



Deliverable 1.3: Initial evaluation, updated requirements and specifications

Aitor Apaolaza, Irina Bienia, Annalouise Maas, Michael Wiese, Franziska Günther, Sabine Barthold, Markel Vigo

30/03/2018

Work Package 1: Requirements and use case development

**TraininG 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	31/03/2018
Actual date of delivery	30/03/2018
Deliverable number	1.3
Deliverable name	Initial evaluation, updated requirements and specifications
File	MOVING_D1.3_v1.0.tex
Nature	Report
Status & version	Final & v1.0
Number of pages	64
WP contributing to the deliverable	1
Task responsible	UMAN
Other contributors	EY, TUD
Author(s)	Aitor Apaolaza, Irina Bienia, Annalouise Maas, Michael Wiese, Franziska Günther, Sabine Barthold, Markel Vigo
Quality Assessors	Chrysa Collyda, Vasileios Mezaris, CERTH
EC Project Officer	Hinano Spreafico
Keywords	User studies, initial evaluation, requirements

Executive Summary

This deliverable “D1.3: Initial evaluation, updated requirements and specifications” updates the list of requirements identified in D1.1, and presents the results of the user studies initially implemented in D1.2. The list of requirements is reviewed in Section 2, and the state of all requirements updated indicating their implementation progress, their cancellation, or their postponement. The final design of the studies for each use case is presented in Section 3. Although the provided tasks needed to be adapted for each use case, the rest of the study has been kept the same for consistency. The results of the analysis are presented in Section 4, presenting the results of the usability metrics reported by the participants, and providing a summary of the gathered feedback. A report on the evaluation, and a list of recommendations for next steps are provided in Section 5

Table of contents

Executive Summary	3
Abbreviations	7
1 Introduction	8
1.1 History of the document	8
1.2 Purpose of the document	8
1.3 Structure of the document	8
2 Updated Requirements	8
2.1 Requirement: search field	8
2.2 Requirement: faceted search	10
2.3 Requirement: data sources	12
2.4 Requirement: search list	14
2.5 Requirement: visualisation	15
2.6 Requirement: document search and analysis	22
2.7 Requirement: video search and analysis	24
2.8 Requirement: Adaptive Training Support	24
2.9 Requirement: community	25
2.10 Requirement: user management	26
2.11 Requirement: accessibility of the platform	28
3 Design of the user studies	28
3.1 Metrics	29
3.2 Apparatus	30
3.3 Use case 1: Research on business information by public administrators	30
3.4 Use case 2: Managing and mining research information	31
4 Analysis of the use cases	31
4.1 Use case 1: Research on business information by public administrators	32
4.2 Use case 2: Managing and mining research information	35
5 Evaluation report and next steps	42
5.1 Current state of the MOVING platform's usability	42
5.2 Recommendations for next steps	43
6 Conclusions	46
References	47
A APPENDIX	48
A.1 Tutorial	48
A.2 Questionnaires	55
A.3 Task notes	56
A.4 Use case 1 tasks	57
A.5 Use case 2 tasks	61

List of Figures

1	SUS questionnaire adjective ratings scale	30
2	Use case 1: Distribution of expected vs perceived difficulty for each of the tasks.	33
3	Use case 1: Ease of use boxplots.	33
4	Use case 1: Distribution of the reported SUS scores.	33
5	Use case 1: Distribution of the completion times for each task.	34
6	Use case 2: Distribution of expected vs perceived difficulty for each of the tasks.	36
7	Use case 2: Ease of use boxplots.	37
8	Use case 2: Distribution of the reported SUS scores.	37
9	Use case 2: Distribution of the reported SUS scores for Computer Science and Business students.	38
10	Use case 2: Distribution of the completion times for each task.	38
11	Distribution of SUS scores for use case 1 and use case 2.	43

List of Tables

1	Document history.	8
2	Requirement: Search field.	9
3	Requirement: faceted search.	10
4	Requirement: data sources.	12
5	Requirement: search list.	14
6	Requirement: visualisation.	15
7	Requirement: document search and analysis.	22
8	Requirement: video search and analysis.	24
9	Requirement: Adaptive Training Support.	24
10	Requirement: community.	25
11	Requirement: user management.	26
12	Requirement: accessibility of the platform.	28
13	Use case 1: Expected and perceived difficulties and ease of use reported for all tasks.	32
14	Use case 2: Expected and perceived difficulties and ease of use reported for all tasks.	36
15	Ease of Use questionnaire	55
16	SUS usability questionnaire	55

Abbreviations

Abbreviation	Explanation
ATS	Adaptive Training Support
ISA	International Standards on Auditing
PEST	Political, economic, social and technical analysis
SUS	System Usability Scale
SWOT	Strengths, weaknesses, opportunities and threats analysis

1 Introduction

1.1 History of the document

Table 1: Document history.

Date	Version
06/02/2018	v0.1: first ToC draft
16/02/2018	v0.2: ToC ready for QA
28/02/2018	v0.3: ToC comments addressed
19/03/2018	v0.4: content ready for QA
23/03/2018	v0.5: document ready for final QA

1.2 Purpose of the document

This document provides an update to the list of requirements presented in D1.1, detailing the current state of each of them. In addition, this document builds upon the initial implementation of user studies presented in D1.2. The specification of the studies and their prospective tasks are adapted to carry out two user studies, one for each target end-user group. The outcomes of these studies are presented and a set of updated requirements and recommendations for the MOVING platform are provided.

This document will be followed by the final implementation of user studies and evaluation in D1.4, which will further test the compliance to the specified requirements, providing insight into the usage of the MOVING platform for prolonged periods of time.

1.3 Structure of the document

This document, in Section 1, starts with a revision of the requirements presented in D1.1. Section 2 contains a list of all the updated requirements, indicating their current state. The rest of the document focuses on the user studies carried out to evaluate the platform for each of the target use cases, and is structured into three main sections, reporting the design (Section 3), analysis (Section 4), and report of the studies carried out for each target end-user group (Section 5).

2 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.

2.1 Requirement: search field

Table 2: 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. <i>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.</i>	<i>Implemented; author disambiguation to be added</i>	#TUD040
#002	Title			<i>Implemented</i>	
#003	Key word			<i>Implemented</i>	#TUD006, #TUD015
#004	Combination of words		It should be possible not only to search for one word but for combination of words. <i>Compound words (e.g. computer science, data mining) should be recognised as single not separate search term.</i>	<i>Implemented</i>	#TUD001
#005	Including and excluding words		It should be possible to exclude and include words or terms, e.g., <i>searching for "social NOT workers" or "social –workers" will return different results than searching for "social workers".</i>	<i>Implemented</i>	#TUD002
#008	Advanced search	Multiple search terms	Define one or multiple search terms, e.g. <i>searching for data science AND machine learning or searching for data science AND a specific author.</i>	<i>Work in progress</i>	#EY002
#010	Search extent	Search depth	Refine the search by title, abstract, full-text.	<i>Implemented</i>	#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.	<i>To do</i>	#EY034
#901 (new)	<i>Simple search: Search domains</i>		<i>Different search domains (research, funding, and learning) should be displayed within the search bar as drop-down menu.</i>	<i>Implemented</i>	TUD

2.2 Requirement: faceted search

Table 3: 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).	<i>Work in progress</i>	#TUD038, #EY036
#007	Search term		Exclude one or multiple search terms.	<i>Merged with #005</i>	#EY001
#009	Media type		Searching in and possibility of excluding monographies, journal articles, open access journal articles, conference articles, posters.	<i>Implemented</i>	#TUD035
#012	Datasets		Listing of all databases (sorted by discipline).	<i>Implemented</i>	#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).	<i>Implemented</i>	#EY011
#020	Geographic Region	Location search	Limit the search to certain geographical areas (e.g., the entity's headquarters country).	<i>Work in progress</i>	#EY009
#021	Language		Limit the search to specific languages. By default, the language is in correspondence to the search term defined.	<i>Implemented</i>	#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.	<i>To do</i>	#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.	<i>Work in progress</i>	#EY049
#023b	Author		Refine the search by author by excluding and including authors.	<i>Implemented</i>	#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).	<i>Implemented</i>	#TUD041, #TUD050, #EY010
#026	Citations		Refine the search by the amount of citations of the source.	<i>Backlog</i>	#TUD042, #TUD051
#027	Clicks		Refine the search after the number of clicks on the source through the users.	<i>Cancelled, not objective enough for the search</i>	#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.	<i>To do</i>	#TUD004, #TUD081
#088	Timeline visualisation		A timeline visualisation showing how search results appear chronologically.	<i>Work in progress</i>	#TUD093, #TUD094

2.3 Requirement: data sources

Table 4: Requirement: data sources.

ID	Topic	Key word	Description	Status	Reference to use case
#011	Datasets Databases	<i>Literature databases</i>	Searching and possibility of including or excluding databases (see D1.1). Connection to various literature databases available to include publications of several disciplines.	<i>Work in progress</i>	#TUD011, #TUD012, #TUD016, #TUD089, #EY007
#013	Datasets Data sources		Include or exclude certain data sources.	<i>Implemented</i>	#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.	<i>Merged with #030, technically difficult to include the world wide web</i>	#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.	<i>To do</i>	#EY005
#016		Data source upload	The upload of data sources stored on the local device is restricted due to data privacy issues.	<i>Backlog</i>	#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.	<i>Backlog</i>	#EY055
#024	Survey data		Searching and possibility of including or excluding survey data (GESIS, Statista ¹).	<i>To do</i>	#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.	<i>Work in progress</i>	#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).	<i>Work in progress</i>	#TUD020, #TUD024, #TUD030, #TUD034
#031	Online Encyclopaedia	Wikipedia	Including and excluding Wikipedia in the search.	<i>Cancelled, dataset too big</i>	#TUD025, #TUD026
#032			Comparing Wikipedia articles in different languages.	<i>Cancelled, see #031</i>	#TUD075
#033	Website	archive.org	Including and excluding archive.org in the search.	<i>Cancelled, see #031</i>	#TUD031
#034	Search engines	Google, Bing, Yahoo	Searching and possibility of including or excluding existing search engines.	<i>Work in progress</i>	#TUD026
#035	Library catalogues		Searching and possibility of including or excluding library catalogues (see D1.1).	<i>To do</i>	#TUD027, #TUD028
#036	PDF files	e.g. Plenary protocols of the Bundestag	Including or excluding PDF files in the search.	<i>Cancelled, there is no need to host external PDFs unless uploaded, which is covered by #015</i>	#TUD029

¹<https://de.statista.com> (2017-03-27)

#037	Project partners	International	It should be possible to look for possible project partners from different countries.	<i>Work in progress</i>	#TUD007
#038	Business partners	International	It should be possible to look for business partners via the platform.	<i>Work in progress</i>	#TUD008, #TUD023
#039	Videos		Searching and possibility of including or excluding videos.	<i>Implemented</i>	#TUD036
#040	Data source for funding		Possibility of including or excluding funding databases.	<i>Work in progress</i>	#TUD088
#041	Creative commons	Flickr	Searching in and possibility of excluding or including sources under the licence of creative commons (e.g. Flickr ²).	<i>To do</i>	#TUD113

2.4 Requirement: search list

Table 5: Requirement: search list.

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

²<https://www.flickr.com> (2017-03-27)

#050	Bookmark		Bookmark several search results.	<i>To do</i>	#TUD067
#051	Library connection		Displaying where the source can be found (e.g. in a library) by clicking on it.	<i>To do</i>	#TUD071, #TUD128
#052	Mark results		Marking the results due to high or low relevance on the search topic.	<i>Cancelled due to technical complexity</i>	#TUD072
#053	Open access		Displaying whether the source is open access or not.	<i>Work in progress</i>	#TUD112, #TUD118
#054	Creative commons		Displaying whether the source is under the licence of creative commons or not.	<i>Work in progress</i>	#TUD113, #TUD118
#055	Author Contact		Displaying the contact details of the author of the resource (especially from frequently occurring authors).	<i>To do</i>	#TUD125, #TUD126
#917	View/hide abstracts		<i>A show abstract/hide abstract button that displays more details for each result within the same frame</i>	<i>Backlog</i>	TUD

2.5 Requirement: visualisation

Table 6: 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.	<i>To do</i>	#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.	<i>Implemented</i>	#TUD090
#058	Different types of visualisation		Choosing different options of visualisation.	<i>Implemented</i>	#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.	Work in progress	#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.	Work in progress	#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 nodes 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.	Work in progress	#TUD057, #TUD091, #EY017

#064		Filter	Ability to limit the view to certain types of concepts, for example, persons and organisations only.	<i>Work in progress</i>	#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.	<i>Implemented</i>	#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.	<i>Implemented</i>	#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.	<i>To do</i>	#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.	<i>Implemented</i>	#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).	<i>Implemented</i>	#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).	<i>Implemented</i>	#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.	<i>To do</i>	#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.	<i>Implemented</i>	#EY039
#073	Topic network		Moving the topic network by click and hold.	<i>Implemented</i>	#TUD061, #EY031
#074			Zooming in and zooming out of the network.	<i>Implemented</i>	#TUD061, #EY031
#075	Author network		Visualisation of the authors of the sources.	<i>Work in progress</i>	#TUD064
#076	One network for all		Visualisation of patterns, concepts, people and phrases in one network.	<i>Work in progress</i>	#TUD055
#077	Focus on references in the sources		Visualisation of frequently occurring references in the search results.	<i>Backlog</i>	#TUD062

#078	Tag cloud	General	Visualisation as topic as tag cloud of topics. The tag cloud displays the top keywords extracted from the data.	<i>Implemented</i>	#TUD058, #TUD063, #EY025
#079		Size	The initial size of the tag depends on the tag's frequency within the search results.	<i>Implemented</i>	#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.	<i>Implemented</i>	#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.	<i>Implemented (uRank provides means of re-ranking the results based on user ratings)</i>	#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.	<i>To do</i>	#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.	<i>To do</i>	#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.	<i>Backlog (there is no link between the user's profile and the search results)</i>	#EY041
#085	Tag cloud for key words		Visualisation of key words used in the databases and on the platform by users.	<i>Cancelled (there will not be a link between the user profiles and the search results)</i>	#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.	<i>Backlog</i>	#EY065
#087	Tree visualisation		Tree visualisation of search results to see which results are based on another.	<i>Backlog</i>	#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.	<i>Backlog</i>	#EY063
#090		Sensitivity	The user can choose to display only dates mentioned more than X times within the data.	<i>Backlog</i>	#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.	<i>Work in progress</i>	#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.	<i>Cancelled; no longer required due to the platform's current functionalities</i>	#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.	<i>Cancelled (due to technical complexity)</i>	#EY054
#094	Search feature		There should be a search function to locate certain keywords within the visualisations.	<i>To do</i>	#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.	<i>Work in progress</i>	#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.	<i>To do</i>	#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.	<i>To do</i>	#EY042

2.6 Requirement: document search and analysis

Table 7: Requirement: document search and analysis.

ID	Topic	Key word	Description	Status	Reference to use case
#098	Full-text access		Provision of full-text access.	<i>Implemented</i>	#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.	<i>Implemented</i>	#TUD014, #TUD068

#100	Colour of key-words		Automatic displaying the keywords in the text with a colour.	<i>To do</i>	#TUD073, #TUD100
#101	OCR		Text recognition on the platform.	<i>Cancelled; not necessary due to not hosting PDFs</i>	#TUD074
#102	Marking in the text		Make colour marks within texts on the platform.	<i>Work in progress for the meta data</i>	#TUD103
#103	Tagging in the text		Tagging of words, section, phrases with keywords within the text on the platform.	<i>To do</i>	#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.	<i>Backlog</i>	#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.	<i>Backlog</i>	#TUD070
#106	Comparison of documents		Visual comparison of documents side by side.	<i>To do</i>	#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.	<i>Work in progress</i>	#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").	<i>To do</i>	#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.	<i>To do</i>	#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.	<i>To do</i>	#EY061
#111		Abbreviations	As there can be several abbreviations contained in the data, the tool should contain a dictionary to recognise common abbreviations.	<i>Merged with #108</i>	#EY062

2.7 Requirement: video search and analysis

Table 8: Requirement: video search and analysis.

ID	Topic	Key word	Description	Status	Reference to use case
#112	Annotations in videos		Generating annotations for specified videos.	<i>Work in progress</i>	#TUD108

2.8 Requirement: Adaptive Training Support

Table 9: Requirement: Adaptive Training Support.

ID	Topic	Key word	Description	Status	Reference to use case
#113	The-sauri/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.	<i>Cancelled as the Tag cloud fulfils this requirement. Implementation into ATS would be redundant.</i>	#TUD003
#114			The user should be informed about frequently occurring authors during the research.	<i>Cancelled as the network graph fulfils this requirement. Implementation into ATS would be redundant.</i>	#TUD010

#115	Recommendation		Getting recommendations of further steps while conducting the search due to the search by other users.	<i>Work in progress (will be part of the curriculum widget).</i>	#TUD122
#116	Going to discussion page		Recommendation of going to the discussion page of the source/topic.	<i>Work in progress (will be part of the curriculum widget).</i>	#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.	<i>Implemented</i>	#EY052

2.9 Requirement: community

Table 10: 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.	<i>Work in progress</i>	#TUD033
#119		Asking questions	Possibility of asking questions.	<i>Work in progress (with the forum function)</i>	#TUD121
#120			It should be possible to inform one about other researchers.	<i>Implemented</i>	#TUD009
#121	Pictures		Sharing found pictures due to the search on the platform with other user of the platform.	<i>Work in progress</i>	#TUD115
#122	Literature		Sharing of found literature due to the search on the platform with other user of the platform.	<i>Work in progress</i>	#TUD116
#123	Ranking		Ranking of found resources due whether or not the result fit to the search query.	<i>Work in progress</i>	#TUD117
#124	Discussion Forum on search results		Possibility of discussing sources found in the search results.	<i>Work in progress (with the forum function)</i>	#TUD076

#125	Discussion Forum on availability of literature		Discussing availability of literature.	<i>Work in progress (with the forum function)</i>	#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).	<i>To do</i>	#TUD124

2.10 Requirement: user management

Table 11: 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.	<i>Work in progress; tracking is implemented</i>	#TUD078
#128	Tagging of queries		Tagging of queries for finding them easier.	<i>To do</i>	#TUD082
#129	Social bookmarking feature		Meta data library for saving and tagging search results with keywords.	<i>To do</i>	#TUD079, #TUD097, #TUD099
#130			Saving the search results.	<i>Cancelled; merged with #129.</i>	#TUD095
#131		Reference management system on the platform	Organising and comparing the search results in a reference management on the platform.	<i>Backlog</i>	#TUD083, #TUD096, #TUD104, #TUD077, #TUD111, #TUD087, #TUD088, #TUD089
#132			Marking the found literature with colours.	<i>Backlog</i>	#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.	<i>To do</i>	#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.	<i>To do</i>	#TUD080
#135	Notes		Writing notes or excerpts on texts/-sources.	<i>Backlog</i>	#TUD084
#136			Copy and paste these notes.	<i>Backlog</i>	#TUD086, #TUD085
#137			Download the notes.	<i>Backlog</i>	#TUD086, #TUD085
#138			Save the notes.	<i>Backlog</i>	#TUD086, #TUD085
#139		Reference management system outside the platform	Exporting the search results into reference management system, which is not included on the platform.	<i>Cancelled; merged with #133.</i>	#TUD098
#140	Upload of resources		Uploading resources.	<i>Work in progress</i>	#TUD110
#141	Downloading resources		Downloading resources.	<i>Work in progress</i>	#TUD109
#142	Managing access of uploaded resources		Managing access of uploaded resources into open access or not.	<i>Work in progress</i>	#TUD110
#143	Collaborative text creation		Creating of texts, which can be accessed collaboratively.	<i>Cancelled due to technical complexity</i>	#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).	<i>Work in progress</i>	#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.	To do	#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).	To do	#EY015
#902	Search history		Display user's last 5 searches on search start page.	To do	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.	To do	#TUD007

2.11 Requirement: accessibility of the platform

Table 12: 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.	Work in progress	#TUD107

3 Design of the user studies

Following from the implementation presented in D1.2, the evaluation of the MOVING platform requires the design of two use case scenarios, with a common set-up. Although the tasks to be carried out differ between scenarios, we have kept them as similar as possible by gathering the same metrics, and using the same apparatus.

Before starting the study, participants have been provided with a tutorial document (see Appendix A.1). This tutorial covers the basic elements of the MOVING platform, so participants are already familiar with the tool, palliating learning effects.

Due to the ever changing nature of the project, the implementation of the user studies presented in Section 4 from D1.2 needs to be adapted. For each of the use case scenarios the necessary changes to the tasks are explained, detailing the target profile for the participants, and explaining the actual tasks provided to the users.

3.1 Metrics

Participants' success rate as well as the time taken to carry out each task has been gathered for all tasks. Participants have also been asked to answer a set of questionnaires throughout the study. Questionnaires gathering the expected and perceived difficulty have been employed to test if the platform lived up to participants' expectations. Additionally, participants have also been asked to answer a questionnaire measuring the ease of use after the execution of the tasks, and a more extensive questionnaire has been used at the end of the study to measure the overall usability of the platform. Qualitative information has also been gathered by a researcher present in the room at the time of the study.

Effectiveness and efficiency. These metrics serve as a basic way to determine how difficult it was for the participants to carry out the given tasks. Although efficiency provides a direct metric (i.e. the amount of time taken to complete a task), effectiveness can cover varying degrees beyond the binary success/failure. Among failed executions, participants might not be able to find a way to carry out the task, while other participants managed to carry it out partially. This information is of particular interest to find out which steps of the execution are particularly problematic for participants.

Expected vs Perceived difficulty. In order to make sure the platform matches participants expectations, the expected difficulty was reported before each task (Albert & Dixon, 2003). As recruited participants are expected to be familiar with similar information acquisition tools (such as Google, or Web of Science), as well as familiar with the MOVING platform (a tutorial is provided before the study), they are assumed to be able to provide a measurement of the expected difficulty for the given task.

In order to gather this measurement, for each of the tasks, the participants are asked to read the task, to then report how difficult they expect the task to be (*How difficult do you expect this task to be?*, i.e. before). Then, after carrying out the task, they are asked to report how difficult they perceived the task to be (*How difficult was this task?*, i.e. after). For both questions the participants were asked to answer using a 7-point likert scale ranging from "Very Difficult" to "Very Easy".

Ease of Use. After performing each of the tasks, participants have been asked to fill in a ease of use questionnaire (Brinkman, Haakma, & Bouwhuis, 2009). This questionnaire contains 6 questions (randomised order, see Table 15 in Appendix A.2) and is answered using a 7-point likert scale ranging from "Extremely unlikely" to "Extremely likely". The answers are then averaged across all questions for each participant.

Usability questionnaire. At the end of the study participants have been asked to fill in a SUS questionnaire (Brooke et al., 1996). This questionnaire contains 10 statements, with alternating positive and negative wording, and participants report their degree of agreement with them using a 5-point Likert scale ranging from "Strongly disagree" to "Strongly agree" (see Table 16 in Appendix A.2). The questionnaire's outcome is a single score (SUS score) obtained by subtracting "1" to participants' answers to positive statements, and subtracting the participant answer to "5" for each negative statement, and then adding all resulting scores.

This score is useful to compare the perceived usability between two different systems, or between two different user groups. Additionally, the corresponding categorical score can be reported, based on the average responses from participants for other systems (Bangor, Kortum, & Miller, 2009). Figure 1 shows the equivalence between the SUS scores and acceptability ranges, adjective ratings, and a school grading scale. According to this equivalence, a SUS score of 65 would be considered marginally acceptable, receive a "D", and be between "OK" and "Good" respectively for each equivalence.

Think-Aloud. Qualitative data complements quantitative recordings, providing additional context that helps understand the identified problems in the interface. We have used the Think-Aloud protocol (Ericsson & Simon, 1980) to gather participants' feedback throughout the study. A researcher has been present during the execution of the tasks, eliciting feedback from the participants as they carried out the tasks. The document shown in Appendix A.3 has been used to keep the notes taken by the researcher consistent.

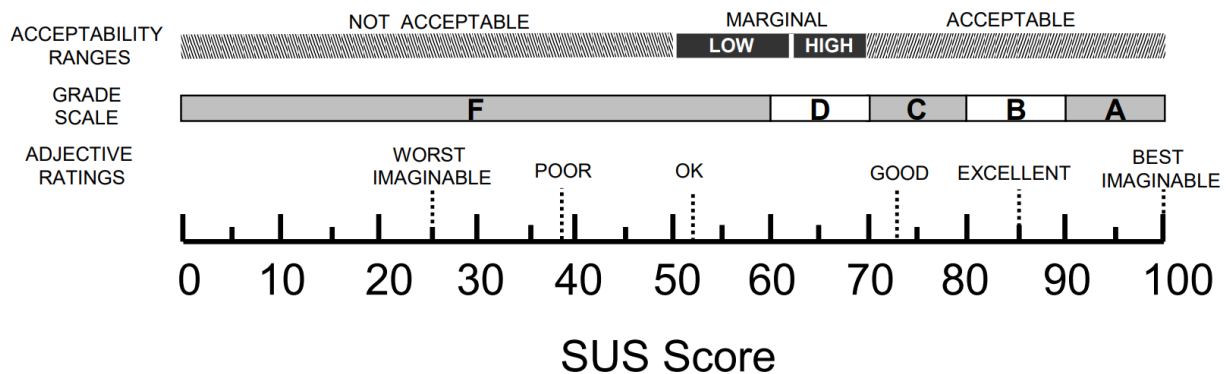


Figure 1: A comparison of the adjective ratings, acceptability scores, and school grading scales, in relation to the average SUS score (Bangor et al., 2009)

3.2 Apparatus

The same set-up has been used for both use cases. The screen and audio during the execution of the tasks has been recorded using OBS Studio³ so the interaction could be replayed to gather further insights. Interaction data has also been recorded using the UCIVIT tool already deployed in the publicly available MOVING platform (see D3.1: Technologies for MOVING data processing and visualisation, Section 4).

3.3 Use case 1: Research on business information by public administrators

3.3.1 Updated task requirements

In the original implementation of the evaluation for this use case, SWOT and PEST analyses involving free interaction from participants were considered as prospective tasks. However, maximising the number of features to be evaluated in this time constraint evaluation required narrowing down the interaction from the participants to the goal modules. Furthermore, the content currently provided by the MOVING platform does not fully support these analyses, so content agnostic tasks were employed instead. The use of search features, filters, and visualisations to explore the data form the core of the final tasks. The final tasks recreate various small scale scenarios of a public administrator carrying out the originally planned analysis tasks.

3.3.2 Participants

The study took place in the offices of EY in Essen. EY professionals have been recruited as the optimal target audience to evaluate this use case (n=7). The tasks designed for this scenario are similar to the tasks these professionals would carry out when on their job.

3.3.3 Tasks

Appendix A.4 lists the full tasks provided to the participants. Each task has been designed so a particular set of features from the MOVING platform would be used. However, participants are free to use the platform in any way they saw fit. The key features expected to be used by the participants for each task have been highlighted.

Task 1: Understand the entity. This task simulates the task an officer would carry out to get an initial overview about a new topic they need to research. The participant is expected to use the **Top Concepts** visualisation to identify topics related to the original search, use the **filters** to narrow down the search, and use the **Concept Graph** to explore relations between the top results.

Task 2: Country by country reporting. Officers are sometimes interested in analysing results per country. As country is not supported as a filter, participants are expected to use **Advanced Search** to look for reports and then use **uRank** to filter them using the keywords corresponding to the indicated countries. This way users can filter the results by any category, not having to rely on existing filters.

³<https://obsproject.com/>

Task 3: SWOT/PEST analysis. Instead of asking participants to carry out a full SWOT/PEST analysis, one of the key steps has been isolated and used as a task. This task requires participants to make use of **Advanced Search** to find a particular annual report containing a strategy plan for the given company.

3.4 Use case 2: Managing and mining research information

3.4.1 Updated task requirements

Due to the ever changing nature of the project, the implementation presented in D1.2 need to be adapted. In this use case, most of the features are ready for evaluation, except from the date filter, which was a requirement for one of the tasks. However, the original tasks required the participant to draw conclusions from the data provided by the platform. Issues with the current state of the data, such as duplicate results, and a mismatch between the search keywords and the results, make an integrity check of the data necessary before the search results can be evaluated with users. Therefore, the original tasks have been altered so MOVING features can be evaluated using content agnostic tasks.

3.4.2 Participants

To find suitable participants for the evaluation of this use case, we have focused on recruiting PhD students. The study took place in the School of Computer Science in the University of Manchester, making it possible to recruit not only computer science students (n=15), but also PhD students from the School of Business (n=5).

3.4.3 Tasks

Appendix A.5 lists the tasks provided to the participants. Similarly to use case 1 (see above) participants are expected to use certain features in each task, which have been **highlighted** below, but they are free to use the platform in any way they want.

Task 1: Get an overview of a topic. This task simulates a young researcher getting an initial overview about a new topic he/she needs to research. The participant is expected to use **Basic Search** to search for a main topic, and then to use **filters** to narrow down the search to a set of subject areas. Finally, the participant would use the **Concept Graph** visualisation to explore the relations between the top results.

Task 2: Find related topics. In this task the participant is free to use any of the available visualisations (**uRank**, **Tag cloud** and **Top concepts**) to identify keywords related to a given search. Once identified, the participant is expected to use **Boolean Operators** in the **Basic Search** in combination with the identified keywords to carry out a new search.

Task 3: Find learning materials. This task simulates a young researcher looking for additional learning resources after getting some insight into the desired topic. The participant is expected to use **Basic Search** and three different **filters**, so the appropriate learning resources can be identified.

4 Analysis of the use cases

Information regarding each one of the use cases is presented here. For each use case, various metrics have been gathered during the study, such as demographics, questionnaires and effectiveness. In the case of the perceived difficulty, expectation and perceived difficulty, reports have been paired and plotted in a single image as shown in Figure 2.

At the end of each use case, a summary of the feedback gathered from users is presented. Next to each feedback, the number of occurrences of that feedback has been indicated (i.e. how many participants mentioned that feedback). The feedback has been classified into five categories: presentation, features, data, relevant quotes and relevant strategies. This categorisation is maintained in later sections for the sake of consistency.

4.1 Use case 1: Research on business information by public administrators

4.1.1 Demographics

This evaluation focuses on the target end-user group of use case 1, financial professionals and public administrators. As a way to match this target group as close as possible, the study has taken part in the EY premises in Essen (Germany). Seven EY employees external to the MOVING project have been recruited (two female, age: mean=28.29 sd=5.19).

4.1.2 Results

Perceived difficulty. The difference between the expected (before carrying out the task) difficulty and the perceived one (after having carried out the task) has been found to be significant (Wilcoxon paired test, $V=32.5$, $p=.010$). The full list of the reported values for each task is shown in Table 13. The order of the tasks did not have an effect in the reported expected (Kruskal-Wallis $\chi^2(2)=1.066$, $p>.5$) and perceived difficulty (Kruskal-Wallis $\chi^2(2)=1.031$, $p>.5$).

As a way to explore these differences Figure 2 shows the correspondence of the expected vs perceived difficulty for each participant and task, using colours to indicate the various sections. The bottom right corner has been coloured in red, as it corresponds to the tasks the participants expected to be easy, but ended up finding difficult to carry out. Occurrences in this area can be considered *critical* issues that should be resolved as soon as possible. The top left corner has been coloured in green, as it corresponds to the tasks the participants expected to be difficult, but ended up finding easy to carry out. Occurrences in this area can be considered the *strong points* of the platform, as they make seemingly difficult tasks easy to carry out. The bottom left corner and the top right corner have been coloured in yellow and blue respectively, and correspond to the tasks where the expected and perceived difficulty remained the same. Although these occurrences are not surprising (participants guessed the difficulty of the task correctly) the occurrences in the yellow area can be considered as opportunities for improvement.

No critical issues could be seen in this case, apart from some outliers. In general it can be seen that participants found the tasks easy to carry out, even if they expected them to be “Difficult” or “Fair”. For tasks 2 and 3, participants perceived the tasks to be easier than they expected. In the case of task 1, there can be seen some exceptions where the perceived and expected difficulty were a match.

Ease of use. Figure 3 shows the distribution of the reported Ease of use 7-point Likert scale values for each task and Table 13 lists the averages for each task. Although an increasing tendency can be seen from task 1 to task 3, it has been found to be non-significant.

Usability. Figure 4 shows the density plot of the seven SUS score gathered for this use case. A vertical red line indicates the mean at 65.36 (sd=16.16). Following the SUS Score adjective rating introduced in Section 3.1 (see Figure 1), this score indicates the participants considered the usability of the platform to be marginally acceptable, between “OK” and “Good”.

Effectiveness and efficiency. Figure 5 shows the distribution of the completion times for each of the tasks. We found the differences between the tasks significant (mixed anova: $F(2,12)=4.896$, $p=.028$). A post-hoc Tukey test revealed the third task is significantly shorter than the other two (Tukey comparison: task3-task1 ($z=-2.460$, $p=.042$) and task3-task2 ($z=-2.905$, $p=.011$)), it should be noted that the tasks are independent from each other (see subsection 3.3.3 for full list of tasks), as they were designed to focus on different aspects of the interface.

Table 13: Use case 1: Expected and perceived difficulties and ease of use reported for all tasks.

	Expected Difficulty		Perceived Difficulty		Ease of Use	
	mean	sd	mean	sd	mean	sd
Task 1	4	1.15	4.71	1.5	5	1.34
Task 2	4.43	0.79	5.43	0.98	5.57	0.79
Task 3	4.29	1.25	5.43	1.5	5.76	0.98
All tasks	4.24	1.04	5.19	1.33	5.44	1.06

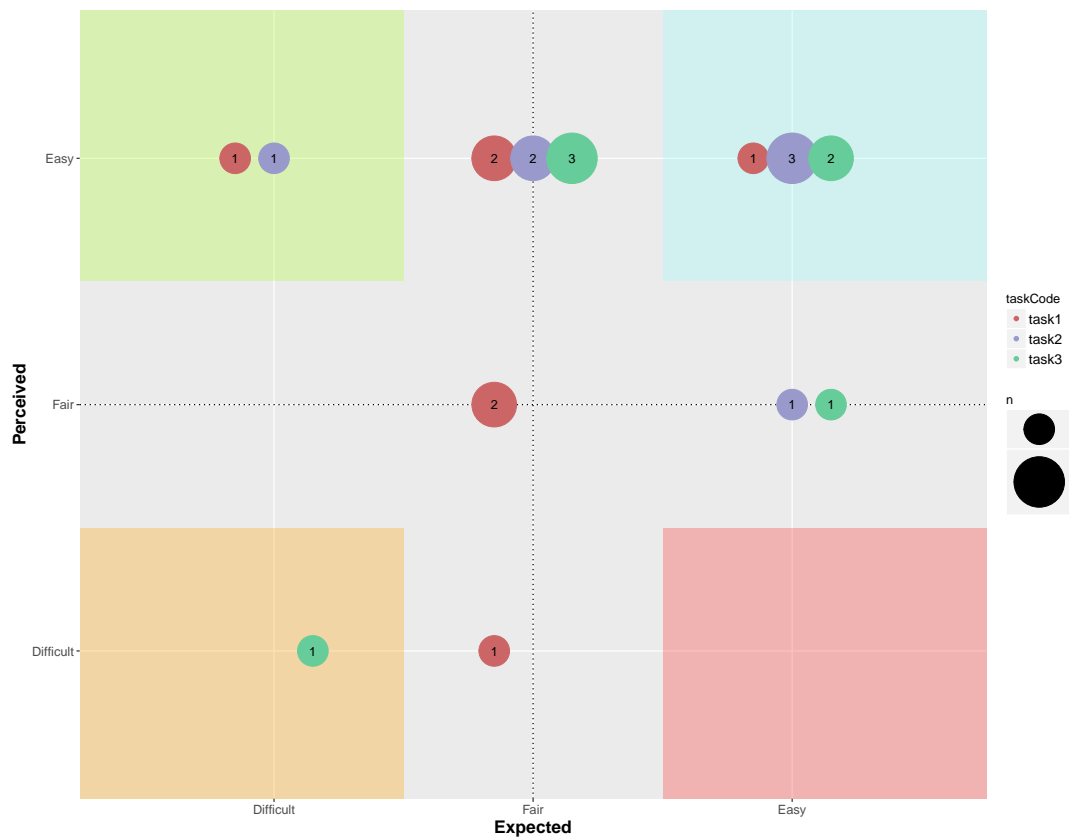


Figure 2: Use case 1: Distribution of expected vs perceived difficulty for each of the tasks.

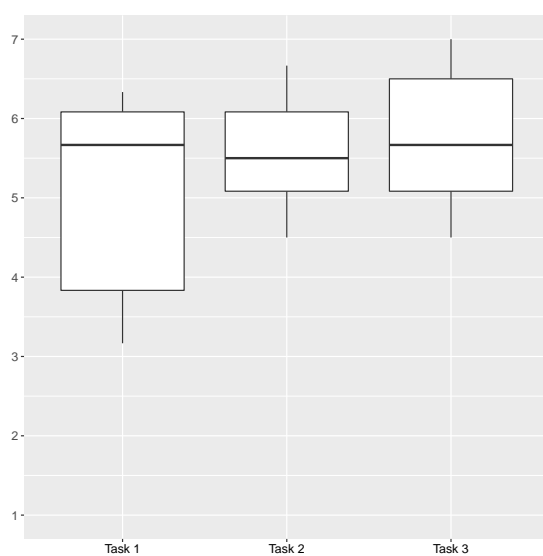


Figure 3: Use case 1: Ease of use boxplots.

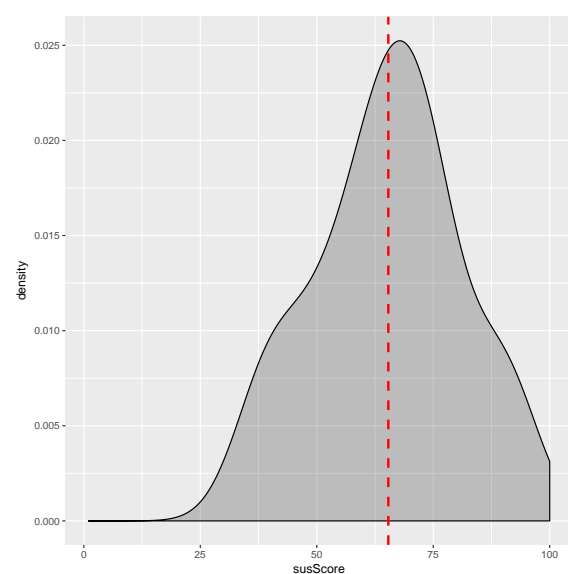


Figure 4: Use case 1: Distribution of the reported SUS scores.

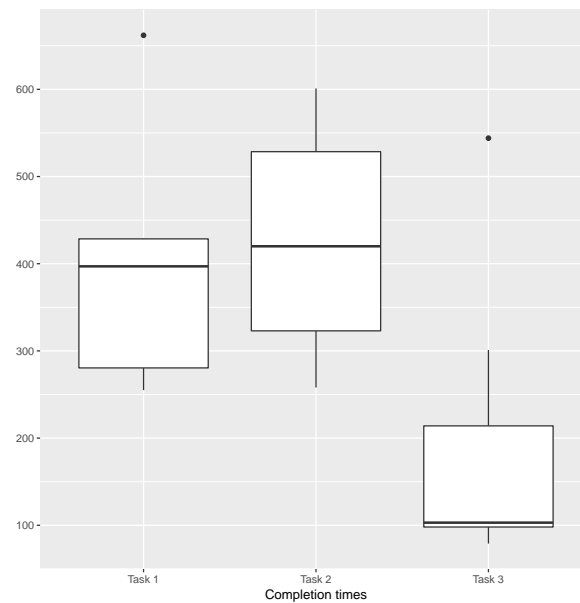


Figure 5: Use case 1: Distribution of the completion times for each task.

Apart from the success rate for each tasks, the followed strategies by each participant has been gathered. This provided insight about particular interface aspects that have been found problematic, and the ones that helped the participants succeed. In **task 1** (median= 397, sd=141.0008), five participants succeeded while two of them failed to select on of the required filters. In **task 2** (median= 420, sd= 136.3513), only one participant failed, when selecting the incorrect result from the list. Finally, in **task 3** (median= 103, sd=172.4185), all participants succeeded.

Participant feedback. This section summarises the feedback gathered from users during the execution of the tasks. It has been classified into five main categories. In Section 5.1.2 observations regarding this feedback are presented, and in Section 5.2 recommendations following the same set of categories are given.

- **Presentation.** Participants have perceived the platform to be slow, and have identified some problems with interface elements.

- **General**

- * It is not clear where to click to open filters, as the click targets are really small (1).
- * It takes a long time to load the results (1).

- **uRank**

- * Participants complained that the overall impression was that the use of uRank slowed down the platform (3).

- **Concept Graph**

- * The tool shows some information when the user hovers a node, which obscures the node, and disrupts the interaction (2).
- * It was difficult to identify newly opened nodes, so a participant suggested changing their colour (1).

- **Features**

- **General**

- * Unrequested page changes, resetting filters and changing search mode (3).

- **uRank**

- * Many participants complained that the search functionality did not work. The cause was that the highlight was short, and the tag was out of the view, not giving enough time to scroll down to find it (7).

- * Several participants mentioned they would like to be able to unselect a single tag instead of all of them (2).
- * A participant suggested the inclusion of a date filter (1).

– Data

○ General

- * Filters were not working correctly, showing participants results from various languages, even if “English” was selected (1).
- * One participant suggested that “Social media & Web” category would be too broad (1).

○ uRank

- * Abstracts were not available (1).
- * Some URLs were the same for more than one result (1).
- * URLs were presented in plain text, instead of links, forcing users to copy and paste them into the browser’s address bar (1).

– Relevant quotes. There were several instances of participants not being sure how to use the tool, and making explicit references to the need of further support, and explanations about the platform’s purpose.

- “I am not entirely sure if the search is case sensitive. It shouldn’t make a difference, unless you use quotation marks.”
- “If I would need support I would just leave and go to Google.”
- “If search fields weren’t given I would have not probably used them like this.”
- “The intention behind the platform is not clear (maybe more on the learning page about it).”

– Relevant strategies. Participants have been seen to click on “Search” as a way to go back to the start.

- Clicking on “Search” in the top bar to reset the state of the search (4).
- Using “Ctrl+F” to find a particular tag (1).

4.1.3 Conclusion and observations

Participants in this use case made heavy use of the visualisations, resulting in useful feedback concerning the Concept Graph and uRank in particular. Participants perceived the platform to be easier to use than they expected, and although there are problems to be addressed, the perception about the platform’s ease of use and usability has been positive.

Even if the number of participants has been low, their profile is an exact match with the use case, adding value to their feedback. One of the main issues identified in this use case has been the unexpected changes in the state of the Web page, along with an overall low performance and non-working features (filters, keyword selection and search functionalities). Errors regarding the presentation of the data, such as missing abstracts, and duplicated results have also been noticed by participants.

Participants of this use case have been particularly interested in additional support for using the platform. Even though there is room for improving the user interface, additional help functionalities would be appreciated by users of this use case.

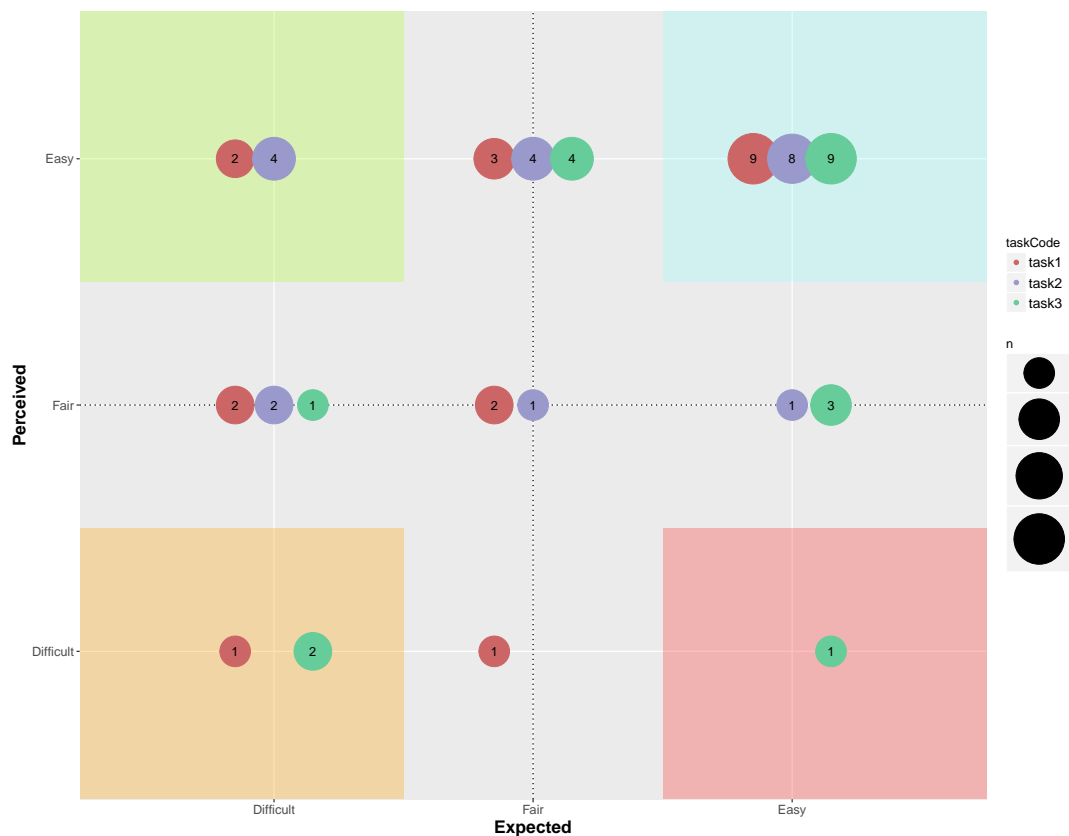
4.2 Use case 2: Managing and mining research information

4.2.1 Demographics

This evaluation focuses on the target end-user group of use case 2, junior researcher in the field of social sciences, humanities, economics, science and technology studies, computer science, or engineering. To evaluate this use case, we recruited PhD students from the University of Manchester. Twenty participants took part in the study (12 female, age: mean = 28.6 sd=4.49), fifteen from the School of Computer science and five from the Business School.

Table 14: Use case 2: Expected and perceived difficulties and ease of use reported for all tasks.

	Expected Difficulty		Perceived Difficulty		Ease of Use	
	mean	sd	mean	sd	mean	sd
Task 1	4.35	1.35	5.15	1.18	5.74	0.9
Task 2	4.25	1.16	5.2	0.83	5.74	0.72
Task 3	5	1.52	5.05	1.57	5.73	0.72
All tasks	4.53	1.37	5.13	1.21	5.74	0.77

**Figure 6:** Use case 2: Distribution of expected vs perceived difficulty for each of the tasks.

4.2.2 Results

Perceived difficulty. The difference between the expected (before carrying out the task) difficulty and the perceived one (after having carried out the task) has been found to be significant (Wilcoxon paired test, $V=320.5$, $p=.005$). The full list of the reported values for each task is shown in Table 14. The order of the tasks did not have an effect in the reported expected (Kruskal-Wallis $\chi^2(2)=3.792$, $p>.05$) and perceived difficulty (Kruskal-Wallis $\chi^2(2)=0.001$, $p>.05$).

As a way to explore these differences Figure 6 shows the correspondence of the expected vs perceived difficulty for each participant and task, using colours to indicate the various sections.

There aren't any critical issues. It can be seen that both task 1 and task 2, which made heavy use of the visualisations, was generally perceived to be easy, even though several participants expected it to be difficult. Task 3, where participants were required to use several filters to find the correct resource, had generally matching expected and perceived difficulties.

Ease of use. Figure 7 shows the distribution of the reported Ease of use 7-point Likert scale values for each task and Table 14 lists the averages for each task. The reported scores can be seen to remain consistently positive, with non-significant differences.

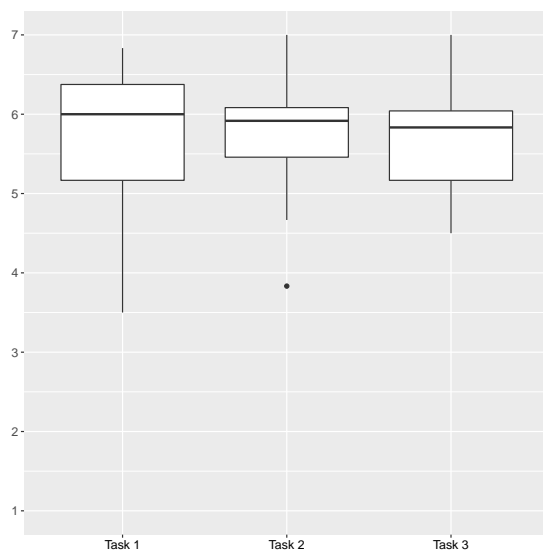


Figure 7: Use case 2: Ease of use boxplots.

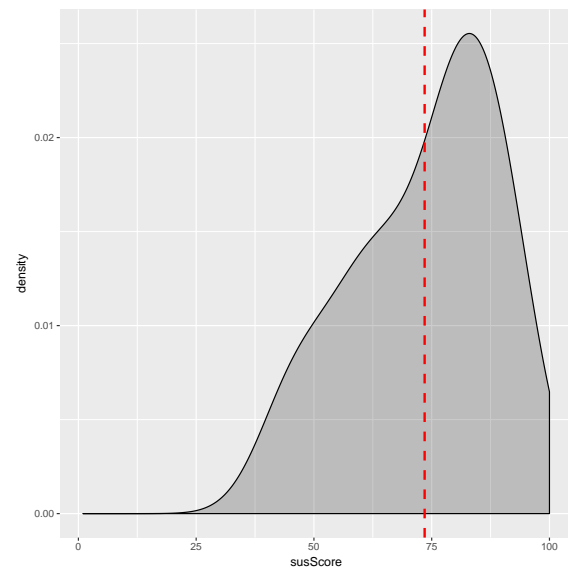


Figure 8: Use case 2: Distribution of the reported SUS scores.

Usability. Figure 8 shows the density plot of the twenty SUS score gathered for this use case. A vertical red line indicates the mean at 73.50 ($sd=15.20474$). Following the SUS Score adjective rating introduced in Section 3.1, this score indicates the participants considered the usability of the platform to be “Good”. As in this use case students from two different domains (Business school and Computer Science) have been recruited, the responses from both domains have been compared in Figure 9. The difference between Business students (mean=74.50, $sd=16.62077$) and Computer Science (mean=73.17, $sd=15.30834$) in the distribution has not been found to be significant (Wilcoxon paired test, $W=39.5$, $p>.5$)

Effectiveness and efficiency. Figure 10 shows the distribution of the completion times for each of the tasks. In task 2 participants were given the freedom to explore as few or many visualisations as they wanted to identify related keywords, hence the outliers. No significant differences have been found between the distributions (mixed anova: $F(2,38)=2.423$, $p>.05$).

Apart from the success rate for each tasks, the reasons behind the failure have been gathered. This provided insight into participants' own understanding about how those particular tasks should be tackled. (see subsection 3.4.3 for full list of tasks), In **task 1** (median= 256, $sd=166.5614$), thirteen participants carried out the task as expected, six participants used booleans instead of or to complement the filters to carry out the task, and only one participant failed to carry out the task by not using any filters at all. In **task 2** (median= 307.5, $sd= 135.9386$), thirteen participants carried out the task as expected, while three participants used lowercase booleans, one participant used the plus symbol “+” instead of a boolean, and three of them failed by not using booleans at all. Finally, in **task 3** (median= 204.0, $sd=165.2495$), twelve participants carried out the task as expected, three used alternative strategies, combining filters, booleans and visualisations to find the correct result, one carried out correctly, but selected the wrong result due to a technical shortcoming in the platform. For this task three participant failed by not selecting a missing filter.

Participant feedback. This section summarises the feedback gathered from users during the execution of the tasks. It has been classified into five main categories. In Section 5.1.1 observations regarding this feedback are presented, and in Section 5.2 recommendations following the same set of categories are given.

- **Presentation.** Participants had problems understanding the interface due to the lack of visual cues. In the case of Concept Graph, participants did not notice which nodes had already been opened.
 - **General**
 - * It is not clear where to click to open filters, as the click targets are really small (14).
 - * Web pages took a long time to load (3).
 - * No visualisation had any loading feedback (2).

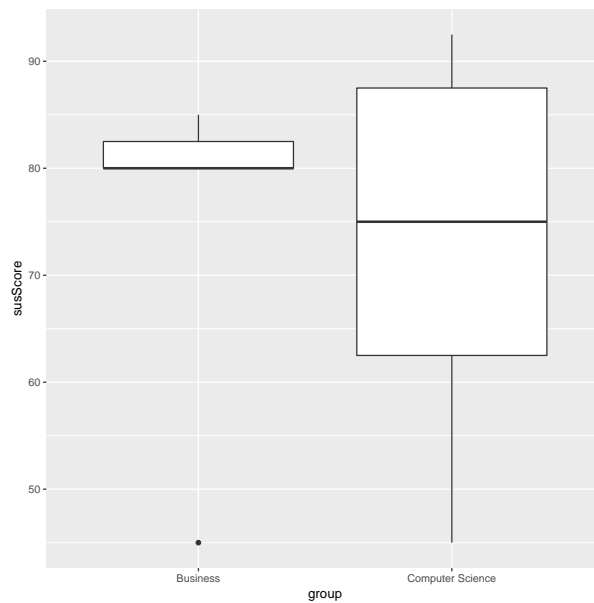


Figure 9: Use case 2: Distribution of the reported SUS scores for Computer Science and Business students.

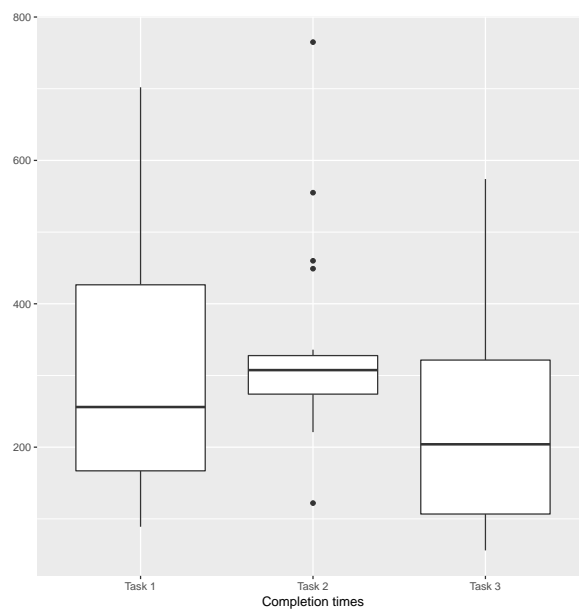


Figure 10: Use case 2: Distribution of the completion times for each task.

- **Concept Graph** Participants found this visualisation particularly innovative and useful, but it was complex, and participants would get easily overwhelmed.
 - * Many Participants clicked on already opened nodes thinking there was still more information to appear (5).
 - * The hovering activated description of the node hides the node, interfering with the interaction (2).
 - * Differences in sizes are not noticeable (1).
 - **uRank** This visualisation has been commonly regarded as a tag cloud (as it shows the top related keywords) with further capabilities to obtain insight into the documents.
 - * Tags were supposed to be clicked, but the interface suggested dragging them (1).
 - * Some participants did not understand the colours for the tags, and suggested having a legend (1).
 - **Tag cloud**
 - * Tag sizes were not recognisable (5).
 - * The colours were found to be too childish (1).
 - * A participant did not initially understand the sorting of the tags, thinking they were sorted by position, rather than size (1).
- **Features.** The unrequested change in the state of the page has been the most common complaints. Participants also requested additional features to manage large amounts of data, and to interact with the visualisations more effectively.
- **General**
 - * MOVING is designed so all the page is refreshed every time there is a request to the server. This feature showed to be problematic to participants. For example, every time a filter is selected, a search request is automatically triggered. As the information contained in the page is not stored, all information from input text fields, as well as the state of other interface elements, such as dropdown elements, is reset (16).
 - * As participants do not know exactly what options are available in the filters and advanced search (although stated in the provided tutorial), many of them looked for filter options in the advanced search (10). Some of them suggested adding, or replacing advanced search options with custom fields, adapted to their search (9).
 - * The number of items available to select as filters was large, making participants struggle to find them. Participants suggested adding custom sorting features, or a global search (7).
 - **Concept Graph**
 - * Participants frequently got overwhelmed when more information than expected appear after clicking on a node. Some sort of undoing the last selection, closing other nodes, or resetting the state of the visualisation to the starting point was requested (7).
 - * A participant showed interest in narrowing down their search to a subset of nodes, discarding the rest of them (1).
 - * The amount of nodes in screen was found to be problematic to manage. Participants struggled to move nodes one by one, and one of them wanted to have more ways to select more than node at a time, such as selecting all nodes related to a given one (1).
 - **Tag cloud**
 - * Participants suggested modifying the default sorting of the tags, using a decreasing sorting, or placing the main tag in the middle (6).
 - * The tags directly related to the search were significantly more frequent than the rest, making the visualisation difficult to interpret. A way to automatically or manually remove certain tags was suggested (3).
 - * Some participants tried clicking on tags in order to add them to their current search, or discover new information, such as relation between shown tags. This same behaviour has been noticed in the Top Concepts visualisation (2).
- **Data.** Ambiguity in the category labels, as well as the lack of information in visualisations have been the main identified problems in this category.

- **General**
 - * Ambiguity in filter categories. Participants confused venue with dataset. They suggested renaming it to sources of data, or just sources (6).
 - * Relevance is not transparent, and a participant did not want to rely on it (1).
- **Concept Graph**
 - * Add more information to descriptions, such as related keywords (2).
 - * Provide an overview of all information, such as opening all descriptions at once, or showing a summary next to the visualisation (2).
 - * Show small pieces of information next to nodes, such as year, to differentiate them (1).
- **Top keywords**
 - * Too many similar keywords. Merge similar keywords (e.g. whistle blower and whistleblower) and make them not case sensitive (iot and IoT) (3).
- **Relevant quotes.** Participants spoke positively about the platform, explicitly mentioning their favourite features, and found it intuitive, thanks to its similarity with other similar tools. However, participants struggled to interpret certain metrics given by the platform, and some of them avoided using key features that were necessary to carry out the tasks.
 - **Positive comments.** Participants have found the platform useful, with a convenient design, and novel visualisations.
 - * “I like it, I very much like it, it reminds of Web of Science, but richer as it includes other kind of data that goes beyond publications”.
 - * “I think this tool would be useful (...). Finding related keywords is really useful (...). I would use it to find learning resources”.
 - * “I think is a friendly tool, and I liked the concept graph and the uRank, those are the two that I really liked.”.
 - * “I think that I would like to use this system frequently. Yes, especially the concept graph”.
 - * “I like the use of tabs to organise the interface”.
 - **Familiarity.** Some participants quickly developed some familiarity with the tool, that eased the execution of the tasks.
 - * “I already knew I had to use the Concept Graph when I saw you were asking for prolific years”.
 - * “I find myself more confident than in task 1” (participant was carrying out task 2).
 - * “I will just do it as in Google” (participant proceeds to use quotes for their search).
 - * “I found Tag cloud useful, as it’s familiar”.
 - * “I expect the task to be difficult because I tried to do this with Google Scholar for a systematic review already”.
 - * “We will be digging further into the tool so I expect this task to get more difficult”.
 - **Help and Support**
 - * “If I had known subject areas appear later, I would not have used them as keywords”.
 - **Discovering related keywords.** Participants had problems interpreting the metrics for the related keywords.
 - * An end-user said: “The top keywords are not really useful” when he/she noticed that the top keywords were the same keywords used for the search.
 - * “I could not differentiate between keywords and subject areas”.
 - * An end-user said: “I cannot judge how relevant the number is. This says 5, but among how many documents, what happens if this is only for one document, that’s not relevant”. while he/she was looking at the frequency count reported by one of the visualisations.
 - * “What sort of insight should this give me, the occurrence? Then you have the number, 5, 7, I thought it would be more, because for example in uRank Internet it occurs 54 times. Oh wait, it says 54 items that have IoT in them. And then I go to Top keywords and there is 3. Unfortunately I cannot interpret what the number represents”.
 - * “I thought it would have changed it according to their frequency, as I cannot tell if the font size is different or the same” (Tag cloud).

- * "It says how many times the work was repeated but... Where? In articles? In which domain?"
- **Personal preferences.** Some participants expressed their preference between the various features in the platform.
 - * "Tag cloud offers similar results to uRank, but uRank looks more useful than the others. Tag cloud looks useful, but doesn't provide further insight".
 - * "Out of the uRank and Tag cloud, uRank is the best".
 - * "The Tag cloud is similar to the uRank".
 - * "Really like this one (uRank). Is more clear and gives me a lot of information. I can choose the keywords I am interested on. More interactive than top keywords. (...) Especially if I am new to the domain, this one (uRank) would be very useful".
 - * "Shows concepts (Tag cloud), but how are they related to my search of internet of things? In uRank I get too many duplicates, but I also get the description for the results".
 - * "Top concepts is more to see how frequent they are, but uRank is really helpful if you are new".
 - * "I can choose between the three? I will only use top concepts then".
 - * "Actually I want to use uRank because it was the more straightforward one".
 - * "I am not that used to visualisations so this task was easier than the other 2, I prefer text".
- **Boolean operators and filters.** Some participants explained how they avoid using boolean operators and filters in their searches.
 - * "If I am looking for something I'd just say 'trade unions in Chile', I'd never put AND, OR, maybe is because I am not a good searcher. (...) I feel that when I use the word AND or OR it can include those words as part of the query, (...) because I don't understand the technological language".
 - * "I'm just not in the habit of using boolean operators".
 - * "I didn't like the filters. In my daily life I don't use filters. When I shop online I never use filters." When suggested that the lack of understanding of how it works internally could be the cause, participant replied "I have trust issues with filters, not only with these tools".
- **Relevant strategies.** Participants have used features included in the web browser, and alternatives to the use of filters and the instructed booleans, that could be natively supported by the MOVING platform.
 - **Filters**
 - * Using keywords and booleans for subject areas instead of filters (10).
 - * "Ctrl + F" to find the appropriate tag (2).
 - **Search**
 - * Use auto-complete for the search. This particular feature was natively supported by the browser based on previous searches (10).
 - * Used alternatives instead of booleans, such as lowercase booleans, or the "+" symbol (5).
 - * Didn't use booleans at all (4).
 - * Use of quotes (3).
 - **Visualisations**
 - * "Ctrl + C" to copy a tag (2).
 - * "Ctrl + F" to find a tag in uRank (1).
 - * Even though the task explicitly recommended to open the visualisations in several tabs, one participant still wanted to use a single tab (1).

4.2.3 Conclusion and observations

Participants in this use case have made use of the filters, the Concept Graph visualisation, and had the option to use as many of the remaining visualisations. Participants expected tasks involving the use of visualisations to be more difficult than they actually were, and overall all tasks have been considered positively regarding their ease of use.

The number of participants for this use case (20 from two different schools) has facilitated gathering extensive feedback about the platform's functionalities. Although content specific tasks had to be avoided,

participants expressed their satisfaction with the variety of content to be provided by the platform. The addition of innovative visualisations, like the Concept Graph and uRank, have been perceived as key features that separate the platform from similar information seeking tools. Although MOVING is a sophisticated platform that would benefit from further support functionalities, participant ranked its usability positively, and found the interface intuitive.

The sudden unrequested changes to the state of the page, and problems in the interaction with the visualisations are key issues identified in this use case. Additionally, a set of unexpected uses of the interface have been identified, which could be supported by the platform. For example, the use of alternative boolean operators, and the suggestion of filter keywords in basic searches.

5 Evaluation report and next steps

As an outcome of the user study, a report on the platform's current usability for each use case is presented. Based on these discoveries, a set of recommendations to improve the platform is presented.

5.1 Current state of the MOVING platform's usability

5.1.1 Use case 1: Research on business information by public administrators

Results of the analysis of this use case show that even though participants considered the platform to be useful, there are some critical issues to attend. The tool has been reported to be slow, and there have been issues that interrupted participants' actions. Unrequested updates to the interface, hiding relevant information on the screen (Concept Graph node information), resetting the state of the page (filter selection), have been seen to confuse users, distracting them from the execution of the task.

Several participants were under the impression that they were not using the tool correctly. Additional feedback in the form of descriptions of the various interface elements, is necessary to prevent participants from needing to look for external help resources, particularly in the case of visualisations. Feedback for the loading visualisations, and when automatic updates are triggered (such as uRank highlighting searched tags) would also help guiding the users, as they can avoid making them think something is not working.

Although the differences between the expected and perceived difficulty has shown no critical issues (see Figure 2), there are occurrences where the perceived difficulty was equal or higher (more difficult) than expected. Task 1 (three occurrences of participants perceiving it fair, or difficult), where a combination of visualisations and filters is necessary, would benefit from better ways to handle the Concept Graph and ways to find the correct filter. In tasks 2 and 3, apart from three outliers, where the perceived difficulty was marginally higher (expected easy, perceived fair) or matched (expected difficult, perceived difficult), the perceived difficulty has been lower overall, indicating that the uRank and the Advanced Search had a positive reception.

5.1.2 Use case 2: Managing and mining research information

Results of the study show that participants appreciated the features provided by the platform. Even though the data content itself was not within the scope of the evaluation, the possibility of exploring a wide range of sources has been of interest to the young researchers taking part in the study.

Some of the most recurring comments involved problematic yet easy to fix problems. For example, most of the participants (14 out of 20) would have benefitted from enlarging the interactive area of the filter dropdown. Other issues have been clearly caused by the amount of data offered in the platform. Showing long lists of filtering options and keywords to participants have been found to be problematic. Participants struggled to navigate through them, and while some of them resorted to external features to find the desired word ("Ctrl+F"), most of the participants ended up systematically checking each one of the options. Participants found the variety of offered data appealing, so as more data is yet to come, the MOVING platform needs to carefully consider how to make the results of the search as accessible as possible to participants.

The lack of transparency from the platform arose as a problem when some participants expressed their preference not to use booleans or filters. As they did not know exactly how the tool would include this options into the search, they decided to ignore them completely.

The comparison between the expected and perceived difficulty has show not critical issues. However, task 3 has been found to be more difficult than expected for some participants. This task makes heavy use of the filters, and participants' feedback has shown that finding the correct filter was problematic. In the case of task

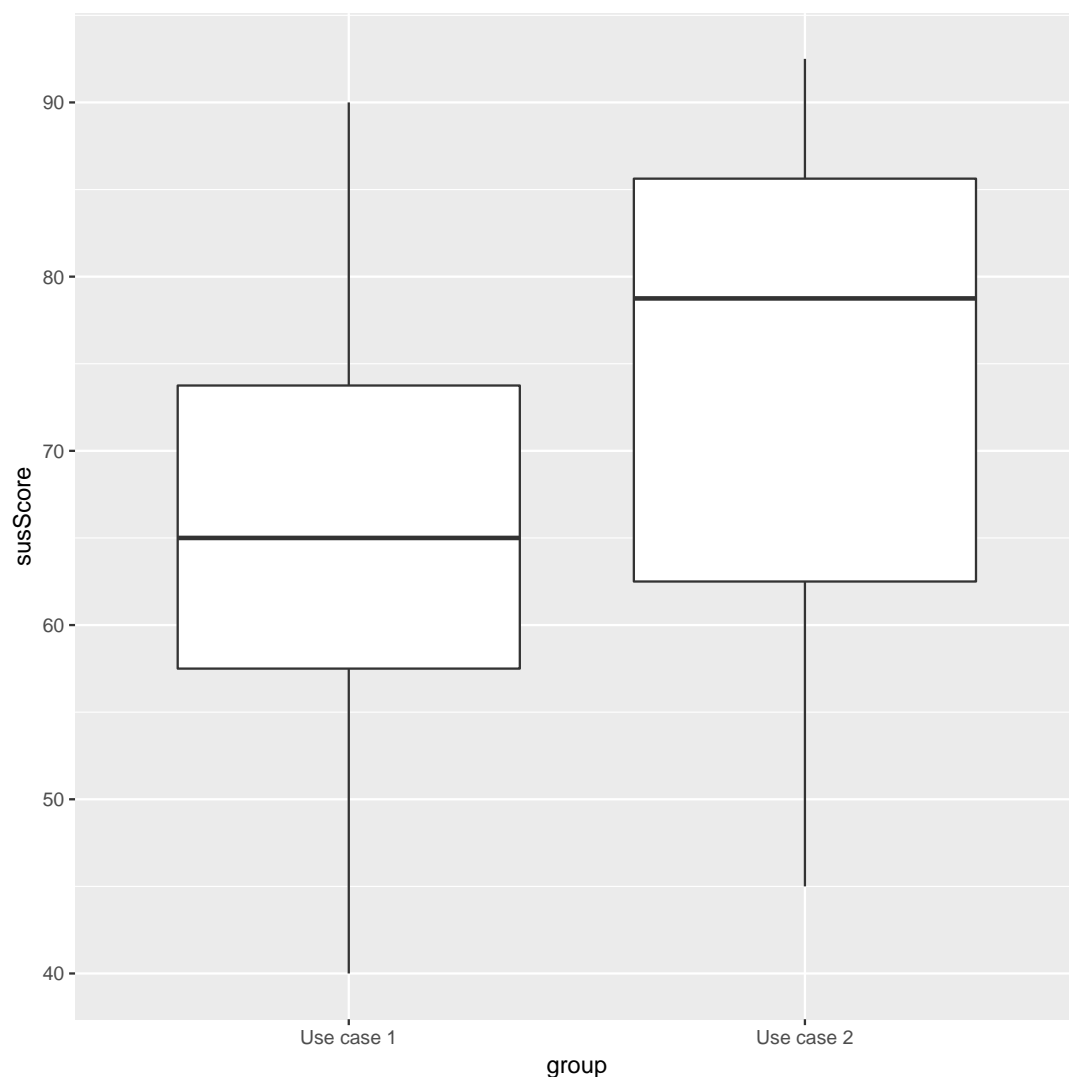


Figure 11: Distribution of SUS scores for use case 1 and use case 2.

2, results indicate that the visualisations used to identify keywords (Top concepts, Tag cloud, and uRank) had a positive reception.

The results of the SUS scores of both use cases have been compared in Figure 11. Although the score received in use case 1 (mean=65.36, sd=16.17) is lower than in use case 2 (mean=73.50, sd=15.20), the difference between the reported scores has not been found to be significant (t-test p-value=0.27).

5.1.3 Summary

The evaluation for both use cases has been positive, with participants rating the provided features positively. Use case 1 participants have been found to be more critical of the platform. This could have been caused by the direct relation of the tasks to their day to day job which would increase their expectations, or that the tasks were more difficult. For example, a participant suggested the inclusion of a date filter (already defined as the requirement #006 in Section 2.2), which would have helped finding the correct annual report in the execution of task 3 (see Appendix A.4). However, the reported perceived difficulty has not been found to be significant (Wilcoxon test, $W=656$, $p>.5$).

5.2 Recommendations for next steps

A set of updated requirements will be provided categorised as presentation, features, and data. For each update, the corresponding code for the requirement reported in D1.1 will be mentioned.

5.2.1 Presentation

R1: Increase size of clickable elements. Particularly in the case of the selection of filter dropdowns, 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 2.5, 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.

5.2.2 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 2.1) 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 user 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.

5.2.3 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.

6 Conclusions

This deliverable updates the list of requirements presented in D1.1. For each requirement we have indicated if its implementation is finished, yet to be processed, or on-going. Some requirements needed to be cancelled, in which case the reason is specified, or postponed.

The design of the user studies has been kept consistent for the two evaluated use cases. Even though the tasks were necessarily different to reflect each use case, the set-up and reported metrics have been kept the same. Metrics regarding the ease of use, differences between expected and perceived difficulty, and the overall usability of the system have been gathered. The original tasks suggested in D1.1 have been adapted to suit the inherent time constraints of user evaluations, and have been made data agnostic. The interaction has been recorded for both use cases, and a researcher has been present during the execution of the tasks to gather participants' feedback. This feedback has provided insight into the interaction problems the participants encountered, helping us to design recommendations for next steps.

Overall participants have evaluated MOVING platform's usability positively. Participants' perceived difficulty has been consistently lower than they expected, and reported ease of use and usability scores have been positive. Participants' feedback has been particularly useful to identify their motivations, and identify opportunities for further improvements and adaptation to prospective users.

Although no critical issues have been identified, important issues need to be tackled to improve the platform's usability. Participants need further support when employing the tool, with further help options, more understandable (and non-ambiguous) feature description, and more transparency, so they can understand and trust the provided results. Finally, a list of recommendations classified into presentation, features, and data has been gathered, and when possible, linked to the corresponding requirement.

References

- Albert, W., & Dixon, E. (2003). Is this what you expected? the use of expectation measures in usability testing. In *Proceedings of the usability professionals association 2003 conference, scottsdale, az.*
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3), 114–123.
- Brinkman, W. P., Haakma, R., & Bouwhuis, D. G. (2009). The theoretical foundation and validity of a component-based usability questionnaire. *Behav. Inf. Technol.*, 28(2), 121–137. Retrieved from <http://dx.doi.org/10.1080/01449290701306510> doi: 10.1080/01449290701306510
- Brooke, J., et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4–7.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological review*, 87(3), 215.

A APPENDIX

A.1 Tutorial

MOVING platform

MOVING is an innovative training platform that enables users from all societal sectors (companies, universities, public administration) to fundamentally improve their information literacy by training how to choose, use and evaluate data mining methods in connection with their daily research tasks and to become data-savvy information professionals.

Search

The screenshot shows the 'Simple search' interface. At the top, there's a header 'Simple search'. Below it, a search bar contains a red 'Research' dropdown button, a text input field with the placeholder 'Search for', and a red search button with a magnifying glass icon. Below the search bar, there are two links: 'Simple search' and 'Advanced search'. At the bottom, a small text line provides document counts: 'Total document count: 2145095 Publications: 1075200 Videos: 21614 EU funding: 47831 HTML webpages: 437936 Social media posts: 0 Other: 562514'.

The basic search of MOVING offers access to the search system. **Boolean operators such as AND, and OR can be used to search for multiple combinations of concepts.**

The dropdown on the left can be used to select **the domain of your search**. You might want to *Research* some topic, look for some *Learning material*, or find some *Funding*.

Advanced Search

The screenshot shows the 'Advanced search' interface. At the top, there's a header 'Advanced search'. Below it, there's a 'Domain' dropdown menu with 'Research' selected. There are four text input fields: 'Title:', 'Abstract:', 'Fulltext:', and 'Person:'. Below these fields is a large red search button with a magnifying glass icon and the text 'Search'. At the bottom, there are two links: 'Simple search' and 'Advanced search'.

Advanced search offers further personalisation into the search query. Here you can describe your query in more detail, specifying which keywords should be found in which fields of the data. Please be aware that some of the results from crawled pages might be missing some of the data. E.g. some documents might be missing the author, even though this author can be found in the resulting page.

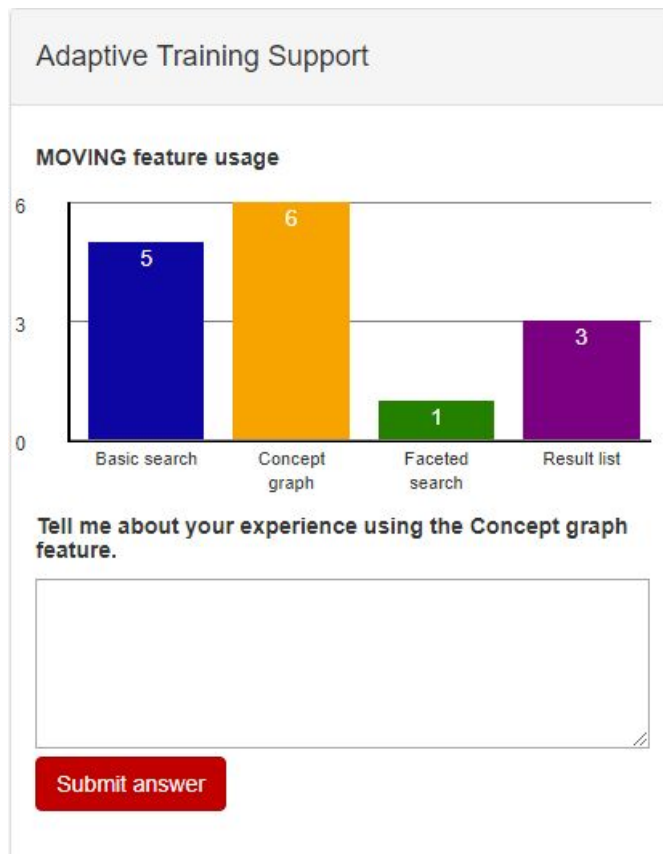
Search results

Once a search has been carried out, a similar screen to the one above is shown.

Left area

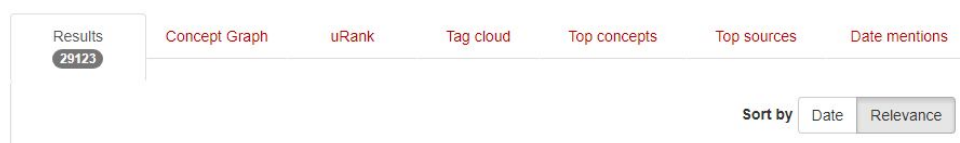
The left area complements the search, allowing you to **narrow down the search** using various filters. These filters can narrow down the search to a particular document type, or focus on a subset of dataset sources (see below). It also shows the number of results **for the original search** (the use of facets does not affect this count) for each displayed group. Each time you select a facet, please wait, as it can take some seconds to reload. Please note, multiple selected facets from one group are combined using an OR filter. This way, the document has to match at least one of the selected facet options. Multiple facets from different groups are combined using an AND filter. For example, (Author1 OR Author2) AND (Doctype2). Furthermore, the author facets are the top 10 authors with most documents retrieved in the current search.

Right area



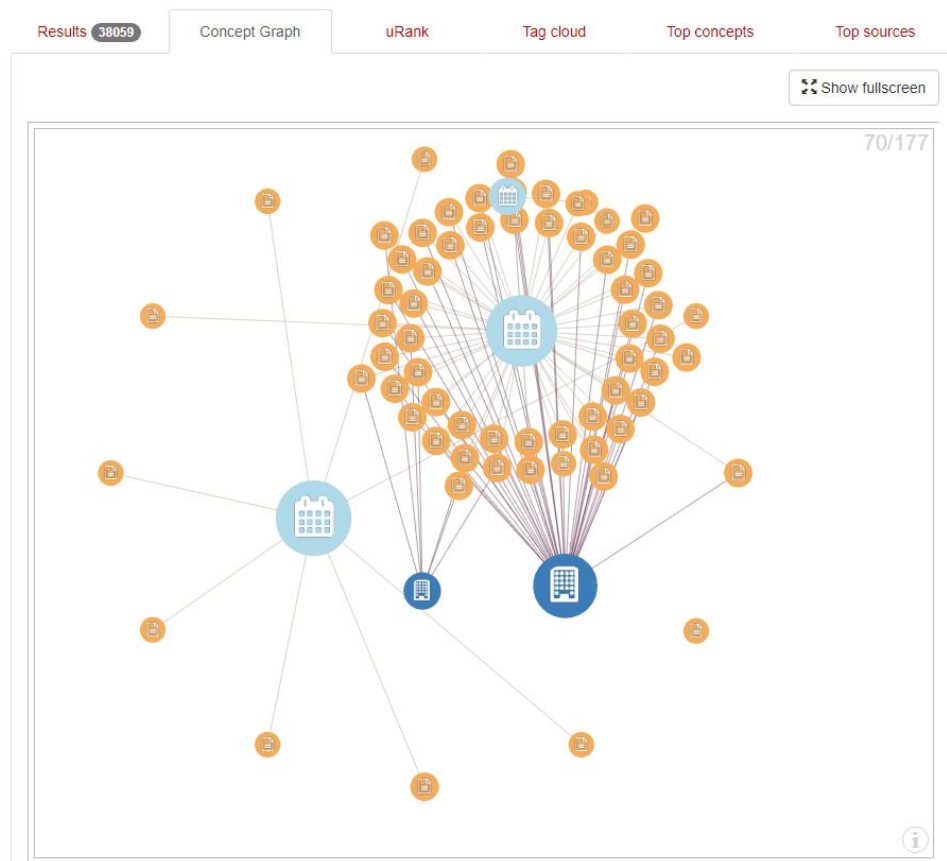
On the right area of the results you will see the Adapting Training Support. This widget tracks your interaction and displays helpful information that will help improve your search habits.

Visualisations



Just below the search dialog, a set of tabs is available to access various features that will help you obtain further insight into the results of your search, highlighting relations between the various concepts.

Concept Graph



Once opened, only the most relevant nodes from the results will be shown. Their size depends on the number of related results for each of them (the bigger the more documents related to it). The upper right corner indicates how many nodes are opened out of the total count. Hovering over them provides you more information about each of them. [This video](#) gives you an overview of the features. Make sure the subtitles are active.

Types of nodes:



Document type: Represents a result from the search.



Author: Represents an author of a document.



Affiliation: Indicates the affiliation of a particular Author. A single author might have more than one affiliation.

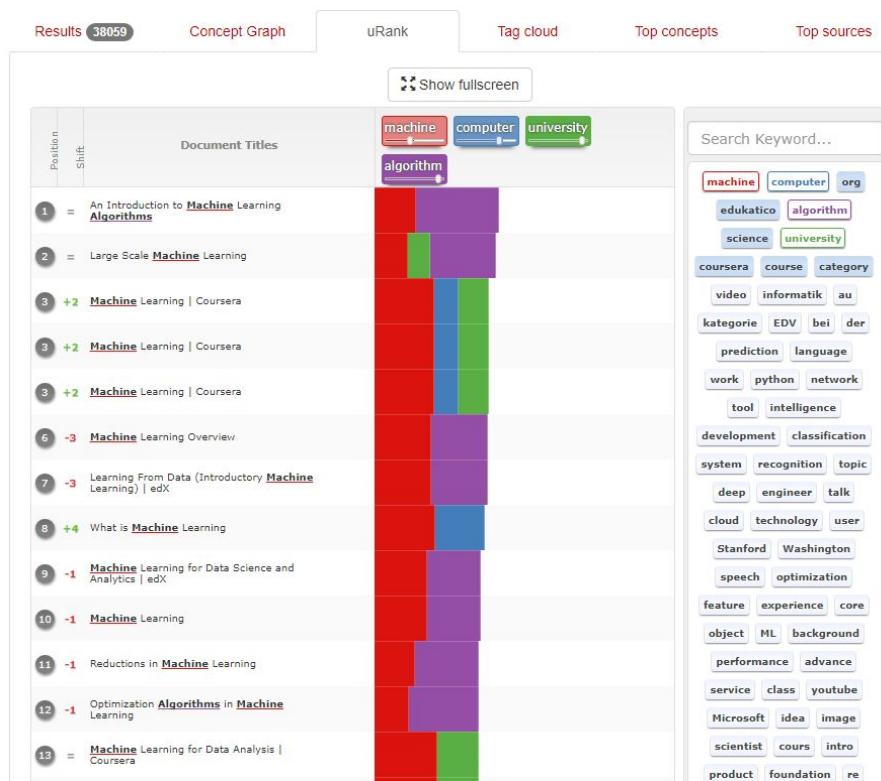


Date: Indicates the date of a particular document. If it's a year rather than a specific date, it will commonly point to several nodes, as there will be many documents sharing the same year.

Interaction with the nodes:

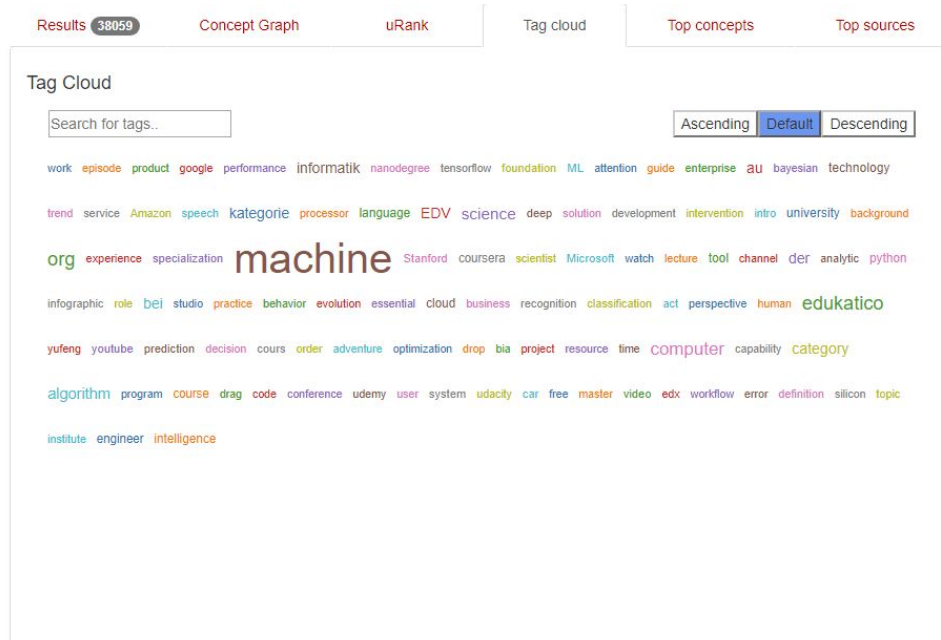
- Left click: open all nodes directly connected to this node.
- Right click: open a ring menu. Selecting a ring will open nodes connected in as many degrees as the level of the ring.
- Double click: open the document in a new tab.

uRank



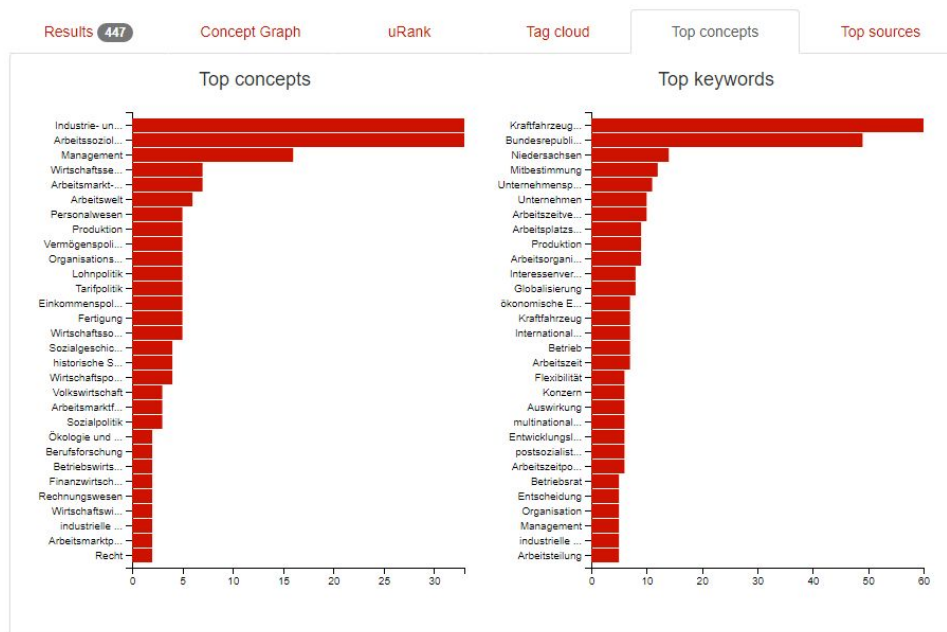
In this visualisation, you can select keywords, and the listed documents will be ranked according to their relevance according to your selection. The weight of each selected keyword can be adjusted using the scroll underneath. The shift column indicates how many positions they gained/lost from their original position.

Tag cloud



The tag cloud gives you an overview of the most important keywords in the results. You can look for a tag using partial names (“on” will show both “Amazon” and “conference” tags) as well as sorting them alphabetically.

Top concepts and Top sources



These visualisations give you an overview of the most popular concepts, keywords, and sources of the 100 most relevant results. You can hover the bars to see the full name of each concept or keyword.

Concepts: they might not be that frequent, as not all documents have a set concept.

Keywords: extracted from the text.

Sources: The resource from which the results came from. For some searches, the source might not have been set, which can result in an empty figure.

A.2 Questionnaires

Table 15: Ease of Use questionnaire

	Extremely unlikely 1	Quite unlikely 2	Slightly unlikely 3	neither 4	Slightly likely 5	Quite likely 6	Extremely likely 7
I would find the tool to be flexible to interact with	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
My interaction with the tool would be clear and understandable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I would find it easy to get the tool to do what I want it to do	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
It would be easy for me to become skilful at using the tool	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I would find the tool easy to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Learning to operate the tool would be easy for me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Table 16: SUS usability questionnaire

	Strongly disagree 1	2	3	4	Strongly agree 5
I think that I would like to use this system frequently	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I found the system unnecessarily complex	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I thought the system was easy to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I think that I would need the support of a technical person to be able to use this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I found the various functions in this system were well integrated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I thought there was too much inconsistency in this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I would imagine that most people would learn to use this system very quickly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I found the system very cumbersome to use	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I felt very confident using the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I needed to learn a lot of things before I could get going with this system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

MOVING

Success?

Bugs

Recommendations

Complaints

Strategies

Quotes (or timestamp of the video recording)

Other

A.4 Use case 1 tasks



Before carrying out the study, please take a look at the [tutorial](https://moving.mz.test.tu-dresden.de/).
<https://moving.mz.test.tu-dresden.de/>

In this study you are going to use the MOVING platform to carry out some simple knowledge acquisition tasks. For each of these tasks you will be asked to fill in a before and after questionnaire, as well as a final questionnaire that will allow us to evaluate the usability of the platform.

You will play the role of a financial professional performing risk assessment procedures for an entity of interest. To simulate this scenario, you will carry out three typical research tasks. For this purpose, you will use the data and text-mining features of the MOVING platform. This platform has indexed various multimedia databases, and complements this information by continuously crawling a set of websites. This way, the tool provides you access to more than 2m documents.

The screenshot displays the MOVING platform's search interface. On the left, there is a sidebar with filter options: Year of publication, Persons, Document Types, Datasets, Language, Venue, Subject Areas, and Access. The main content area shows search results for 'machine learning'. The results are sorted by Relevance and include several entries, each with a title, publication date, and abstract. For example, the first result is 'Machine learning project | Machine Learning' published on 2017-12-18. On the right side of the interface, there is a section titled 'Adaptive Training Support' which includes a bar chart showing 'MOVING feature usage' with values for Basic search (5), Concept graph (6), Faceted search (1), and Result list (3). Below the chart is a text input field for 'Tell me about your experience using the Concept graph feature' and a 'Submit answer' button.



Task 1: Understand the entity

1. Search for **Volkswagen** in simple search
2. Use Top Concepts to look at the keywords visualization to better understand the business of the company. State out loud what are the top three keywords.
3. Use the datasets filter to select the social media results, the document type filter to select web (HTML)/Organization and the language filter to select English.
4. Use the concept graph to find within the relations between the nodes how the results relate to each other. Go to concept graph and use the expansion functionality by right clicking on some of the nodes. Identify which year has a number bigger or equal to five connections with entities and documents.

How difficult do you expect this task to be?

Before you start carrying out the task, remember to answer the questionnaire that the experimenter will show you.

Please carry out the task now

Evaluating the ease of use and the perceived difficulty of the task

Once you can finish the task, please answer the the questionnaire that the experimenter will show you.



Task 2: Country by country reporting

1. Search in advanced mode, for ***country-by-country reporting*** in the title field and for ***financial report*** in the fulltext field.
2. Use uRank to find the top three results on country-by-country reporting in the UK, Denmark and Greece (select tags on the right side to fulfill the task). Select the top one for each of them to access the abstract and the source of the document.

How difficult do you expect this task to be?

Before you start carrying out the task, remember to answer the questionnaire that the experimenter will show you.

Please carry out the task now

Evaluating the ease of use and the perceived difficulty of the task

Once you can finish the task, please answer the the questionnaire that the experimenter will show you.



Task 3: SWOT/PEST analysis

Identify Volkswagen's strategy for 2018.

1. Search in advanced mode for **digital strategy** in the title field and for **volkswagen** in the abstract field
2. Find the annual report for 2015 and read about the 2018 strategy.

How difficult do you expect this task to be?

Before you start carrying out the task, remember to answer the questionnaire that the experimenter will show you.

Please carry out the task now

Evaluating the ease of use and the perceived difficulty of the task

Once you can finish the task, please answer the the questionnaire that the experimenter will show you.

A.5 Use case 2 tasks



Before carrying out the study, please take a look at the [tutorial](https://moving.mz.test.tu-dresden.de/).
<https://moving.mz.test.tu-dresden.de/>

In this study you are going to use the MOVING platform to carry out some simple knowledge acquisition tasks. For each of these tasks you will be asked to fill in a before and after questionnaire, as well as a final questionnaire that will allow us to evaluate the usability of the platform.

You will play the role of a junior researcher interested in the emerging interdisciplinary research field of the “Internet of Things”, wanting to get an overview of existing research and publications on the topic. To simulate this scenario, you will carry out three typical research tasks.

For this purpose, you will use the data and text-mining features of the MOVING platform. This platform has indexed various multimedia databases, and complements this information by continuously crawling a set of websites. This way, the tool provides you access to more than 2m documents.

The screenshot displays the MOVING platform's search interface. On the left, there is a sidebar with filters for 'Year of publication', 'Persons', 'Document Types', 'Datasets', 'Language', 'Venue', 'Subject Areas', and 'Access'. The main search area shows a search bar with 'machine learning' entered. Below the search bar, there are tabs for 'Simple search' and 'Advanced search'. The search results are displayed in a list format, showing the title, publication date, and abstract for several documents related to 'Machine Learning'. On the right side, there is a section titled 'Adaptive Training Support' which includes a bar chart showing 'MOVING feature usage' and a text input field for user feedback.

Feature	Usage Count
Basic search	5
Concept graph	6
Faceted search	1
Result list	3



Task 1: Get an overview of a topic

As a young researcher new to the area, you will first want to get an overview. You are particularly interested on the technical advancements made on the internet of things as well as how it affects the development of cities. To narrow down your search, you will search for the topic "Internet of Things" for the subjects areas of "Science", "Technology", and "Architecture".

Then, you will use the "concept graph" to explore the relationships between documents, authors and publication years. In this step, try to find a particularly prolific year for the topic you are researching. Look for a year in which there are at least 3 of the most relevant documents from your search.

How difficult do you expect this task to be?

Before you start carrying out the task, remember to answer the questionnaire that the experimenter will show you.

Please carry out the task now

Evaluating the ease of use and the perceived difficulty of the task

Once you can finish the task, please answer the the questionnaire that the experimenter will show you.



Task 2: Find related topics

Now that you got an overview of the topic, you will try to discover other possible topics related to your original search. You can then use them as suggestions to narrow down or extend your original search, thus supporting an iterative exploratory approach.

Start by searching for the topic "Internet of Things" again. You are now going to use some of the available visualisation features to find related keywords. You can use uRank, Tag cloud, and Top concepts. As these visualisations can take a while to load, feel free to open them in separate tabs (using the middle button of the mouse).

After identifying some new related concepts, carry out the original search for Internet of Things, adding one or many of the newly identified concepts. Remember that we are not looking for combinations of all the keywords, but rather the occurrence of several keywords in the same document.

How difficult do you expect this task to be?

Before you start carrying out the task, remember to answer the questionnaire that the experimenter will show you.

Please carry out the task now

Evaluating the ease of use and the perceived difficulty of the task

Once you can finish the task, please answer the the questionnaire that the experimenter will show you.



Task 3: Find learning materials

Now that you got an overview and were able to find related topics, you are going to narrow down the search to find a particular resource. You are interested in learning more about the topic, so you will look for an online learning resource on the topic, in your language.

Look for “Internet of Things” again, but focus on documents of type Web (HTML)/Learning, in English, and retrieved from Social Media and the Web. From the results, select the most relevant one.

How difficult do you expect this task to be?

Before you start carrying out the task, remember to answer the questionnaire that the experimenter will show you.

Please carry out the task now

Evaluating the ease of use and the perceived difficulty of the task

Once you can finish the task, please answer the the questionnaire that the experimenter will show you.