–13:4). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2461121.2461133 doi: 10.1145/2461121.2461133

Apaolaza, A., & Vigo, M. (2017, June). Wevquery: Testing hypotheses about web interaction patterns. *Proc. ACM Hum.-Comput. Interact.*, *1*(EICS), 4:1–4:17. Retrieved from http://doi.acm.org/10.1145/3095806 doi: 10.1145/3095806

Aula, A., & Nordhausen, K. (2006). Modeling successful performance in web searching. *Journal of the american society for information science and technology*, *57*(12), 1678–1693.

Barthold, S., Bienia, I., Blume, T., Gottfried, S., Günther, F., Fessl, A., ... Wiese, M. (2017). *Deliverable 1.2: Initial implementation of user studies* (Tech. Rep.). MOVING.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. Retrieved from http://www.jstor.org/stable/2346101

Bienia, I., Fessl, A., Günther, F., Herbst, S., Maas, A., & Wiese, M. (2017). *Deliverable 1.1: User requirements and specification of the use cases* (Tech. Rep.). MOVING.

Blume, T., Böschen, F., Galke, L., Saleh, A., Scherp, A., Schulte-Althoff, M., ... AG, S. H. (2018). *Deliverable 3.1: Technologies for moving data processing and visualisation v1.0* (Tech. Rep.). MOVING.

Collyda, C., Mezaris, V., Herbst, S., Grunewaldm, P., Köhler, T., Fessl, A., ... Skulimowski, A. (2017). *Deliverable 6.2: Data management plan* (Tech. Rep.). MOVING.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319–340.

Fessl, A., Thalmann, S., Pammer-Schindler, V., Saleh, A., Nishioka, C., Scherp, A., ... and, F. G. (2019). *Deliverable 2.1: Initial conceptual framework, curricula and technical prototypes for adaptive training support* (Tech. Rep.). MOVING.

Gottfried, S., Grunewald, P., Pournaras, A., Collyda, C., Fessl, A., Hasitschka, P., ... Scherp, A. (2017). *Deliverable 4.1: Definition of platform architecture and software development configuration* (Tech. Rep.). MOVING.

Gottfried, S., Pournaras, A., Collyda, C., Mezaris, V., Backes, T., Wertner, A., ... Saleh, A. (2017). *Deliverable 4.2: Initial responsive platform prototype, modules and common communication protocol* (Tech. Rep.). MOVING.

Gunther, F., Barthold, S., Herbst, S., Zawidzki, J., Helbig, A., Bienia, I., ... Zdolsek, T. (2019). *Deliverable 2.3: Final conceptual framework, curricula and moving mooc for community building* (Tech. Rep.). MOVING.

Günther, F., Barthold, S., Bienia, I., Maas, A., Wiese, M., Fessl, A., ... Draksler, T. Z. (2019). *Deliverable 2.2: Updated curricula and prototypes for adaptive training support and introductory moving mooc for community building* (Tech. Rep.). MOVING.

Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., & Hsu, M. (2001). Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *proceedings of the 17th international conference on data engineering* (pp. 215–224).

KENDALL, M. G. (1938, 06). A NEW MEASURE OF RANK CORRELATION. *Biometrika*, *30*(1-2), 81-93. Retrieved from https://dx.doi.org/10.1093/biomet/30.1-2.81 doi: 10.1093/biomet/30.1-2.81

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating trainings programs: The four levels. 3rd ed.* Berrett-Koehler Publ., Inc.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583-621. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/01621459.1952.10483441 doi: 10.1080/01621459.1952.10483441

Lagun, D., & Lalmas, M. (2016). Understanding user attention and engagement in online news reading. In *Proceedings of the ninth acm international conference on web search and data mining* (pp. 113–122).

Lalmas, M., O'Brien, H., & Yom-Tov, E. (2014). Measuring user engagement. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, *6*(4), 1–132.

Lorenz, R., Pournaras, A., Collyda, C., Mezaris, V., Zielinski, A., Šimić, I., ... Vagliano, I. (2017). *Deliverable 4.3: Final responsive platform prototype, modules and common communication protocol* (Tech. Rep.). MOVING.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55–60).

Nishioka, C., & Scherp, A. (2016). Profiling vs. time vs. content: What does matter for top-k publication recommendation based on twitter profiles? In *2016 ieee/acm joint conference on digital libraries (jcdl)* (pp. 171–180).

Park, S. Y., et al. (2009). An analysis of the technology acceptance model in understanding university students' behavioral intention to use e-learning. *Educational technology & society*, *12*(3), 150–162.

Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*, 240–242. Retrieved from `http://www.jstor.org/stable/115794`

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. Retrieved from `http://www.jstor.org/stable/1412159`

Vagliano, I., Abdel-Qader, M., Blume, T., Böschen, F., Galke, L., Saleh, A., ... Mutschke, P. (2018). *Deliverable 3.2: Technologies for moving data processing and visualisation v2.0* (Tech. Rep.). MOVING.

Vagliano, I., Blume, T., Galke, L., Mai, F., Saleh, A., Pournaras, A., ... Mutschke, P. (2019). *Deliverable 3.3: Technologies for moving data processing and visualisation v3.0* (Tech. Rep.). MOVING.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186–204.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83. Retrieved from `http://www.jstor.org/stable/3001968`

A APPENDIX

A.1 Learning-how-to-search widget engagement comparisons

Table 8: P-values resulting from the use of the Wilcoxon test followed by a Bonferroni correction.

	Engagement										Use of functionalities										
	activeTimeEpi	awayTimeEpi	totalTimeEpi	activeTimeTab	awayTimeTab	totalTimeTab	mousedown_result-item	episodeCount	searchCount	scrollCount	mousedown_search-tab-top-sources	mousedown_search-tab-urank	mousedown_search-tab-top-concepts	mousedown_search-tab-tag-cloud	mousedown_search-tab-results	mousedown_search-tab-concept-graph	mousedown_ats_right	mousedown_search_advanced_link	mousedown_search_simple_link	mousedown_search-button	mousedown_q
Group A: task order	0.89	0.74	0.74	0.89	0.74	0.74	0.78	0.74	0.74	0.80	0.89	0.67	0.74	0.74	0.74	0.74	0.89	0.70	0.89	0.70	0.70
Group B: task order	0.43	0.16	0.16	0.53	0.16	0.16	0.83	0.22	0.36	0.10	0.76	0.32	0.76	0.16	1.00	0.43	0.43	0.32	0.83	0.32	0.22
Group A: period comparison	0.82	0.86	0.82	0.82	0.82	0.82	0.88	0.88	0.86	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.86	0.82	0.82	0.88	0.86
Group B: period comparison	0.67	0.67	0.67	0.67	0.67	0.67	0.89	0.67	0.67	0.67	0.77	0.67	0.87	0.80	0.67	0.67	0.99	0.89	0.67	0.67	0.67
Group A: time outside tasks period comparison	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
Group B: time outside tasks period comparison	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.98	0.91	0.98	0.91	0.98	0.91	0.98	0.91	0.91	0.91	0.91

A.2 Learning-how-to-search widget evaluation questionnaires

.

A.2.1 Pre-Questionnaire

Participant Information Sheet & Consent Form

In order to be conform with the GDPR (General Data Protection Regulation) and the Data Protection Act (2018), we kindly ask you to read the participant information sheet below, before you give us your consent to participate in the MOVING evaluation.

Approval date: 25th of June 2018

This document gives you all information about the conducted study. Please read this sheet carefully and ask questions about anything that you don't understand or want to know in more detail.

1. *Title of the project*
   MOVING: TrainInG towards a society of data-saVvy inforMation prOfessionals to enable open leadership INnovation (http://moving-project.eu/)

2. *What is this study about? (Project and Study Purpose)*
   We are investigating a widget on the MOVING platform that visualizes usage of different search interfaces on the MOVING platform with the goal to motivate users to explore alternative search interfaces. The underlying rationale is, that by now the one-line-input to search engines is so standard, that we can assume users' search behaviour to be operationalized. This means, that users may be reluctant to explore alternatives even though these may be suited better to their context of use / search task.

   Within this project, we want to conduct the following study:

   Title: Motivating Users to Explore Alternative Search Interfaces

   We have designed and implemented a widget, consisting of two parts, the search behavior visualization and the reflection guidance. The search behavior visualization mirrors back the feature usage of a user on a search platform. The reflection guidance part presents reflection prompts adapted to the user's search history, e.g. how long the user is already on the platform and the search activities conducted. For example, while a new user gets a sentence starter that makes her aware of a new, not used functionality on the platform, an experienced user receives a questions on how a feature might have influenced the own search performance. This part of the widget aims to breakdown operationalized behavior to the level of conscious action again, thereby opening up the possibility for re-shaping own behavior.

3. *Who is running the study? (Investigators)*
   The study is carried out by the following researchers:

   - Angela Fessl, Know-Center GmbH, afessl@know-center.at
   - Viktoria Pammer-Schindler, Know-Center GmbH, vpammer@know-center.at

   and

   - Markel Vigo, University of Manchester, markel.vigo@manchester.ac.uk
   - Aitor Apaolaza, University of Manchester, aitor.apaolaza@manchester.ac.uk

4. *Why have I been invited to participate in the study? (Eligibility)*
   You have been invited to take part in this research study because you are either an employee of the Know-Center GmbH or a student a the Universtity of Manchester.

5. *What would I be asked to do if I take part? (Overall Description of Participation)*
   As participant of this study, you will be asked to

   - fill in a pre-, an in-between and a post-questionnaire,
   - conduct prescribed search tasks on the MOVING platform and using the developed widget accordingly,
   - send emails with the task results to the researchers,
   - participate eventually in an interview in order to get deeper insights on your experience

There are no right or wrong answers, every answer given is important for our research.

6. *What is the duration of the research? (Length of Participation)*
   Altogether, the duration of the evaluation will last from the 25th of June to the 8th of July and lasts about 10-15minutes per day.

7. *What are the risks associated to the study? (Risks of Participation)*
   Aside from providing your time, we do not expect any risks/costs associated with taking part in this study.

8. *What are the benefits associated to the study? (Benefits of Participation)*
   We cannot guarantee or promise that you will receive any direct benefits from being in the study. Your input will be used as for investigating if the developed widget does fulfil its purpose.

9. *Is there any compensation/payment/incentives? (Compensation/Payment/Incentives)*
   No, you will not receive any compensation/payment/incentives.

10. *What happens if I do not want to take part or change my mind? (Volunteer Statement)*
    It is up to you to decide whether or not to take part. If you decide to participate in the study, you may withdraw from it at any time without giving a reason and with detriment to yourself.

11. *Will the collected information about me be kept confidential? (Confidentiality Statement)*
    By providing your consent, you agree that we are collecting personal information about you for the purposes of this research study. This information will only be used for the purposes outlined in this Participant Information Sheet. Your information will be stored securely and your identity/information will be kept strictly confidential. Study findings may be published, but all data for analysis will be anonymised. In reporting on the research findings, we will not reveal the names of any participants. At all times there will be no possibility of you as individuals being linked with the data. Although every effort will be made to protect your identity, there is a risk that your participation (but no individual data) might be identifiable in publications due to the nature of the study and/or the results.

12. *What if something goes wrong? (Formal complaint about the conduct)*
    If you want to make a formal complaint about the conduct of the study, please contact:

    – Angela Fessl, Know-Center GmbH, afessl@know-center.at
    – Aitor Apaolaza, University of Manchester, aitor.apaolaza@manchester.ac.uk

Thank you for reading this information sheet and for considering taking part in this research.
If you are happy to participate please complete the consent form below:

– [CS1] I confirm that I have read the participant information for the above study. [yes/no]
– [CS2] I understand that my participation in the study is voluntary and that I am free to withdraw at any time without giving a reason and without detriment to myself. I understand that it will not be possible to remove my data from the project once it has been anonymised and forms part of the data set. I agree to take part on this basis. [yes/no]
– [CS3] I consent access to the data collected within the questionnaires. [yes/no]
– [CS4] I consent access to the activity log data and data explicitly given within the MOVING platform. [yes/no]
– [CS5] I consent access to the mails send to the researchers submitting the tasks. [yes/no]
– [CS6] I agree that any data collected may be published in anonymous form in academic books, reports or journals. [yes/no]
– [CS7] I agree that the researchers may retain my contact details in order to provide me with a summary of the findings for this study. [yes/no]
– [CS8] I agree to take part in the above evaluation. [yes/no]

Demographics
The possible answer options are presented in the square bracket at the end of each question.

- [D1] Please state your gender [Female/Male]
- [D2] Please enter your age [Only numbers]
- [D3] Please state your profession [Student/Employee/Other] If student:
  - [DS3.1] Please enter the name of your field of study.
  - [DS3.2] Which degree are you aiming at? [Bachelor/Master/PhD/Other]
  - [DS3.3] Enter your year of study.

  If employee
  - [DE3.1] Please enter the domain of your company you are working in:
  - [DE3.2] Please enter your job / position in your company:
  - [DE3.3] Please add the number of years you are working in this position:

  If other
  - [DO3.1] Please describe your profession in detail below:

Background Information

- [BI1] How many years of computer experience do you have?
- [BI2] How many years of web experience do you have?
- [BI3] What is your frequency of using search engines? [Daily/Alsmost Daily/Couple of times a week/Weekly/Less than weekly]
- [BI4] What is your frequency of using computers? [Daily/Alsmost Daily/Couple of times a week/Weekly/Less than weekly]
- [BI5] What is your frequency of using the web? [Daily/Alsmost Daily/Couple of times a week/Weekly/Less than weekly]
- [BI6] How would you evaluate your search skills on a scale from1 (unskilled) to 5 (skilled)? [Very skilled/Skilled/Netural/Unskilled/very unskilled]

A.2.2 In-between Questionnaire

All questions are to be rated on a 5-point Likert scale from 1 = strongly disagree to 5 strongly agree. If an other answer option is used it is stated at the end of the question in square brackets.

- [EX1] The platform visualizes relevant information to reconstruct my search behaviour.
- [EX2] The platform motivates me to think about my own search behaviour.
- [EX3] The platform motivates me to change my typical search behaviour.
- [EX4] The platform motivates me to try out different search interfaces and visualisations (ConceptGraph, uRank, TagCloud, Top Concepts, Top Sources).
- [EX5] The platform motivates me to try out other functionalities like the advanced search or the available filters.
- [EX6] The platform raises my engagement on the usage of the platform's functionalities, interfaces and visualisations.
- [EX7] What are your first impressions about the MOVING platform? [open question]
- [EX8] How would you describe your first experiences on the MOVING platform? [open question]

A.2.3 Post-Questionnaire

All questions are to be rated on a 5-point Likert scale from 1 = strongly disagree to 5 strongly agree. If an other answer option is used it is stated at the end of the question in square brackets.

Usability of the widget

- [US1] It was very easy to use the widget
- [US2] I like the look & feel of the widget.
- [US3] The widget was easy to understand.
- [US4] The widget was fun to use.
- [US5] The user experience with the widget was very comfortable.
- [US6] The widget was free of bugs and errors.

Usefulness, satisfaction, long-term usage of the widget

- [SAT1] I am satisfied to use the widget.
- [UF1] I think the widget is useful for exploring different functionalities on the MOVING platform.
- [UF2] I think using the widget would increase my search performance.
- [UF3] I think using the widget would increase my productivity.
- [LT1] I would like to use the widget continuously as part of my student's/work life.
- [LT2] It is useful for me to continue using the widget in my student's/work life.
- [LT3] It is beneficial for me to continue using the widget in my student's/work life.
- [SUL1] Do you have further comments for us regarding the usability, usefulness or satisfaction of the widget? [open text]

Attitude to the widget

- [AT1] Using the widget is a good idea.
- [AT2] I am in favour of using the widget for my search activities.

Widget Specific Questions

- [WS1] The widget visualizes relevant information to reconstruct my search behaviour.
- [WS2] The widget motivates me to think about my search behaviour.
- [WS3] The widget motivates me to change my typical search behaviour.
- [WS4] The widget motivates me to try out different search interfaces and visualisations (ConceptGraph, uRank, TagCloud, Top Concepts, Top Sources).
- [WS5] The widget motivates me to try out other functionalities like the advanced search or the available filters.
- [WS6] The widget raised my engagement with the usage of the platform's functionality.

Learning Outcomes

- [LO1] I made a conscious decision on how to search in the future.
- [LO2] I did gain a deeper understanding of my search behaviour.
- [LO3] Do you have any further comments for us regarding your learning outcome with help of the widget? [open text]

Search Behaviour

- [SB1] I intend to use the widget in the future.
- [SB2] I made a conscious decision to change my search behaviour.
- [SB3] I made a conscious decision to stay with my former search behaviour.
- [SB4] I am confident that the decision I made on my search behaviour is the right one.
- [SB5] I am motivated to change my search behaviour.
- [SB6] I am confident that I can change my search behaviour.
- [SB7] I will recommend the widget to my friends and colleagues.
- [SB8] Could you quickly describe how you plan to change your search behavior? [open text]

Technological Self-Efficacy

- [TS1] I feel confident in using software like the MOVING platform.
- [TS2] I feel confident in using software like the widget.
- [TS3] I have the necessary skill for using such a widget.
- [TS4] I know quite a lot about the web and search engines.
- [TS5] What further features would you like to have in the widget? [open text]

Technological Self-Efficacy

- [FC1] If you have any further comments, please write them below.[open text]

A.3 Additional interaction documentation

Table 9: ATS interaction annotation

Description	Location	Type	ID	Class
Main container	Inside sidebar_right	div	ats_right	grid_ats
Widget header	Top bar of the ATS	div	ats-icon-bar	panel-heading
Links to select widget curriculum	Inside header	a	curriculum_new	ats-mode-button
Links to select widget learning how to search	Inside header	a	learning-how-to-search	ats-mode-button
Links to select widget overall progress	Inside header	a	curriculum	ats-mode-button
Title	On top of the chart	h5	ats-widget-title	widget-title
ATS content	Container for the ATS chart	div	learning-how-to-search-container	ats
Feature usage bars	Inside ats	div	learning-how-to-search-chart	graphics
Feature usage bars labels	Inside ats, under bars	div	learning-how-to-search-labels	graphics-labels
Text describing the form	Under bars' labels	div	learning-how-to-search-prompt-text	prompt-text
Input form (not in use?)	Under prompt-text	div	learning-how-to-search-prompt-input-form	form-group

Table 10: Overall Progress annotation table

Description	Location	Type	ID	Class
Main container		div	ats_right	grid_ats
Widget header	Top bar of the ATS	div	ats-icon-bar	panel-heading
Links to select widget curriculum	Inside header	a	curriculum_new	ats-mode-button
Links to select widget learning how to search	Inside header	a	learning-how-to-search	ats-mode-button
Links to select widget overall progress	Inside header	a	curriculum	ats-mode-button
Title	On top of the chart	div	ats-widget-title	widget-title
Widget content (including chart and progress)		div	chart	
Nested pie chart	Interactive chart	g	skills-slice-group	
Skill progress	Under interactive chart	div	skillProgressExplanation	skillProgressExplanation
Core progress bar	Under interactive chart	div	coreskillprogress	coreskillprogress
Core skill label	Under interactive chart	div	coreskill	
Subskillprogress bar	Under interactive chart	div	subskillprogress	subskillprogress
Subskillprogress label	Under interactive chart	div	subskill	
Description of skills	Under widget content (#chart)	div	skill-description	
Skill description items	List of items describing progress for each X skill	div	knowledge_assessment_X	current-skill-level-header

Table 11: Curriculum widget annotation table

Description	Location	Type	ID	Class
main container		div	ats_right	grid_ats
widget header	top bar of the ATS	div	ats-icon-bar	panel-heading
Links to select widget curriculum	Inside header	a	curriculum_new	ats-mode-button
Links to select widget learning how to search	Inside header	a	learning-how-to-search	ats-mode-button
Links to select widget overall progress	Inside header	a	curriculum	ats-mode-button
Title	on top of the chart	div	ats-widget-title	widget-title
Curriculum prompt container	directly under title	div	ats-curriculum-prompt-container	prompt-container
Learn more button in prompt	Inside curriculum prompt container	a	ats-curriculum-learn-more	btn-learn-more
Progress bar container	under the prompt container	div	ats-curriculum-progress-bar-container	progress-bar-container
Progress bar title	over progress bar	span	ats-curriculum-progress-label	
Progress bar	under the prompt container	div	ats-curriculum-progress-bar	
Form for the reflective question	Same place as the curriculum container prompt. The two are swapping places, always one is displayed	form	ats-answer-form	
A reflective question	Over the input field in the form	span	ats-curriculum-reflective-question	
The input field for entering the answer to the reflective question	The input field in the curriculum widget	textarea	ats-answer-text	
Submits the answer to the reflective question	under the reflective question input field		ats-curriculum-submit-reflective-question	btn btn-submit

Table 12: Recommender system annotation table

Description	Location	Type	ID	Class
Main container	Under the ATS widget		rs-panel	card
Header	Top of the RS	h4	rs-header	
Container of list of recommendations	Under the header	div	rs-body	card
Go up button	Above the list of recommendations	a	rs-prev-btn	
Go down button	Above the list of recommendations	a	rs-next-btn	
List of recommendations			rs-list	
The div that persists the json data to iterate over		div	rs-json-div	
To be used to associate the collapsable area with the more link	div	rs-accordion		
Reference to the collapsable area		div	rs-collapsable-area	
Recommendation item		div	elasticsearch document ID	card
Items within recommendation items (see below)			inherited from list of recommendations	recommended-item

B Updated Requirements

The following tables provide an update on the state of implementation of the user requirements presented in D1.1. In order to organise the discussions on the integration of requirements between project partners, various Kanban-style boards were built in Trello (for more information please check D1.2, Session 2.1). Here is a short description of what each status description represents:

- To do: The requirement has still to be processed.
- Work in progress: The requirement is currently being processed.
- Implemented: The requirement was processed and successfully integrated into the platform.
- Cancelled: The requirement could not be processed. Reasons for this are given for the respective requirements.
- Backlog: These requirements are considered important, but not feasible during the project period. In the case of a follow-up project, these requirements are taken up again.

In the case of new requirements, e.g. emerging from the findings of ongoing focus group interviews, these are identified by the ID #9XY and they are italicised. Moreover, if the description of the requirement has been changed, this is indicated by the italic font.

B.1 Requirement: search field

Table 13: Requirement: Search field.

ID	Topic	Key word	Description	Status	Reference to use case
#001	Author		*There should be search field where to search for an author, a title or a key word when searching for resources. In case of a search for authors a list of disambiguated author names should be returned, respectively a list of items with disambiguated author names.*	*Implemented*	#TUD040
#002	Title			*Implemented*	
#003	Key word			*Implemented*	#TUD006, #TUD015
#004	Combination of words		It should be possible not only to search for one word but for combination of words. *Compound words (e.g. computer science, data mining) should be recognised as single not separate search term.*	*Implemented*	#TUD001

#005	Including and excluding words		It should be possible to exclude and include words or terms, *e.g., searching for "social NOT workers" or "social –workers" will return different results than searching for "social workers".*	*Implemented*	#TUD002
#008	*Advanced search*	*Multiple search terms*	Define one or multiple search terms, *e.g. searching for data science AND machine learning or searching for data science AND a specific author.*	*Implemented*	#EY002
#010	Search extent	Search depth	Refine the search by title, abstract, full-text.	*Implemented*	#TUD039, #EY008
#018	Auto-complete		A recommendation/auto-completion feature supports the specification of the search term(s). This will support the auditor in: (a) searching for the correct entity and (b) identifying similar (and therefore possibly related) entities right from the beginning of the analysis.	*Implemented*	#EY034
#901 (new)	*Simple search: Search domains*		*Different search domains (research, funding, and learning) should be displayed within the search bar as drop-down menu.*	*Implemented*	TUD

B.2 Requirement: faceted search

Table 14: Requirement: faceted search.

ID	Topic	Key word	Description	Status	Reference to use case
#006	Date		Refining the search after the publishing date of the resource. Specify the relevant date range (e.g., the period after the entity's formation).	*Implemented*	#TUD038, #EY036
#007	Search term		Exclude one or multiple search terms.	*Merged with #005*	#EY001

#009	Media type		Searching in and possibility of excluding monographies, journal articles, open access journal articles, conference articles, posters.	*Implemented*	#TUD035
#012	Datasets		Listing of all databases (sorted by discipline).	*Implemented*	#TUD032
#019	Dynamic Interface		Based on the search settings, the remaining search criteria become enables or disabled (e.g., when a file on the local device is included that has no metadata, the depth of the search cannot be set to metadata).	*Implemented*	#EY011
#020	Geographic Region	Location search	Limit the search to certain geographical areas (e.g., the entity's headquarters country).	*Implemented*	#EY009
#021	Language		Limit the search to specific languages. By default, the language is in correspondence to the search term defined.	*Implemented*	#EY012
#022	Concepts	Concept Type	Ability to limit the search to certain types of concepts, for example, persons (who are not necessarily authors) and organisations only.	*Implemented*	#EY037
#023a	Industry	Industry Type	Limit the search to certain industries, when applicable (e.g., retail, automotive, airlines). This can support the auditor in identifying the relevant laws and regulations.	*Implemented*	#EY049
#023b	Author		Refine the search by author by excluding and including authors.	*Implemented*	#TUD040

#025	Discipline		Refining the search by including or excluding scientific disciplines. Limit the search to certain disciplines, when applicable (e.g., when selecting a database that covers multiple disciplines).	*Implemented*	#TUD041, #TUD050, #EY010
#026	Citations		Refine the search by the amount of citations of the source.	*Cancelled due to technical complexity*	#TUD042, #TUD051
#027	Clicks		Refine the search after the number of clicks on the source through the users.	*Cancelled, not objective enough for the search*	#TUD046
#028	Search template		It should be possible that the user can adjust the faceted search due to his/her needs. E.g., the specific user searches mostly within the social sciences, so he/she saves this for the faceted search.	*Cancelled, not useful, since the facets of the search are determined by the search results*	#TUD004, #TUD081
#088	Timeline visualisation		A timeline visualisation showing how search results appear chronologically.	*Implemented*	#TUD093, #TUD094
#906	Checkboxes		*Search filters should have check boxes so that users can select/exclude multiple items on each filter; there should be a check box "all" or "select all" for every filter*	Implemented	
#907	Limit display of filter options		*Each filter should show only 4 or 5 items at once and have the option "show more" to see all hidden options in a drop-down menu bar; For example, when there are more than 5 different document types, show the 5 categories (books, articles, videos, reports, with the most hits and hide others under "show more".*	Implemented	

#908	Display of filters		*User should have an overview over the selected filters at all times; menu bar with filter selections should be permanently visible to the user while browsing the results list.*	Implemented	
#909	Reset filters		*The user should have the option to reset all the filters "reset search"*	Implemented	TUD

B.3 Requirement: data sources

Table 15: Requirement: data sources.

ID	Topic	Key word	Description	Status	Reference to use case
#011	Datasets Databases	*Literature databases*	Searching and possibility of including or excluding databases (see D1.1). Connection to various literature databases available to include publications of several disciplines.	*Implemented*	#TUD011, #TUD012, #TUD016, #TUD089, #EY007
#013	Datasets Data sources		Include or exclude certain data sources.	*Implemented*	#EY003
#014		World Wide Web	Data sources should include both the world wide web and the possibility to define one or multiple specific websites to include.	*Merged with #030, technically difficult to include the world wide web*	#EY004
#015		Various extension types	Browse for files stored on the local device to include in the analysis. Those files can be of various file formats, e.g., .pdf, .doc, .doc, .rtf, .txt, .xls, .xlsx, .csv, .htm, .html.	*Implemented*	#EY005
#016		Data source upload	The upload of data sources stored on the local device is restricted due to data privacy issues.	*Backlog*	#EY006

#017		Intranet	We should consider enabling the connection of the MOVING platform to files stored in the intranet. However, the tool may not upload those files to the internet at any stage of the analysis.	*Backlog, merged with the upload functionality on the EY branch*	#EY055
#024	Survey data		Searching and possibility of including or excluding survey data (GESIS, Statista[16]).	*Backlog, high effort, and other requirements had priority*	#TUD037
#029	Social media		Searching in and possibility of including or excluding social media (Twitter, Facebook, ResearchGate, academia.eu, XING, LinkedIn, Blogs). This requirement is in connection to #030. The social media websites are added to the crawler list.	*Implemented, Twitter, Google+, Blogs*	#TUD017, #TUD018, #TUD019, #TUD020, #TUD021, #TUD022, #TUD024, #TUD123, #EY035
#030	Websites		Searching in and possibility of including and excluding websites (see D1.1). The searching is made possible through crawling technology (see #029).	*Implemented*	#TUD020, #TUD024, #TUD030, #TUD034
#031	Online Encyclopaedia	Wikipedia	Including and excluding Wikipedia in the search.	*Cancelled, dataset too big*	#TUD025, #TUD026
#032			Comparing Wikipedia articles in different languages.	*Cancelled, see #031*	#TUD075
#033	Website	archive.org	Including and excluding archive.org in the search.	*Cancelled, see #031*	#TUD031
#034	Search engines	Google, Bing, Yahoo	Searching and possibility of including or excluding existing search engines.	*Cancelled, no possibility to get access to these search engines*	#TUD026
#035	Library catalogues		Searching and possibility of including or excluding library catalogues (see D1.1).	*Cancelled, high effort, and other requirements had priority*	#TUD027, #TUD028
#036	PDF files	e.g. Plenary protocols of the Bundestag	Including or excluding PDF files in the search.	*Implemented*	#TUD029

[16] `https://de.statista.com` (2017-03-27)

#037	Project partners	International	It should be possible to look for possible project partners from different countries.	*Implemented*	#TUD007
#038	Business partners	International	It should be possible to look for business partners via the platform.	*Implemented*	#TUD008, #TUD023
#039	Videos		Searching and possibility of including or excluding videos.	*Implemented*	#TUD036
#040	Data source for funding		Possibility of including or excluding funding databases.	*Implemented*	#TUD088
#041	Creative commons	Flickr	Searching in and possibility of excluding or including sources under the licence of creative commons (e.g. Flickr[17]).	*Backlog, priority was on documents*	#TUD113

B.4 Requirement: search list

Table 16: Requirement: search list.

ID	Topic	Key word	Description	Status	Reference to use case
#042	List		Displaying the search results as a list.	*Implemented*	#TUD060, #EY051
#043	Selecting		Picking up the relevant search results by clicking on it.	*Implemented*	#TUD045, #TUD065
#044	Display frequently cited reference per source		Displaying frequently cited references per source through mouse over.	*Cancelled due to technical complexity*	#TUD047
#045	Funding deadline		Displaying the deadline for a funding opportunity and possibility of rearranging the results by deadline.	*Cancelled due to complexity*	#TUD048
#046	Funding contact		Displaying the contact for a funding opportunity when you mouse over.	*Cancelled due to technical complexity*	#TUD048
#047	Duplicates		Avoiding duplicates due to search in different databases.	*Implemented*	#TUD053
#048	Checkbox		Displaying which resources, I already searched/looked at through a checkbox.	*Cancelled due to technical complexity*	#TUD049, #TUD059
#049	Different tabs		Open search results in different tabs.	*Implemented*	#TUD066

[17] `https://www.flickr.com` (2017-03-27)

#050	Bookmark		Bookmark several search results.	*Implemented*	#TUD067
#051	Library connection		Displaying where the source can be found (e.g. in a library) by clicking on it.	*Cancelled, high effort, and other requirements had priority*	#TUD071, #TUD128
#052	Mark results		Marking the results due to high or low relevance on the search topic.	*Cancelled due to technical complexity*	#TUD072
#053	Open access		Displaying whether the source is open access or not.	*Implemented*	#TUD112, #TUD118
#054	Creative commons		Displaying whether the source is under the licence of creative commons or not.	*Cancelled, other requirements had priority*	#TUD113, #TUD118
#055	Author Contact		Displaying the contact details of the author of the resource (especially from frequently occurring authors).	*Cancelled due to technical complexity*	#TUD125, #TUD126
#912	Funding		*Update RSS on Horizon2020 and adding filters*	*Implemented*	
#913	Number of results per page		*User should be able to select number (e.g. 20, 50, 100) of results per page that will be shown at once in results list*	*Implemented*	EY, TUD

B.5 Requirement: visualisation

Table 17: Requirement: visualisation.

ID	Topic	Key word	Description	Status	Reference to use case
#056	Different meanings of search term		It should be possible to see via the visualisation whether the search term/word has different meanings due to different disciplines.	*Cancelled due to technical complexity*	#TUD005
#057	Connections between different types of sources		It should be possible to see connections between different types of sources, e.g. journals, archives, conferences, books.	*Implemented*	#TUD090
#058	Different types of visualisation		Choosing different options of visualisation.	*Implemented*	#TUD056

#059	Top Concepts	General	Two bar charts (one for entities, one for other concepts) that show the concepts that best describe the data. The length of each concept's bar depends on the relevance of the concept. Additional information should be displayed when clicking on a bar. For example: When clicking on a concept bar, the titles of the documents should be displayed from which the concept was extracted, for keywords and sources similar.	*Implemented*	#EY043
#060		Sensitivity	There should be an option to change the number of bars (i.e., concepts) displayed in each bar chart.	*Cancelled (to reduce UI complexity); the number of bars will be limited to a maximum of 25-30 by default*	#EY044
#061	Top sources	General	A bar chart that shows the most relevant data sources for the given search query. The length of each concept's bar depends on the relevance of the source.	*Implemented*	#EY045
#062		Sensitivity	There should be an option to change the number of bars (i.e., sources) displayed in the bar chart.	*Cancelled (to reduce UI complexity); the amount of bars will be limited to a maximum of 25-30 by default*	#EY046
#063	Topic network	Concept Graph	Visualisation as topic network. Concept graph to visualise topics around a specific keyword. The graph consists of notes and edges linking those nodes. The nodes are the most relevant concepts within the data and the edges link the concepts (nodes) that frequently co-occur within the data.	*Implemented*	#TUD057, #TUD091, #EY017

#064		Filter	Ability to limit the view to certain types of concepts, for example, persons and organisations only.	*Implemented*	#EY038
#065	Network	Navigation	The user should be able to navigate through the network, diving deeper into areas of interest and expanding the network in relevant directions. To increase the efficiency of the analysis, there should be an indication for the user how many documents/entities/locations/concepts will become visible when expanding the network accordingly.	*Implemented*	#EY018
#066		Completeness	There should be an indicator of the degree of completeness on the current view. This should be based on the additional information that can be obtained by extending the network.	*Implemented*	#EY019
#067		Filter	There should be a general option to change the displayed number of nodes and edges. With the minimum settings, the graph only shows the most relevant nodes/edges.	*Implemented*	#EY020
#068		Size of nodes, thickness of edges	By default, the size of the node depends on the concept's relevance (e.g., frequency of occurrence within the data) and the thickness of the edges depends on the number of co-occurrences of the connected nodes. We should consider adding alternatives, e.g., sizing the nodes by in-degree or out-degree.	*Implemented*	#EY021

#069		Co-occurrence of concepts	Ability to specify the co-occurrence of concepts that is displayed in the graph (e.g., concepts connected by edges are included in the same text, paragraph or sentence).	*Implemented*	#EY022
#070		Hovering	When hovering the mouse over nodes and edges, there should be additional information about the attributes of the node/edge (e.g., number of occurrences, data source with the most occurrences).	*Implemented*	#EY023
#071		Statistics and measures	The network development should go hand in hand with statistical measures, e.g. centrality, concentration, density, shortest path, community clustering. These measures will be determined in more detail in D1.3.	*Implemented*	#EY024
#072		Colour	The colour of the nodes corresponds to the "entity identification" algorithm, e.g., person = red, organisation = green, location = blue, another concept = black.	*Implemented*	#EY039
#073	Topic network		Moving the topic network by click and hold.	*Implemented*	#TUD061, #EY031
#074			Zooming in and zooming out of the network.	*Implemented*	#TUD061, #EY031
#075	Author network		Visualisation of the authors of the sources.	*Implemented*	#TUD064
#076	One network for all		Visualisation of patterns, concepts, people and phrases in one network.	*Implemented*	#TUD055
#077	Focus on references in the sources		Visualisation of frequently occurring references in the search results.	*Backlog, other requirements had priority*	#TUD062

#078	Tag cloud	General	Visualisation as topic as tag cloud of topics. The tag cloud displays the top keywords extracted from the data.	*Implemented*	#TUD058, #TUD063, #EY025
#079		Size	The initial size of the tag depends on the tag's frequency within the search results.	*Implemented*	#EY026
#080		Position	The position of the tag within the tag cloud depends on the relevance. The initial position of the tag within the tag cloud depends on the keyword's relevance. However, there should also be an option to order the tags alphabetically.	*Implemented*	#EY027
#081		Rating of tags	When reviewing the initial tag cloud, the user can "rate" the different tags according to his research question/area of interest. Rating tags leads to a refresh of the tag cloud taking into account these ratings.	*Implemented (uRank provides means of re-ranking the results based on user ratings)*	#EY028
#082		Sensitivity	There should be an option (e.g., a scroll bar) to adjust the sensitivity of the analysis. When changing the sensitivity, the size of the tag cloud changes as the required frequency/relevance of a concept to be included in the tag cloud changes.	*Implemented*	#EY029
#083		Colour	The colour of the tag corresponds to the "entity identification" algorithm, e.g., person = red, organisation = green, location = blue, other concept = black.	*Implemented*	#EY040

#084		Learning	The platform should learn from the user behaviour. For example, when a number of users rated a tag (e.g., a related organisation) as "high", the standard rating/weight for this tag in future search queries should adjust automatically.	*Cancelled (there will not be a link between the user profiles and the search results)*	#EY041
#085	Tag cloud for key words		Visualisation of key words used in the databases and on the platform by users.	*Cancelled (there will not be a link between the user profiles and the search results)*	#TUD054
#086	Tagging		The user should be able to tag/mark concepts for further analysis and thus be able to export a summary containing the tagged concepts and (a list of) the underlying data sources.	*Cancelled (differently implemented - this requirement is fulfilled by the Top Properties – Concepts visualisationas it allows further analysis of the concepts of interest*	#EY065
#087	Tree visualisation		Tree visualisation of search results to see which results are based on another.	*Backlog, other requirements had priority*	#TUD092
#089	Date mentions	General	This visualisation does not show the data but a calendar. The font size of the days in the calendar depends of the frequency those dates are mentioned within the data. When a day is not mentioned within the data, it is not visible in the visualisation.	*Cancelled (differently implemented - this requirement is fulfilled by the size of the "Year" node in the Concept Graph Visualisation)*	#EY063
#090		Sensitivity	The user can choose to display only dates mentioned more than X times within the data.	*Cancelled (requirement is no longer of interest, see #089)*	#EY064

#091	Help		Each visualisation should contain a help icon where the user can obtain instructions about the current visualisation and tips how to adjust and interpret it. We can also consider integrating this feature into the adaptive training support.	*Implemented*	#EY032
#092	Document pane		All visualisations should provide a document pane listing the documents according to the current search settings. The document pane should contain the (default) option to include all documents in the visualisation. The user should be able to select one, multiple or all documents. Upon selecting documents, all visualisations should update accordingly.	*Cancelled; no longer required due to the platform's current functionalities*	#EY030
#093	Filter	Geographical area	Based on the search results, a world map displays any geographic data contained by markers on the map. The user can limit the search results to specific locations by selecting and unselecting them.	*Cancelled (due to technical complexity)*	#EY054
#094	Search feature		There should be a search function to locate certain keywords within the visualisations.	*Implemented*	#EY056

#095	Drill down		At any stage of the analysis, the user needs to drill down to the detailed data. When selecting a concept (e.g., a tag in the Tag cloud or a node in the network graph), the user should be able to receive a list of the corresponding documents, including the links to open these documents.	*Implemented*	#EY057
#096	Export		To facilitate the review from engagement executives, the user should be able to export the visualisations. In the export, the work done within the program should be documented, i.e., the search settings and the steps performed to adjust the visualisation.	*Implemented*	#EY058
#097	Over time development		The network and Tag cloud visualisation also show the evolution, development and degeneration of concepts over time (taking into account the document dates) and thus allow the assessment of the relevance of a concept.	*Implemented*	#EY042

B.6 Requirement: document search and analysis

Table 18: Requirement: document search and analysis.

ID	Topic	Key word	Description	Status	Reference to use case
#098	Full-text access		Provision of full-text access.	*Implemented*	#TUD102
#099	Full-text search		It should be possible for the user to search for information and keywords within the full-text of a resource he has found.	*Implemented*	#TUD014, #TUD068

#100	Colour of keywords		Automatic displaying the keywords in the text with a colour.	*Backlog, other requirements had priority*	#TUD073, #TUD100
#101	OCR		Text recognition on the platform.	*Implemented*	#TUD074
#102	Marking in the text		Make colour marks within texts on the platform.	*Implemented*	#TUD103
#103	Tagging in the text		Tagging of words, section, phrases with keywords within the text on the platform.	*Backlog, other requirements had priority*	#TUD105
#104	Linking of key words within one text		Automatic linking of key words (same words) within the text and through clicking on it jumping to the next phrase or paragraph with this key word.	*Backlog, other requirements had priority*	#TUD069
#105	Linkage of key word within more texts		Automatic linking of key words (same words) within texts on my search list and through clicking on it jumping to the next phrase or paragraph with this key word.	*Backlog, other requirements had priority*	#TUD070
#106	Comparison of documents		Visual comparison of documents side by side.	*Backlog, other requirements had priority*	#TUD087, #TUD088, #TUD089, #TUD111
#107	Entity identification	(Sub-) document analysis	Entity identification in preparation for the visualisations: Extraction of entities, locations, persons and other top concepts from the data.	*Implemented*	#EY016
#108		Abbreviations	The concept identification algorithm should recognise abbreviations and treat the abbreviated and the written-out word as one concept. This especially applies to laws and regulations (e.g., "HGB" and "Handelsgesetzbuch").	*Backlog*	#EY050

#109		Uniqueness	As the entity identification is key to the EY use cases, we should consider including "dictionaries" from DUNS or ISIN to take advantage of the uniqueness of those systems.	*Backlog*	#EY060
#110		Excel files	As journal entry descriptions usually contain multiple words, the program should be able to separate multiple words stored in MS Excel cells.	*Backlog (currently only implemented PDFs)*	#EY061
#111		Abbreviations	As there can be several abbreviations contained in the data, the tool should contain a dictionary to recognise common abbreviations.	*Merged with #108*	#EY062
#914	Entity Extraction		When analysing the management report and the consolidated financial statements of a company, i.e. "Villeroy & Boch AG" the entity "organization" should be unique and named "Villeroy & Boch AG".	*Implemented*	

B.7 Requirement: video search and analysis

Table 19: Requirement: video search and analysis.

ID	Topic	Key word	Description	Status	Reference to use case
#112	Annotations in videos		Generating annotations for specified videos.	*Implemented*	#TUD108

B.8 Requirement: Adaptive Training Support

Table 20: Requirement: Adaptive Training Support.

ID	Topic	Key word	Description	Status	Reference to use case

#113	Thesauri/Synonyms		The user should be informed about which words or search terms frequently occur with each other through the adaptive training support for the MOVING platform.	*Cancelled as the Tag cloud fulfils this requirement. Implementation into ATS would be redundant.*	#TUD003
#114			The user should be informed about frequently occurring authors during the research.	*Cancelled as the network graph fulfils this requirement. Implementation into ATS would be redundant.*	#TUD010
#115	Recommendation		Getting recommendations of further steps while conducting the search due to the search by other users.	*Cancelled due to technical complexity*	#TUD122
#116	Going to discussion page		Recommendation of going to the discussion page of the source/topic.	*Implemented*	#TUD076
#117	Search list		When searching for specific paragraphs or laws and regulations, the ATS should remind the user that it might be useful to sort the search list by publication date in order to address recent changes and comments that might be relevant.	*Implemented*	#EY052
#915	Curriculum Widget	Open learning environment in new tab	*When clicking on learn more, the learning environment should be opened in a new tab. When coming back from the learning environment the page should be refreshed. IMPORTANT: The page should only be refreshed, if the user previously opened the learning environment and NOT on any page focus.*	*Implemented*	
#916	Curriculum Widget	Placing EY Learning Track 4 on top of learning tracks in learning environment	*EY user will see Learning track 4 on top of the left side navigation bar in the learning environment*	*Implemented*	

B.9 Requirement: community

Table 21: Requirement: community.

ID	Topic	Key word	Description	Status	Reference to use case
#118	Network	Contacting other researchers	It should be possible to contact other researchers via the platform.	*Implemented*	#TUD033
#119		Asking questions	Possibility of asking questions.	*Implemented*	#TUD121
#120			It should be possible to inform one about other researchers.	*Implemented*	#TUD009
#121	Pictures		Sharing found pictures due to the search on the platform with other user of the platform.	*Implemented*	#TUD115
#122	Literature		Sharing of found literature due to the search on the platform with other user of the platform.	*Implemented*	#TUD116
#123	Ranking		Ranking of found resources due whether or not the result fit to the search query.	*Cancelled due to decision that users should not be able to manipulate the search index. Users can use the uRank functionality to analyse search results according to their personal preferences (e.g. combination with other keywords).*	#TUD117
#124	Discussion Forum on search results		Possibility of discussing sources found in the search results.	*Implemented*	#TUD076
#125	Discussion Forum on availability of literature		Discussing availability of literature.	*Implemented*	#TUD127
#126	Author details		When searching for an author, it should be displayed whether the author has created a profile on the platform, which publications the author has published, where the author is cited and who the author cites (which is done automatically).	*Cancelled due to the decision to have no connection between the user data base and the document data base*	#TUD124

B.10 Requirement: user management

Table 22: Requirement: user management.

ID	Topic	Key word	Description	Status	Reference to use case
#127	Tracking queries		Tracking of queries on search terms, search results, used databases. *Save search queries as social bookmarks on user page.*	*Tracking is implemented, saving search queries has been backlogged*	#TUD078
#128	Tagging of queries		Tagging of queries for finding them easier.	*Backlog, other requirements had priority*	#TUD082
#129	Social bookmarking feature		Meta data library for saving and tagging search results with keywords.	*Backlog, other requirements had priority*	#TUD079, #TUD097, #TUD099
#130			Saving the search results.	*Cancelled; merged with #129.*	#TUD095
#131		Reference management system on the platform	Organising and comparing the search results in a reference management on the platform.	*Backlog due to technical complexity, and other requirements had priority*	#TUD083, #TUD096, #TUD104, #TUD077, #TUD111, #TUD087, #TUD088, #TUD089
#132			Marking the found literature with colours.	*Backlog, other requirements had priority*	#TUD106
#133	BibTex plugin		BibTex plugin to transfer the reference automatically, which I found on the platform into a document outside the platform. Other referencing plugins, such as RIS, Text and EndNote, could also be added.	*Backlog, other requirements had priority*	#TUD114
#134	Documentation of search terms		Tracking of which search term was used and in which manner (Boolean operation); display search path on the search page: show used keywords for Title/Abstract/-Full Text; Boolean Operators AND OR NOT.	*Implemented*	#TUD080
#135	Notes		Writing notes or excerpts on texts/-sources.	*Backlog, other requirements had priority*	#TUD084

#136			Copy and paste these notes.	*Backlog, other requirements had priority*	#TUD086, #TUD085
#137			Download the notes.	*Backlog, other requirements had priority*	#TUD086, #TUD085
#138			Save the notes.	*Backlog, other requirements had priority*	#TUD086, #TUD085
#139		Reference management system outside the platform	Exporting the search results into reference management system, which is not included on the platform.	*Cancelled; merged with #133.*	#TUD098
#140	Upload of resources		Uploading resources.	*Cancelled due to copyright issues*	#TUD110
#141	Downloading resources		Downloading resources.	*Cancelled due to copyright issues*	#TUD109
#142	Managing access of uploaded resources		Managing access of uploaded resources into open access or not.	*Cancelled due to copyright issues*	#TUD110
#143	Collaborative text creation		Creating of texts, which can be accessed collaboratively.	*Cancelled due to technical complexity*	#TUD119, #TUD120
#144	Search profile management	Save settings	Save search settings into search profiles or favourites for later use (e.g., the next year's understanding of the entity and the environment).	*Backlog*	#EY013
#145		Import/Export settings	Option to import and export search settings in order to share them with colleagues or use them for documentation purposes. Sharing inside the MOVING platform will be sufficient.	*Implemented*	#EY014
#146		Predefined settings	Provide pre-defined search settings for the different uses of the platform. For example, for the ISA 315 scenario, the profile should always include the company website and a hint from the adaptive training support also to include the latest management report and notes to the financial statements available (as files from the local device).	*Merged with #144*	#EY015

#902	Search history		Display user's last 5 searches on search start page.	*Implemented*	TUD
#903	In search for partners	Profile page	Add a button/field within the user profile for indicating whether the person is in search for partners.	*Implemented*	#TUD007

B.11 Requirement: accessibility of the platform

Table 23: Requirement: accessibility of the platform.

ID	Topic	Key word	Description	Status	Reference to use case
#147	Accessibility to the platform from different devices		Accessing the platform from different devices.	*Implemented*	#TUD107

B.12 User Study Requirements

As a result of the study carried out in D1.3 (Apaolaza et al., 2018), a set of updated requirements were provided categorised as presentation, features, and data. For each update, the corresponding code for the requirement reported in D1.1 (Bienia et al., 2017) is mentioned.

B.12.1 Presentation

R1: Increase size of clickable elements. Particularly in the case of the selection of filter dropbowns, the clickable elements are redundant, and too small. Instead, the interactivity should be moved to the container element, which has been shown to meet the expectations of the participants in the study.

R2: Avoid unrequested obscuring dialogues. In the case of Concept Graph, an unrequested information dialogue appears when the participant hovers a node. As the shown information is generally large, this dialogue obscures the hovered node completely, hindering further interaction. This visualisation also contains a help dialogue with the same behaviour, not used by participants in this study. Unless disclosing this information is shown in an unobtrusive way (i.e. not hindering possible interactions in any way) users should be able to control when and how the information is shown to them in a explicit way. A similar problem occurs when an error is triggered in the Top Concepts visualisation. A visible alert is shown to the user, indicating the lack of results for one of the visualisations. If the notification of such an error is necessary, it should be done in a way that does not interfere with the interaction.

R3: Additional help functionalities. A way to support newcomers to the platform is necessary. For example, the ambiguities among the filters reported by the participants (e.g. participants mistaking venue and dataset) can be tackled by adding a description to each of them. Further help functionalities, such as a user triggered overlay describing the functionality of each available feature, would prevent participants resorting to external sources. This recommendation is related to the requirement *#091* in Section B.5, but can be extended to the entire platform.

R4: Feedback in visualisations. Several participants thought that the visualisations were not working, due to the extended loading period without any visible feedback. Once loaded, participants had problems interpreting the information shown to them. In the case of Concept Graph, there is no discernible difference between closed and opened nodes, and if the number of connections is similar, the difference in size is not noticeable. The differences in size between the keywords presented in Tag cloud also needs to be made more explicit, as it codes all the information in the size of the keywords, but participants failed to distinguish them.

In uRank, participants thought the search functionality was not working, as the list of tags was too long, and the feedback indicating the highlighted keyword insufficient.

B.12.2 Features

R5: Maintain page state. One of the key issues that participants found when interacting with the platform was the unrequested update of the state of the page. When carrying out a search, every time a new filter is selected, the page carries out the new search, and resets the state of all page inputs. As all the information is contained in the URL, the visible state of the input elements (e.g. filters opened/closed, text inputs) in the page are reset. This unrequested update confused participants, and made them lose any data not contained in the URL. Any unrequested change to the state of the user's interaction should be avoided.

R6: Misleading or non-working features. Although certain elements' descriptions can be found ambiguous (see "R3: Additional help functionalities" in section above) and require more specific labels, other descriptions have been found to be misleading. In the case of Tag cloud, the sorting works alphabetically (as the result of requirement *#080* but without informing the user), which contradicts the logic of the Tag cloud, which is based on frequency, and there is no description about how the "default" sorting works. In the case of the use of booleans in the Basic Search, the addition of the conjunctive operator "AND" resulted in a larger number of results. For both these cases, the problem can be a misleading instruction to the participant, or a non-working feature that needs to be fixed.

R7: Communication between modules. Several participants expected that their interaction with a module would have an effect in the search they were carrying out. For example, participants selected tags in the visualisations expecting to either see related information, add them to the search, or show related results or keywords. Adding a connecting feature between modules would help making the platform more coherent, and complement the use of filters to narrow down the search, so all modules can contribute to the iterative refinement of the search.

R8: Performance. Several participants have complained about the speed of the platform, which was particularly problematic when exploring multiple visualisations. Some participants tackled the loading time problems by opening multiple tabs, however, most of them ended up going back to previously loaded visualisation at some point. Although performance improvements are expected as the platform matures, the use of temporary local storage in users' browsers can be considered to reduce loading times when going back to the same visualisation.

R9: Lack of transparency. Some participants have expressed "trust" issues with regards to automatic metrics and the use of metadata. Participants were not sure about what comprises the "relevance" metric. One participant was reluctant to use filters, as including these would require all results to contain the necessary metadata, which would possibly discard interesting results. Although the participants admitted that this problem is general to most information seeking tools, a way to provide some information about how the automatic metrics for all the modules are computed, and information about how many documents contain the correct metadata for each filter would increase participants' trust in the platform.

R10: Natively supported auto-complete. Several participants made use of the auto-complete feature from the used browser. This feature only included the search history from previous tasks (search history was not accessible between participants). A feature supporting auto-complete including user search history, or related keywords, would add to the platform's value. This recommendation has already been defined as a requirement (see *#018* in Section B.1) but remains to be implemented.

R11: Features to handle more nodes. Although useful, the Concept Graph visualisation can become overwhelming when several nodes are opened. Several participants tried closing back nodes, and asked for ways to reset the state of the visualisation to its original state. Another participant asked for better ways to move a group of connected nodes, so they could be isolated, or even removing them from the visualisation. Finally, an overview of the displayed nodes, and a way to show some information next to the nodes, or colour code them, so they can be compared without hovering them would ease the exploration of large amounts of concurrent nodes.

R12: Removing non-relevant keywords. Although the use of filters, and uRank keywords have been shown helpful to narrow down participants' searches, some participants have complained that the inclusion of particularly frequent keywords, in some cases similar to the ones used in the search, obscure less frequent yet relevant ones. A way to automatically (if all results contain the keyword) or manually remove (giving the use the feature to select individual keywords) would help narrowing down searches more effectively, while increasing the usefulness of the visualisations. This feature would complement requirements such as *#081*, *#082*, and *#084*.

R13: Handling large amounts of data. Participants have been seen to be having problems when handling large amounts of search options. The list of available filters to select from becomes unmanageable, forcing users to carefully read each entry. A way of navigating through the filters, with custom sorting, or auto-complete search boxes, is necessary. As the participants are not aware of the possible filter values before the search, an intelligent matching of search keywords and filter options could be suggested to the user. This issue should be considered in any module that presents list of results to the user, such as the keyword list in uRank, which forced participants to scroll down to manually find the desired keyword.

B.12.3 Data

R14: Different results between visualisations. Several participants have noticed that the provided results differed between visualisations. Top Concepts was the only one showing duplicate or empty keywords, and the frequency counts were different. As it was obvious that the data used for the computation was different, participants also expressed their concerns about how many results from the main search had been included in the visualisations. More consistency across the visualisations, with clear indications about what data is being included, is necessary.

