Deliverable 7.3: Activities and outcome of the Pilots, final report

Gerhard Rudinger (APA-IT)

Evlampios Apostolidis, Vasileios Mezaris (CERTH)

Arno Scharl (WLT)

21/12/2018

Work Package 7: Pilots, Testing and User Feedback

InVID - In Video Veritas: Verification of Social Media Video Content for the News Industry

Innovation Action

Horizon 2020, Research and Innovation Programme

Grant Agreement Number 687786

Dissemination level	PU
Contractual date of delivery	31/12/2018
Actual date of delivery	21/12/2018
Deliverable number	D7.3
Deliverable name	Activities and outcome of the Pilots, final report
File	invid_d7.3_v1.0.docx
Nature	Report
Status & version	Final, v1.0
Number of pages	81
WP contributing to the deliverable	WP7
Task responsible	APA-IT
Other contributors	CERTH WLT
Author(s)	Gerhard Rudinger / APA-IT Arno Scharl / WLT Evlampios Apostolidis, Vasileios Mezaris / CERTH
Quality Assessors	Symeon Papadopoulos / CERTH Roberto García González / UdL
EC Project Officer	Alberto Rabbachin
Keywords	Testing, evaluation, user feedback

Abstract

In WP7, the technologies integrated into the InVID platform and applications have been tested and evaluated in various editorial cases and trials. This document reports the results of test cycles 7 to 9, which have been held between M27 and M36 of the project. The first part of the document gives a general description of the test cycle concept in the InVID project. This is followed by an overview of the tested components and applications within the project's test cycles 7 to 9 and also the evaluation methods that have been applied in the test cycles for these components and applications. The main part of the document reports the results of test cycles 7 to 9. This part is structured according to the different components and applications. For each component and application, the focus of the tests, the number of feedback comments and the major outcomes of the test cycles are reported. The different InVID applications received very good feedback from the testers. The InVID tools helped them in the task of verification and rights management for user-generated videos. The last part gives a summary of the pilot testing throughout the project’s lifetime.

Table of contents

List of Figures

List of Tables

1 Introduction

In WP7, the technologies integrated into the InVID platform and applications have been tested and evaluated in various editorial cases and trials. The primary aim has been to collect user feedback both on the tools and components themselves, their usability and appropriateness for various tasks, and the results of the system as a whole (e.g. in terms of reliability and accuracy).

Using an iterative approach over the entire project duration (with overall nine validation cycles), the feedback and results obtained from each cycle of the trials were used to improve the subsequent versions of the applications, platform and components. Tests and evaluations focused on video news emerging from social networks and media websites.

In full accordance with the InVID DoA (Description of Action), this deliverable extends the reporting of D7.1 and D7.2 by covering the pilot testing carried out as part of the project's development and validation cycles 7 to 9, which have run from M28 to M36.

1.1 History of the document

Table 1: History of the document

Date	Version	Name	Comment
1/12/2018	V0.50	Gerhard Rudinger	Initial version
27/11/2018	V0.60	Arno Scharl	Results of the InVID Dashboard evaluation
29/11/2018	V0.70	Evlampios Apostolidis, Vasileios Mezaris	Results of the video sub-shot fragmentation and keyframe extraction evaluation
1/12/2018	V0.80	Gerhard Rudinger	Version ready for QA
10/12/2018	V0.81	Roberto García González	QA
10/12/2018	V0.82	Symeon Papadopoulos	QA
14/12/2018	V1.00	Gerhard Rudinger	Final version

1.2 Glossary of acronyms

Table 2: Glossary of acronyms

Acronym	Explanation
API	Application Programming Interface
JSON	JavaScript Object Notation
UGC	User-Generated Content
UGV	User-Generated Video
(G)UI	(Graphical) User Interface
URL	Uniform Resource Locator
REST	Representational State Transfer
XML	Extensible Markup Language

2 Overview of test cycles seven to nine

2.1 Objectives of WP7

As discussed in D7.1 (see Section 3) and D7.2 (see Section 2), through the work carried out in WP7, the technologies integrated into the InVID platform and applications were tested and evaluated in various editorial cases and trials. This was done with the help of different user groups. The primary aim of the conducted evaluations was to collect the users' feedback about the usability and appropriateness of the exposed tools and components for performing various tasks related to the collection and management of newsworthy user-generated content, and the overall effectiveness of the InVID system in terms of reliability and accuracy.

Driven by the objectives of the InVID project, the performed tests and evaluations were focused on the analysis of newsworthy videos distributed via social networks and media websites, as well as on user-generated content provided by users' communities that had been developed by regional newspapers.

2.2 General description of the test cycles

In InVID we followed an agile development methodology which included nine test and validation cycles. The findings of each evaluation cycle and the feedback collected from the participants (which could be both internal users from the InVID consortium and external users from outside the consortium) were exploited to improve the following versions of the exposed analysis components, integrated applications and the overall InVID platform.

The test cycles 7 to 9 took place from month 28 to month 36 (see Table 3).

Table 3: Time plan of the project's test and validation cycles

Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Test cycle							Pilots prep.			Test cycle 1			Test cycle 2			Test cycle 3			Test cycle 4			Test cycle 5			Test cycle 6			Test cycle 7			Test cycle 8			Test cycle 9		

2.3 Overview of test cycles 7 to 9

2.3.1 Applications tested in test cycles 7 to 9

In general, we have included as many applications and components as possible in each test cycle. In the final phase of the project, some components had already reached a mature state and the further development has covered only minor changes or has already been finished within the project. These components have been skipped in the respective test cycles.

Figure 1 shows the overall InVID architecture diagram. All components that have been evaluated in the test cycles 7 to 9 are marked with a solid red box. Some analysis components that are not directly accessible through individual user interfaces have been tested through the InVID applications that integrate these components.

Table 4 lists the components evaluated in each test cycle.

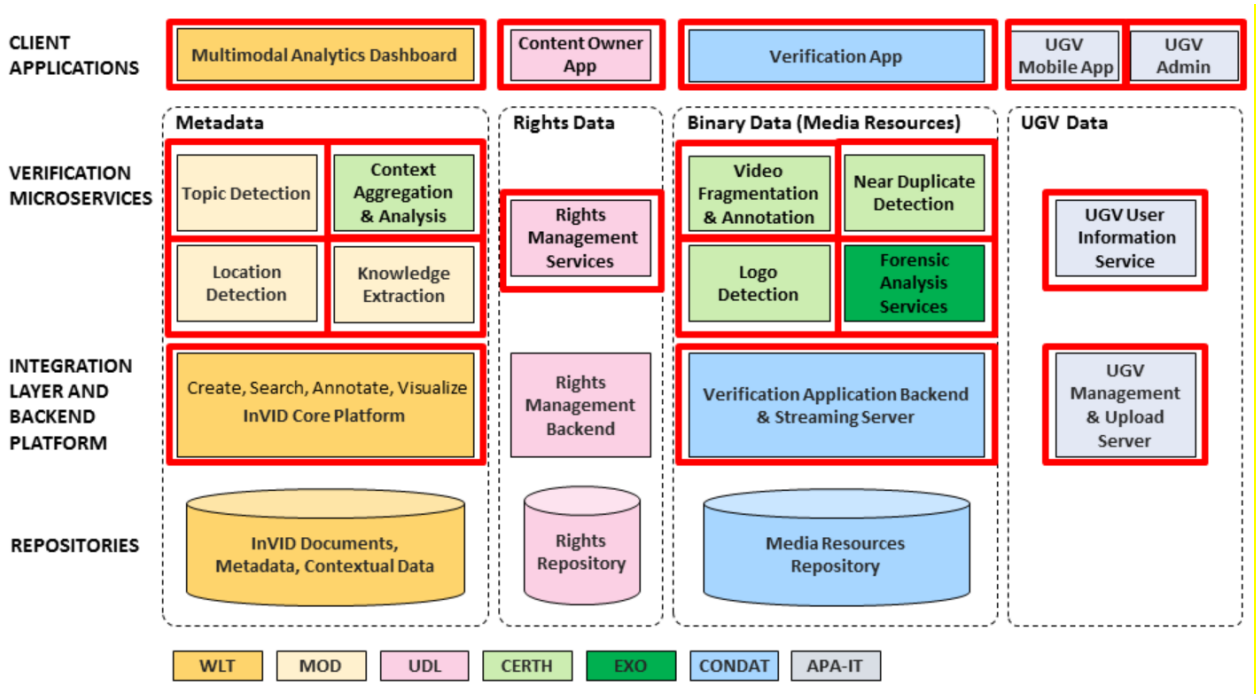

Figure 1: Tested components in test cycles 7 to 9

Table 4: Tested components in test cycles 7 to 9

	Test cycles		
Applications	**7**	**8**	**9**
Video Fragmentation & Annotation Service	Yes	Indirect	Indirect
Web Application for Video Fragmentation & Reverse Image Search	No	Yes	No
Near Duplicate Detection Service	Indirect	Yes	Indirect
Logo Detection Service	Indirect	Indirect	Indirect
Forensic Analysis Service	Yes	Indirect	Indirect
Context Aggregation & Analysis Service	Yes	Yes	Indirect
Rights Management Service	Yes	Yes	Yes
Tool for Social Media Retrieval and Topic Detection	Yes	No	Yes
InVID Multimodal Analytics Dashboard	Yes	No	Yes

InVID Verification Plugin	Yes	Yes	Yes
InVID Verification Application	Yes	Yes	Yes
InVID Mobile Application	Yes	Yes	Yes
InVID Core Platform API	Indirect	Indirect	Indirect

In the table above, "indirect" testing means that the corresponding component has been tested through its integration into the Verification Plugin and the Verification Application.

2.3.2 Testing methods used in test cycles 7 to 9

A detailed overview of existing testing and evaluation methodologies has been given in Section 2 of D7.1. From the reported methods, the following ones have been used in test cycles 7 to 9:

Targets of testing:

- **Integration testing:** Integration and interface testing has been applied to all individual analysis components.
- **System testing:** Testing of the systems as whole has been used for the integrated tools and applications that are accessible mostly via user interfaces, i.e. the Verification Plugin, the Verification Application, the Mobile Application and the Multimodal Analytics Dashboard.

Remark: Unit testing is part of the development process and is not addressed in WP7.

Objectives of testing:

- **Alpha testing:** All exposed applications and components have been tested by users within the consortium.
- **Beta testing:** Parts of the integrated tools and applications, namely the Verification Plugin, the Multimodal Analytics Dashboard, the Verification Application and the Mobile Application, have been tested by potential and/or existing users/customers.
- **Functional testing:** In all test cycles the functionality of the components/applications has been tested.
- **Non-functional testing:** In all test cycles the non-functional aspects of the components/applications, such as their reliability, have been tested.
- **Usability testing**: In all test cycles the testers have evaluated the degree to which the system can be used by specified users with regard to effectiveness, efficiency and satisfaction, in a specified context of use.
- **Performance testing:** The performance of each individual component/application has been tested in the conducted test cycles. The exposed technologies have been evaluated both in a qualitative manner (e.g. is the performance in processing of videos sufficient enough for a journalist?) and a quantitative way on interface level (e.g. measurement of response times at the API level).

- **Regression testing:** In the test cycles, the functionality of the components/ applications has been re-tested to ensure that no defects were introduced as a result of the changes made based on the feedback of previous test cycles.

Testing techniques

- **Exploratory testing:** Besides testing with guidelines, every tester also evaluated each technology in an exploratory mode.
- **Fault-based techniques:** Fault-based techniques have been mainly used in the integration (API) tests. The tests at the API level enable a stable interaction between the different components of the system and thus, also a stable basis for all user interfaces.
- **Scenario testing:** Scenario testing in the test cycles has been based on a typical journalistic workflow for video verification and has provided information on how the exposed and tested technologies help with these tasks.
- **Walkthrough testing:** Test cycle 7 and 9 have included a walkthrough testing via a dedicated session that involved both the testers and the developers of the technologies for the InVID Verification Application.

These testing methods and techniques have been chosen because they are suitable for providing adequately detailed and well-justified feedback on the usability and effectiveness of the evaluated applications and components. The different testing methods have often been used in combination, for example, functional, non-functional and performance testing have been performed in one test sequence.

Table 5 gives an overview of the testing methods that have been used for assessing each component and/or integrated system, in the test cycles of the reporting period. The tests have been performed manually, and in the case of the API tests also automatically with automation tools (JMeter). The automatic API tests ensure the functionality of the tested APIs and reduce time and effort for re-testing the APIs after a change.

Table 5: Use of the different testing methods for the different applications/components (where TC refers to Testing Cycle)

	Targets of testing		**Objectives of testing**						**Testing techniques**			
Application	**Integration testing**	**System testing**	**Alpha testing**	**Beta testing**	**Regression testing**	**Functional testing**	**Non-functional testing**	**Performance testing**	**Exploratory testing**	**Fault-based techniques**	**Scenario testing**	**Walkthrough testing**
Video Fragm. & Annotation	TC7 TC8	TC7 TC8	TC7 TC8	TC7 TC8	TC7 TC8	TC7 TC8	TC7 TC8	TC7 TC8	TC7 TC8	TC7 TC8	TC8 TC9	

Service		TC9		TC9		TC9	TC9	TC9	TC9			
Web App. for Video Fragm. & Reverse Image Search	TC8	TC7 TC8 TC9	TC8	TC7 TC8 TC9	TC8	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC8	TC8	
Near Duplicate Detection Service	TC8	TC7 TC8 TC9	TC8	TC7 TC8 TC9	TC8	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC8	TC8	
Logo Detection Service		TC7 TC8 TC9		TC7 TC8 TC9		TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9			
Forensic Analysis Service	TC7	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7		
Context Aggregation & Analysis Service	TC7 TC8	TC7 TC8 TC9	TC7 TC8	TC7 TC8 TC9	TC7 TC8	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8	TC7 TC8	
Rights Management Service	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9		TC7 TC8 TC9	
Tool for Social Media Retrieval & Topic Detect.		TC7	TC7	TC7 TC9	TC7	TC7 TC9	TC7 TC9	TC7 TC9	TC7 TC9		TC7 TC9	
InVID Multimodal Analytics Dashboard	TC7	TC7	TC7	TC7 TC9	TC7	TC7 TC9	TC7 TC9	TC7 TC9	TC7 TC9		TC7 TC9	
InVID Verification Plugin		TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9		TC7 TC8	
InVID Verification Application	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8	TC7 TC8 TC9	TC7 TC8	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8	TC7 TC8 TC9	TC7 TC9
InVID Mobile Application		TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC8 TC9	TC7 TC9	TC7 TC9	
InVID Core Platform-API		TC7 TC8 TC9										

2.3.3 Testers

The components and applications have been tested by different groups of testers.

- Members of the consortium: Partners with a journalistic background (AFP, DW).
- Members of the companies of the consortium, but external to the project (AFP, DW).
- IT specialists from the companies of the consortium for the testing of the technical interfaces (APIs) (AFP, APA-IT).
- External testers: Users external to both the consortium and the companies of the consortium, e.g. testers from Storyful, BBC, Tiroler Tageszeitung (regional newspaper in Austria), France 24, Newsy.com, Berkeley University, Almasy Alyoum (Daily independent newspaper in Egypt).

Figure 2 shows the gender and age distribution of the testers from the consortium and the companies of the consortium. The external testers were not the same for the different tools of InVID. The demographic data for the external testers is therefore shown together with the testing results in the corresponding sections.

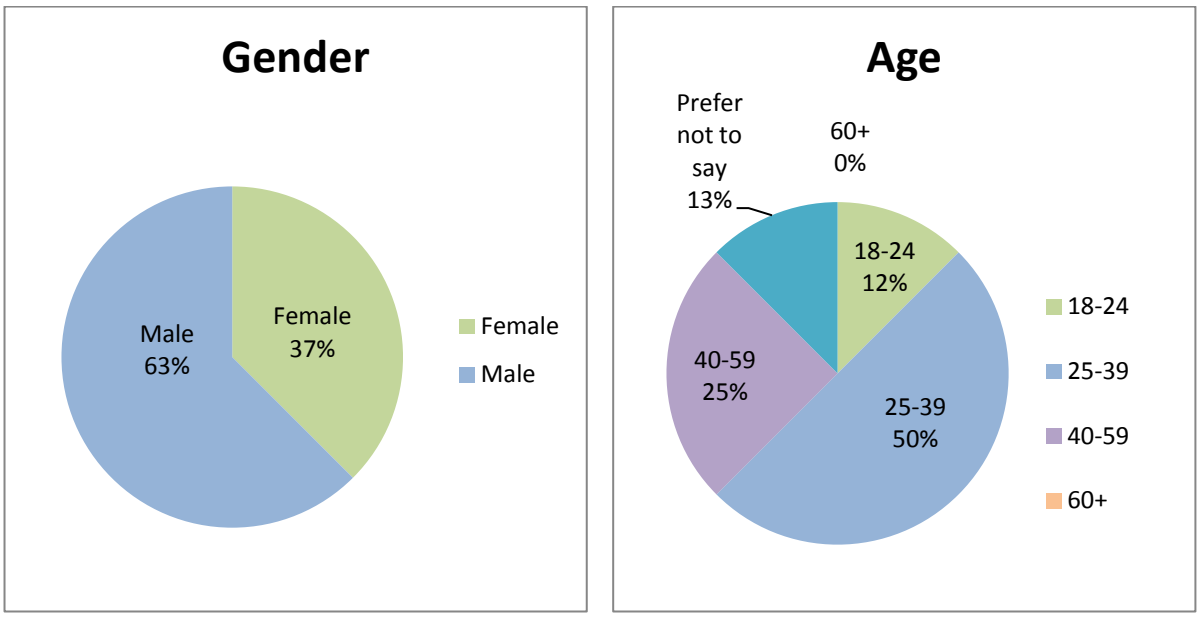

Figure 2: Gender and age of the testers from the consortium and from the companies of the consortium, in total eight testers.

2.3.4 Testing dates

Table 6 shows the time period for the implementation of test cycles 7 to 9. These periods do not include the preparation and the follow-up phases of each test cycle.

Table 6: Testing times of test cycles 7 to 9

Test cycle	Time period
Test cycle 7	2018-05-28 to 2018-06-16
Test cycle 8	2018-09-05 to 2018-09-31
Test cycle 9	2018-11-05 to 2018-11-23

3 Results of test cycles seven to nine

3.1 General remarks

Following a reporting approach similar to the one in D7.2, the results of the three test cycles are structured according to the different components and applications. Each of the following subsections is dedicated to a different component/application. It starts with a short description of the relevant technology, the target group for the component and then provides a description of the performed tests. The number of individual responses/comments about the tested technology is reported at the test cycle level. Finally, the main outcomes of the performance of the evaluated technology are listed.

3.2 Video Fragmentation & Annotation Service

3.2.1 Description of the service

The Video Fragmentation and Annotation Service is a web service (API) that performs a temporal decomposition of a video into three different levels of granularity; scenes (i.e. semantically and temporally coherent segments that correspond to the story-telling parts of the video), shots (i.e. sequences of frames captured uninterruptedly by a single camera) and sub-shots (i.e. visually coherent parts of a video shot; useful when analysing single-shot videos). Consecutively, it identifies the semantics of the video at the most fine-grained level (either shots or sub-shots, depending on the type of analysed video) by detecting a number of high-level visual concepts after analysing one representative key frame per video fragment.

3.2.2 Target groups

The Video Fragmentation & Annotation Service is a web service accessible through a REST API. Therefore it cannot be used directly by users. It is intended to be used by applications such as the Verification Application and Plugin, and the Multimodal Analytics Dashboard.

The target groups for this service (via the aforementioned technologies) are journalists from media organisations, such as publishing houses and broadcasters, members of human rights

organisations and other people dealing with videos from social media, with user-generated videos in general or the verification of videos in general.

3.2.3 Tests

A dedicated test of this service was executed in test cycle 7. The testing of the service's API was performed by IT specialists from AFP and APA-IT. The service had already been in a mature state, therefore no additional direct testing was done in test cycles 8 and 9. Since this service is integrated into the Verification Application, further tests were done indirectly through the testing of the Verification Application in test cycles 7 to 9.

The dedicated tests in test cycle 7 focused on the evaluation of changes and improvements in the API, such as:

- Checking the updated service response after requesting the status of an analysis request under processing.
- Checking the new functionality that allows the user to see the extracted thumbnails (for the shots and sub-shots of the video) in his/her browser.
- Evaluating the time performance of the service in case of Vimeo videos.
- Checking the service's response in case of re-submitting an already submitted and queued video.

Table 7: Number of received feedback comments for the Video Fragmentation & Annotation Service

Test cycle	Feedback comments
Test cycle 7	17
Test cycle 8	--
Test cycle 9	--

3.2.4 Major outcomes of the test cycles

As the Video Fragmentation and Annotation Service is only accessible through its API, all tests have been performed on the communication interface level.

At the beginning of test cycle 7, a tester found a major bug regarding the retrieved thumbnails. The bug was fixed quickly within the test cycle by the developers, which was appreciated by the testers. Other than that, no major bugs were found in the service. Only minor recommendations such as the deactivation of outdated calls and improvements for slightly inconsistent result parameters were made. Further suggestions for improvements were made regarding the indication of the progress of a video-processing, the improvements of the API parameters and the supporting of local videos in the service. These recommendations were taken into consideration and addressed by the developers of this service. The impact of these improvements was observed during the testing of other InVID technologies that integrate this service.

The testers assessed the tool in different aspects in a five point scale with the following range of answer options: “Not good at all”, “Not so good”, “Neutral”, “Good”, “Very good”.

Feedback on the service's documentation:

The documentation was rated as “Good”.

Feedback on reliability:

The reliability of the API was rated as "Neutral" and "Good". The "Neutral" rating was caused by the major bug discovered by the testers at the beginning of the test cycle.

Feedback on response time:

The evaluation of the response time of the service was assessed as “Good".

Feedback on robustness and error handling:

The robustness and error handling capability of the service were marked as "Good". One tester marked the error handling as “Neutral”, but only due to the fact that no errors occurred in the tests of this user and he had therefore found it difficult to assess the error handling.

3.3 Web Application for Video Fragmentation & Reverse Image Search

3.3.1 Description of the service

Shown in Figure 3, the Web Application for Video Fragmentation & Reverse Image Search, accessible through its user interface, allows the user to extract a set of representative key frames from a video, and to search for occurrences of these key frames on the Web, through the reverse image search functionality of the Google search engine.

This service is accessible as a standalone application, but also integrated into the InVID Verification Plugin.

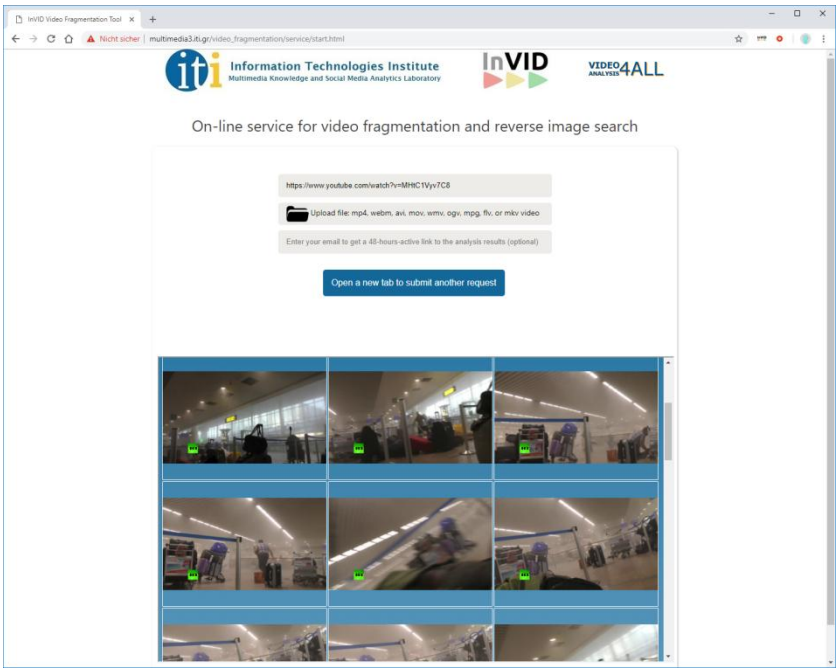

Figure 3: Web Application for Video Fragmentation & Reverse Image Search

3.3.2 Target groups

The target groups for this service are journalists from media organisations such as publishing houses and broadcasters, members of human rights organisations and other people dealing with videos from social media, with user-generated videos in general or the verification of videos in general.

3.3.3 Tests

In test cycle 8, video sub-shot fragmentation and keyframe extraction were comparatively evaluated against two alternative baseline approaches for keyframe extraction by journalists from the companies of the consortium (AFP, DW). The service itself had reached already a mature state; therefore no additional direct testing was done in test cycles 7 and 9. The component is also integrated into the InVID Verification Plugin and has consequently been assessed through the testing of this technology (see Section 3.10).

3.3.4 Results for the evaluation of the video sub-shot fragmentation and keyframe extraction

The InVID approach for video sub-shot fragmentation and keyframe extraction was comparatively evaluated against two alternative baseline approaches for keyframe extraction; one extracting a single keyframe per second, and another one that extracts the reference frames (a.k.a. I-frames) of the mp4 video stream[1]. This benchmarking was conducted with the help of two journalists - one coming from Agence France-Presse (AFP)

[1] Both of these approaches were implemented using the FFmpeg framework that is available at: https://www.ffmpeg.org/

and one coming from Deutsche Welle (DW) - with media verification background, and its focus was bilateral. In particular, it aimed to assess:

- the effectiveness of each tested approach in defining a set of keyframes that represents the visual content of the video without missing any important pieces of information, with the least amount of frames;
- the usefulness / appropriateness of each generated keyframe collection for supporting the task of finding near duplicates of the analysed video on the Web.

Given that the evaluated InVID method is integrated into the web application for reverse video search, this testing allowed to assess how concise and complete the produced collection of keyframes is, and to which extent the generated collection (and thus this web application) facilitates the quick identification of prior occurrences of a given video on the Web.

According to the evaluation protocol each tester was asked to select 10 user-generated videos (at least); these videos could be either online available videos from the Web or local videos from the testers' machines. Experimentation with non-user-generated videos (i.e. edited professional videos) was also permitted. Subsequently, each selected video should be submitted for analysis to:

- the InVID web application for reverse video search that uses the InVID approach for video fragmentation and keyframe selection;
- a variation of this tool that creates a keyframe collection by applying the first alternative and extracts one keyframe per second;
- another variation of this tool that defines a keyframe collection by applying the second alternative and extracts the reference frames (a.k.a. I-frames) of the mp4 video stream.

After analysing each selected video with the above listed technologies, the testers had to answer the following questions.

- Q1: How many keyframes were extracted by each tested approach?
- Q2: Which collection is the most concise and complete one (i.e. represents the visual content of the video without missing any important pieces of information, with the least amount of frames)?
- Q3: If you try reverse image search: which collection helps you the most to quickly identify near duplicates of the video on the Web?
- Q4: if the used videos are publicly accessible, please copy-paste the links at the end of this document.

Their feedback was provided by filling in the following tables. Table 8 contains the evaluation results of the AFP journalist and Table 9 includes the evaluation results of the DW journalist. In the utilised ranking system for answering questions Q2 and Q3, 1 stands for the worse performance and 5 stands for the best.

The videos submitted for analysis and reverse search by each tester, are listed in Table 10.

Table 8: The votes of the AFP journalist regarding the tested approaches for video keyframe extraction and keyframe-based reverse video search.

		Q1: Number of extracted keyframes	Q2: concise and complete					Q3: helps the most in reverse search				
	Method		1	2	3	4	5	1	2	3	4	5
			(worse)				(best)	(worse)				(best)
Video #1	InVID	17				X					X	
	Alt. 1	43		X						X		
	Alt. 2	12			X				X			
Video #2	InVID	6		X					X			
	Alt. 1	17				X				X		
	Alt. 2	4	X					X				
Video #3	InVID	101				X					X	
	Alt. 1	371		X					X			
	Alt. 2	127			X					X		
Video #4	InVID	4				X					X	
	Alt. 1	19					X			X		
	Alt. 2	5				X					X	
Video #5	InVID	9				X				X		
	Alt. 1	29				X				X		
	Alt. 2	46			X				X			
Video #6	InVID	10				X				X		
	Alt. 1	43			X					X		
	Alt. 2	43			X					X		
Video #7	InVID	65				X				X		
	Alt. 1	210			X					X		
	Alt. 2	92			X					X		
Video #8	InVID	13				X				X		
	Alt. 1	46			X					X		
	Alt. 2	45			X					X		
Video #9	InVID	85		X						X		
	Alt. 1	303				X					X	
	Alt. 2	72		X					X			
Video #10	InVID	31			X						X	
	Alt. 1	74			X					X		

	Alt. 2	32			X					X		

Table 9: The votes of the DW journalist regarding the tested approaches for video keyframe extraction and keyframe-based reverse video search.

		Q1: Number of extracted keyframes	Q2: concise and complete					Q3: helps the most in reverse search				
	Method		1	2	3	4	5	1	2	3	4	5
			(worse)				(best)	(worse)				(best)
Video #1	InVID	41				X					X	
	Alt. 1	150		X					X			
	Alt. 2	31				X						X
Video #2	InVID	20					X				X	
	Alt. 1	81		X					X			
	Alt. 2	18					X				X	
Video #3	InVID	6					X				X	
	Alt. 1	20		X					X			
	Alt. 2	7					X				X	
Video #4	InVID	6				X					X	
	Alt. 1	25			X					X		
	Alt. 2	9				X					X	
Video #5	InVID	42				X					X	
	Alt. 1	153	X					X				
	Alt. 2	68		X				X				
Video #6	InVID	14			X					X		
	Alt. 1	52		X				X				
	Alt. 2	21			X					X		
Video #7	InVID	6					X				X	
	Alt. 1	26		X					X			
	Alt. 2	8				X					X	
Video #8	InVID	36					X					X
	Alt. 1	139	X						X			
	Alt. 2	53				X						X
Video #9	InVID	10		X				X				
	Alt. 1	54			X						X	
	Alt. 2	15				X					X	

Video #10	InVID	20				X		X				
	Alt. 1	64		X				X				
	Alt. 2	17				X		X				

Table 10: The submitted videos by the AFP and DW journalists for evaluating the InVID and the two alternative methods for video keyframe extraction and keyframe-based reverse video search.

#	AFP journalist's videos	DW journalist's videos
1	https://www.youtube.com/watch?v=GhxqIITtTtU	https://www.youtube.com/watch?v=okvoLbHlaVA
2	https://www.youtube.com/watch?v=oKQiTUjHlQ4	https://www.youtube.com/watch?v=ziOvZSUwU_c
3	https://www.facebook.com/Oker.Turgut/videos/1708996762482817/	https://twitter.com/AZeckenbiss/status/1033790392037199873
4	https://twitter.com/kengarex/status/1003749477583413249	https://twitter.com/JorgeaHurtado/status/1018125444158279682
5	https://www.youtube.com/watch?v=sza-j0nubNw	https://www.facebook.com/nafisa.alharazi/videos/10156699747657790/
6	https://twitter.com/tprincedelamour/status/843421609159544836	https://www.facebook.com/goodshitgoOdsHitthatssomegoodshitrightthere/videos/347521802658077/
7	https://www.youtube.com/watch?v=r5aBqCniQyw	https://www.youtube.com/watch?v=szKPipLRFsM
8	https://video.twimg.com/ext_tw_video/876820481919397889/pu/vid/360x640/VWTPEvrV8vVJFf4d.mp4	https://www.youtube.com/watch?v=BU9YAHigNx8
9	Thailand_cave_rescue_video (uploaded)	https://www.youtube.com/watch?v=DeUVsmWji8g
10	https://www.youtube.com/watch?v=UTeqpMQKZaY	https://www.youtube.com/watch?v=-sWZuykJy9Q

The feedback received from the AFP journalist showed that the InVID approach exhibits competitive performance compared to the other tested approaches. Concerning the generation of a concise and complete keyframe-based summary of the video content, the InVID algorithm was the highest-voted one in seven cases, and the second best performing one in the remaining three cases. The representation effectiveness of the first alternative, which extracts a single keyframe per second, was positively appreciated by the AFP journalist in four cases where the algorithm was voted as the best (or among the best) performing one(s). The second alternative that selects the I-frames of the video was indicated as the least effective one and marked as the second best in four cases only.

The good ranking of the first alternative approach reveals the AFP journalist's preference in having keyframe collections that sufficiently cover all the details of the video, even if this

entails a compromise regarding the comprehensiveness of the created collection and the existence of information redundancy. As further detailed in his evaluation report, the explanation behind this choice is governed by his news verification background and relies in the fact that some video frames might contain an element that helps to confirm the location, identify a person, a scene, an event or something useful for the verification or debunking of a news video. As a consequence, the appearance of these frames in the keyframe collection, even if near-duplicates of them - that are less informative though - are already included in this collection, is positively assessed. Finally, the keyframe collections generated by the second alternative, even being comparatively-sized with the ones created by the InVID algorithm (see Table 8), proved to be less useful than the other evaluated techniques due to more missing frames that are needed for effectively conveying the reported story in the video.

An example that illustrates the findings reported above is depicted in Figure 4 which contains the generated keyframe collections by the three evaluated algorithms for the analysed video #4. The top left corresponds to the InVID method, the bottom left corresponds to the second alternative and the right-sided one corresponds to the first alternative. The video reports a story about the first woman in Saudi Arabia that receives her driving license, and it is recorded within an office by a (mainly) standing cameraman. The InVID-extracted keyframe collection contains three keyframes that show the provision of the license by the officer to the woman. The keyframe collection created by the second alternative conveys (visually) the same information but exhibits more redundancy, as keyframes #3 and #4 are near-duplicates of keyframes #2 and #5 respectively. Finally, the collection generated by the first alternative covers the story in much more details, but at the cost of much higher duplication of visual information. Nevertheless, the last keyframe of this collection shows a photographer that is also in the room and takes a photo of this event. His appearance in the video does not change or affect the main subject of the video, but it can provide a hint that could help a journalist to verify or debunk this video (e.g. by observing a badge on his uniform that relates to a specific country or army). Hence, the journalist's voting (as shown in Table 8) rewards the existence of this keyframe in the collection, considering it as more important than the information redundancy that this collection presents.

Concluding, the keyframe selection strategy of the first alternative combined with the competitive performance of the InVID approach in most examined cases, indicates the InVID method as the most effective one in generating keyframe-based video summaries that are well-balanced according to the determined criteria for the descriptiveness (completeness) and representativeness (conciseness) of the keyframe collection.

Figure 4: The keyframe collections generated for an AFP-selected video by the three tested approaches; the top left corresponds to the InVID method, the bottom left corresponds to the second alternative and the right-sided corresponds to the first alternative

Concerning the use of the generated keyframe collections by the three evaluated methods to facilitate the identification of near-duplicates of the processed videos on the Web, the InVID method was generally determined as the most useful one. The keyframe collections extracted by this method, were considered as helping the most in reverse video search in three cases, and as equally effective with the collections produced by other approaches in five cases. The first alternative proved to be the second most appreciated method, and this finding is aligned with the journalist's interest, explained previously, to get and use any visual detail of the video that could assist its verification. Finally, the second alternative was ranked as the less effective one since the extracted keyframe collections were denoted as less useful for video reverse search in several of the tested scenarios. These outcomes are consistent to the findings regarding the comprehensiveness of the generated keyframe collections, and show that the InVID developed algorithm and the first alternative can (almost) equally support the users' needs when performing a fragment-level reverse search of a video on the Web.

The collected feedback from the DW journalist clearly indicates the InVID approach as the best performing one in producing a concise and complete keyframe-based summary of the video. In most cases (specifically in nine out of ten) the InVID method got the highest score compared to the other tested approaches. The keyframe collections generated by this algorithm were voted as best (four times) or well (four times) performing ones. A similar, but in some cases less effective, performance was shown by the second alternative which extracts the I-frames of the video. This technique was evaluated as approximately equally performing one with the InVID approach in six cases, while in one case it was voted as the most effective technique. This finding is reasonable if we take into account that this method: a) selects the frames of the video that are the most complete and descriptive ones in terms of visual information (in order to be used as the reference basis for the compression of the

subsequent p- and b-frames of the video) and b) usually results in a small set of keyframes that is comparable in size with the collection of keyframes extracted by the InVID method, as reported in Table 9 and shown in the example of Figure 5 below (left column). The least competitive one was the first alternative that extracts one keyframe per second. This method results to a very fine-grained keyframe-based representation of the video, but the high amount of redundant information (due to the occurrence of near-duplicate frames) reduces the usefulness of this collection when someone tries to quickly discover the video content.

The above described findings are illustrated in Figure 5 which shows the extracted keyframe collections by the three tested approaches for the submitted video #7. Once again, the top left corresponds to the InVID method, the bottom left corresponds to the second alternative and the right-sided one corresponds to the first alternative. As can be seen, the latest one offers a very detailed and complete representation of the video content; however, several keyframes exhibit high visual resemblance, thus resulting in significant information redundancy which, in case of long videos, makes the discovery of the video content a time-consuming process. On the contrary, the left-sided keyframe collections provide a concise but also complete summary of the video content, as they contain all the key parts of the presented story. The collection generated by the second alternative (at the bottom left of Figure 5) includes a couple of near-duplicate frames, and thus was voted as slightly worse than the collection produced by the InVID approach.

As an overall comment, the keyframe selection strategy of the second alternative, in combination with the competitive performance that the InVID method exhibits in most cases, indicates that the developed algorithm for video sub-shot fragmentation and keyframe selection is highly effective in extracting a set of keyframes that represent the visual content of the video without missing any important pieces of information, with the least amount of frames.

Figure 5: The keyframe collections generated for a DW-selected video by the three tested approaches; the top left corresponds to the InVID methods, the bottom left corresponds to the second alternative and the right-sided corresponds to the first alternative

In terms of keyframe-based reverse search for quickly finding near-duplicates of the submitted videos on the Web, the InVID approach and the second alternative were voted as equally performing in seven cases. Moreover, the InVID method was the best performing in one case and the second best performing in two cases. The second alternative was voted as the best in two cases, while the first alternative was marked as the least effective in all tested cases. The latter can be explained by the fact that, even providing a very fine-grained representation of the video content, this collection increases the amount of the time and effort needed to discover the keyframe collection and select the most appropriate keyframes for performing the keyframe-based reverse video search on the Web. These findings are aligned to the ones extracted regarding the conciseness and completeness of the generated keyframe collections, and indicate the InVID method and the second alternative as the best choices for performing a fragment-level reverse search of a video on the Web.

Summing up the collected feedback regarding the quality of the developed video fragmentation approach for creating a concise and complete summary of the video content, and the effectiveness of this visual summary for supporting the task of reverse video search on the Web, we reach the conclusion that this technology is the best trade-off between two desirable but, to some extent, incompatible features. It results in keyframe collections that adequately maintain the visual details of the video content which can be highly-valued for

video verification tasks (being aligned to the AFP journalist's demands), while it secures a concise representation of the presented story, thus allowing the quick discovery of the video content and the sufficiently fine-grained, fragment-level search for finding near-duplicates of the video on the Web (meeting the DW journalist's requirements).

3.4 Near Duplicate Detection Service

3.4.1 Description of the service

The Near Duplicate Detection Service, which is accessible via its API, aims to identify near-duplicate content in a video collection that has been indexed by the InVID platform. It gives a similarity value to indexed videos and therefore allows the detection of near-duplicates of a query image or video. This is a strong indication of prior use of this media item in the past, and thus, evidence that it is not original. The functionality of the Near Duplicate Detection Service is available to the InVID users through its integration into the InVID Verification Application.

3.4.2 Target groups

The Near Duplicate Detection Service is only accessible through an API and cannot be used directly by users. It is intended to be used by applications such as the Verification Application. The target groups for this service (via applications such as the Verification Application) are journalists from media organisations such as publishing houses and broadcasters, members of human rights organisations and other people dealing with videos from social media, with user-generated videos in general or the verification of videos in general.

3.4.3 Tests

A dedicated test of this service was executed in test cycle 8. In test cycles 7 to 9, the evaluations of the accuracy and appropriateness of the analysis results were done by journalists through the testing of the InVID Verification Application. The results are reported with the Verification Application in Section 3.11. The dedicated tests in test cycle 8 focused on robustness, reliability, error handling and response time.

Table 11: Number of received feedback comments for the Near Duplicate Detection Service

Test cycle	Feedback comments
Test cycle 7	--
Test cycle 8	6
Test cycle 9	--

3.4.4 Major outcomes of the test cycles

As the Near Duplicate Detection Service is only accessible through its API, all tests have been performed on the communication interface level. The API was already in a mature state and only minor recommendations were made by the testers like improvements in checking the API parameters for invalid values.

The testers assessed the tool in different aspects in a five point scale with the following range of answer options: “Not good at all”, “Not so good”, “Neutral”, “Good”, “Very good”.

Feedback on the service's documentation:

The documentation was rated as “Good” to “Very good”.

Feedback on reliability:

The reliability of the API was rated as “Good” to “Very good”.

Feedback on response time:

The evaluation of the response time of the service was assessed as “Not so good" for the search function whereas the performance for the other functions of the interface was assessed as “Good”.

In response to the assessment of the search function the developer of the service have implemented a caching mechanism that temporarily stores the results of the search calls. The response of the service is now instant, when the same query video is given and its results exist in the service's cache. Additionally, further measures have been taken in order to decrease the response time of the service, including the heavy parallelization of the search process and the re-indexing of the videos (with a more comprehensive visual codebook) for improved retrieval efficiency.

Feedback on robustness and error handling:

The robustness and error handling of the API was rated as “Good” to “Very good”.

3.5 Logo Detection Service

3.5.1 Description

The Logo Detection Service (accessible both via its API and its UI – see Figure 6) evaluates the existence of a given logo within a video or image. The service integrates an extendable pool of logos of particular interest for journalists, and after processing a video or image, it

provides information about the occurrence of these logos in the video frames or within the image, aiming to assist investigators in identifying the origin of the media item under investigation.

The user interface of the Logo Detection service has been developed only for demonstration purposes and for internal testing within the InVID project. The functionality of the Logo Detection Service is accessible to the InVID users, via the InVID Verification Application.

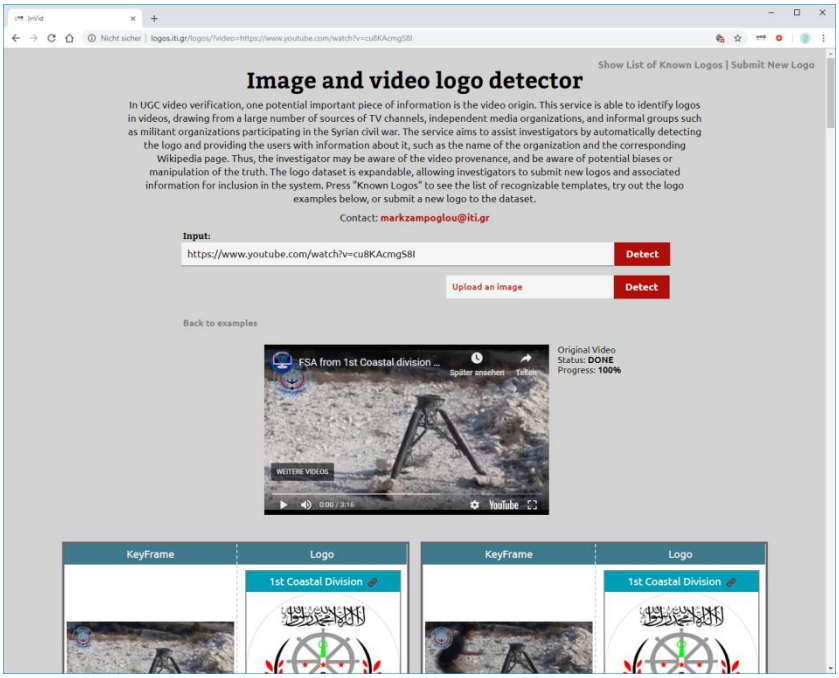

Figure 6: Logo Detection Service

3.5.2 Target groups

The target groups for this service (via the Verification Application) are journalists from media organisations, such as publishing houses and broadcasters, members of human rights organisations and other people dealing with videos from social media, with user-generated videos in general or the verification of videos in general.

3.5.3 Tests

The Logo Detection Service had reached a mature state prior to test cycle 7. Therefore, no explicit tests were done in the test cycles 7 to 9. Since the Logo Detection Service is integrated into the Verification Application, the service has been tested by journalists through the testing of the InVID Verification Application.

3.6 Forensic Analysis Service

3.6.1 Description of the service

The Forensic Analysis Service provides a set of developed forensic filters that assist users in detecting manipulation of videos and is available on API level. The functionality of the

Forensic Analysis Service is available to the InVID users by its integration into the InVID Verification Application.

3.6.2 Target groups

The Forensic Analysis Service is only accessible through an API and cannot be used directly by users. It is intended to be used by applications such as the Verification Application. The target groups for this service (via applications such as the Verification Application) are journalists from media organisations such as publishing houses and broadcasters, members of human rights organisations and other people dealing with videos from social media, with user-generated videos in general or the verification of videos in general.

3.6.3 Tests

A dedicated test of this service was executed in test cycle 7. The testing of the service's API was done by IT specialists from AFP and APA-IT. The tests focused on functionality and the error handling capability of the API. The service had reached already a mature state, therefore no additional direct testing was done in test cycles 8 and 9. However, further tests of the Forensic Analysis Service have been carried out through the testing of the Verification Application which integrates this analysis component.

Table 12: Number of received feedback comments for the Forensic Analysis Service

Test cycle	Feedback comments
Test cycle 7	13
Test cycle 8	--
Test cycle 9	--

3.6.4 Major outcomes of the test cycles

The service is reliable and no major bugs have been found in test cycle 7. Feedback mainly concerned the improvement of error messages and some minor bugs.

The testers assessed the tool in different aspects in a five point scale with the following range of answer options: “Not good at all”, “Not so good”, “Neutral”, “Good”, “Very good”.

Feedback on documentation:

> The documentation of the service was rated from “Neutral” to "Good" due to some missing details. Suggestions for the improvement of the documentation were made.

Feedback on reliability:

> The reliability of the service was considered as "Good".

Feedback on response time:

> The response time of the service was evaluated as "Good" and "Very good".

Feedback on robustness and error handling:

The robustness and error handling capability of the service were marked as "Neutral". The testers suggested further improvements of the error messages of the service.

3.7 Context Aggregation & Analysis Service

3.7.1 Description of the service

The Context Aggregation & Analysis Service (accessible via an API and a user interface – see Figure 7) aims at facilitating the verification of content derived exclusively from the YouTube, Facebook and Twitter platforms. The Context Aggregation & Analysis Service is used by the Verification Application and by the Verification Plugin.

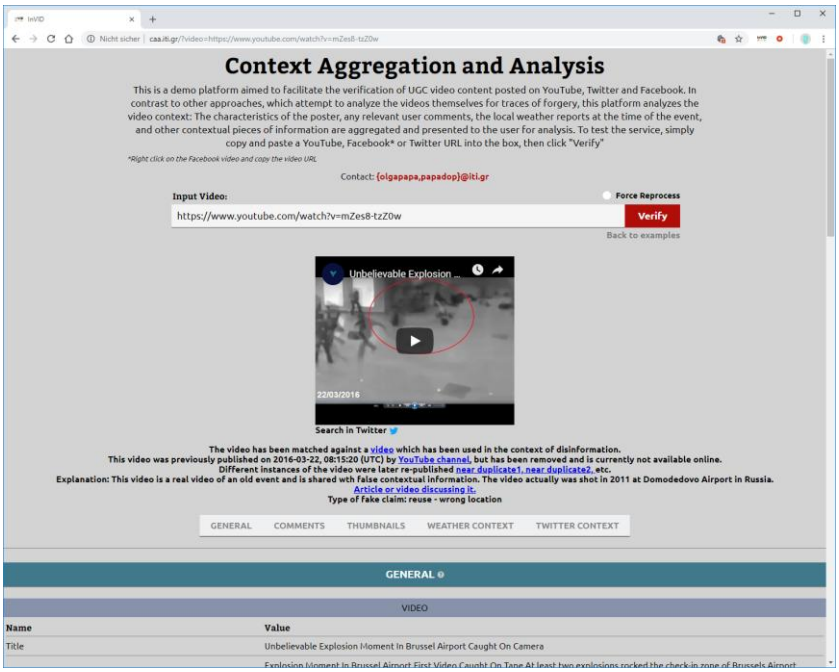

Figure 7: Context Aggregation & Analysis Service

3.7.2 Target groups

The target groups for this service are journalists from media organisations such as publishing houses and broadcasters, members of human rights organisations and other people dealing with videos from social media.

3.7.3 Tests

As mentioned above, this service provides a user interface and an API. Both were tested in the test cycles 7 and 8. Dedicated tests of the user interface were done by testers of the consortium with a journalistic background and also by members of the companies of the consortium. The testing of the service's API was done by IT specialists from AFP and APA-IT.

In test cycle 7, the tests of the user interface and API included:

- Changes in weather context

- Changes in Twitter timeline for YouTube and Facebook videos
- Changes in Twitter timeline for Twitter videos
- Changes in Embedded Twitter video

In test cycle 8, the conducted tests focused on:

- Search comments by keyword (user interface)
- Presentation improvement (user interface)
- Twitter timeline improvements (user interface)
- Response time (API)
- Reliability (API)
- Fault tolerance (API)

Table 13: Number of received feedback comments for the Context Aggregation & Analysis Service

Test cycle	Feedback comments from testing the user interface	Feedback comments from testing the API
Test cycle 7	40	10
Test cycle 8	23	8
Test cycle 9	--	--

3.7.4 Major outcomes of the test cycles

Major results from the user interface tests:

The testers assessed that this tool helps a journalist get more insights into a video based on the contextual data. The added value of this tool for a journalist was rated as "High" and “Very high” in a five point scale with the following range of answer options: “Very low”, “Low”, “Neutral”, “High”, “Very high”.

The new feature “Search comments by keyword” was appreciated by the testers. However, they also marked that if comments are in a foreign language, the search function is of limited help. They also noted that if there are many comments for one video, the tool cannot load them all due to restrictions on the YouTube API. In this case, the search function in comments is only partially helpful.

The weather tool was assessed as very helpful. Several comments in test cycle 7 concerning the user interface lead to an improvement of the interface. This was well received by the testers in test cycle 8. The Twitter timeline improvements for test cycle 8 have also been based on the feedback from test cycle 7. As a consequence, the responses of the testers for the Twitter timeline were clearly more positive in test cycle 8. Furthermore, the testers reported occasional videos where the processing of the video had been taking too long.

The feedback from the test cycles has been taken into consideration in the following development cycle and has been retested by the users in the next test cycle.

Major results from the API tests:

No bugs were found in the test cycles 7 and 8. Minor suggestions for improvements of the parameters and status messages of the API were made by the testers. The testers assessed the API as very good and complete and also stated that the API could easily be integrated into a GUI. The testers assessed the tool in different aspects in a five point scale with the following range of answer options: “Not good at all”, “Not so good”, “Neutral”, “Good”, “Very good”.

Feedback on documentation:

The documentation of the service was rated "Good" and "Very good".

Feedback on reliability:

The reliability of the API was considered as "Very good".

Feedback on response time:

The response time of the service was marked with answers from "Not so good" to "Good". There seemed to be a slight loss in performance with the new API version. The loss of performance was due to the inclusion of an additional service in order to correctly retrieve the video id of a submitted URL. This was necessary to ensure the processing of all submitted videos despite the many variations of submitted video URLs. In response to the feedback of the testers the API calls have been optimized and the response times are at now almost as quick as in the last version of the API.

Feedback on robustness and error handling:

The robustness and error handling capability of the service were labelled as “Good” and "Very good".

3.8 Rights Management Service

3.8.1 Description of the service

The InVID Rights Management Service (accessible both via an API and a user interface – see Figure 8) deals with the copyright aspects related to the reuse of UGC. It helps to discover the owner of an interesting piece of UGC, contact the owner, set a copyright negotiation framework and establish the required rights agreements to reuse the asset. The Rights Management Service is integrated into the Verification Application and the Verification Plugin.

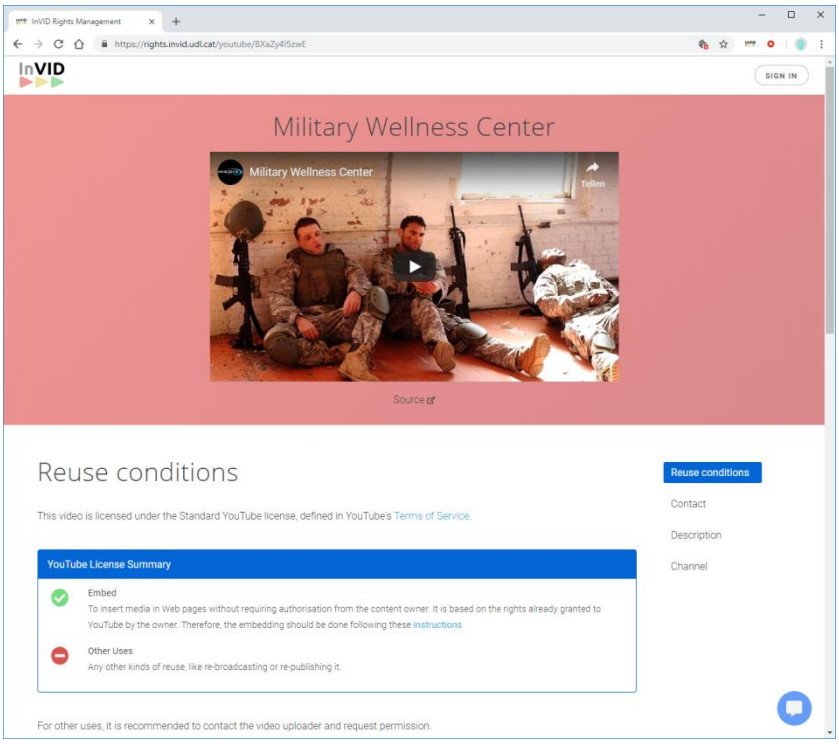

Figure 8: Rights Management Service

3.8.2 Target groups

The target groups for this service are journalists from media organisations such as publishing houses and broadcasters that want to reuse videos from social media and need to know the copyright terms of the video and a contract with the owner of a video to use it.

3.8.3 Tests

The user interface of the Rights Management Service was tested in all three test cycles (7 to 9). The testing of the user interface was done by testers of the consortium with a journalistic background and by testers from the companies of the consortium. As the API for the Rights Management Service had already been in excellent status, it has not been tested again in test cycles 7 to 9.

In test cycle 7, the user interface tests focused on:

- Negotiating a reuse request
- Accepting/rejecting a reuse request
- Creating a reuse request from a template

In test cycle 8, the user interface tests focused on:

- Using a personal account instead of using the same sample account for all testers

In test cycle 9, the user interface tests focused on:

- Requesting a reuse request at the organisational level.

Further tests on this technology have been done by journalists through the testing of the Verification Application and the Verification Plugin which integrate this service.

Table 14: Number of received feedback comments for the Rights Management Service

Test cycle	Feedback comments from testing the user interface
Test cycle 7	26
Test cycle 8	25
Test cycle 9	6

3.8.4 Major outcomes of the test cycles

The testers stated that this tool helps a journalist a lot in clarifying the situation regarding the rights on a newsworthy video. The system was assessed as intuitive and easy to understand. The way of displaying the legal information for a video was appreciated by the testers.

The major feedback from test cycle 7 concerned suggestions to improve the workflow of a reuse request and improvements in the selection possibilities of the request form, e.g. adding a country/region selection, changes in the naming of different choices and a possibility to tell the copyright owner where the requester plans to use the video. The user interface was in a very stable and usable status, only a few bugs were reported, e.g. the saving of individual templates was not possible.

The major feedback from test cycle 8 concerned the registration process for new users and further suggestions for the request form and request process. Regarding the registration process, the users reported problems when the organisation web site had been filled in without “http://” and problems with changing the password. Regarding the request process, a user pointed out the duplicate display of the same information on different pages. The negotiation process worked very well and only a minor bug was found for a scenario where the content owner negotiates with the journalist and the journalist revokes the negotiation. In this case, the reuse request was marked as rejected by the content owner whereas it was the journalist who had revoked the request.

In test cycle 9, no bugs or problems were reported.

3.9 InVID Multimodal Analytics Dashboard and Tool for Social Media Retrieval and Topic Detection

3.9.1 Description of the service

Shown in Figure 9, the InVID Multimodal Analytics Dashboard developed in T5.3 serves as a visual frontend to collect and explore newsworthy user-generated content shared via social media platforms. It exploits the technologies developed in T2.1, T2.2 and T2.3 that allow the user to detect emerging stories across social media channels, identify and retrieve the most relevant content items, analyse the latest trends by topic, geographic region and positive vs.

negative sentiment, and perform a multimodal text- and visual-based search for finding the most adequate media fragments that report a story.

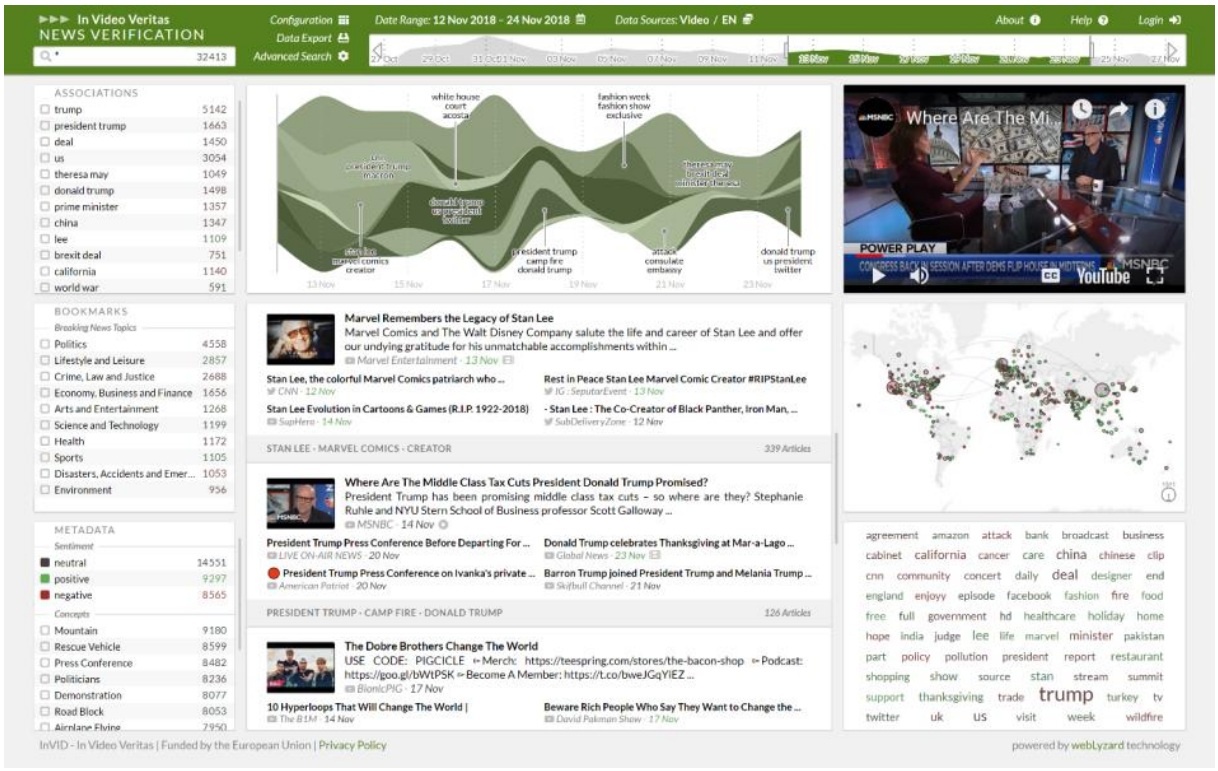

Figure 9: Screenshot of the InVID dashboard as of November 2018

The InVID dashboard is a *Web-based Single Page Application* (SPA) following a multiple coordinated view approach (Hubmann et al., 2009), aiming to provide an experience similar to that of a desktop application. The development process followed a series of iterative deployments, followed by feedback cycles after each deployment. This user-centered iterative process, shown in Figure 10, promotes usability throughout the whole development lifecycle (Matera, M et al, 2006).

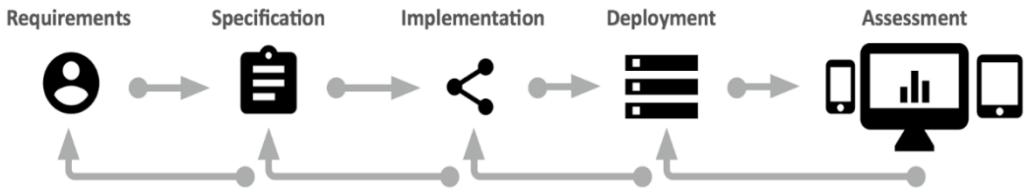

Figure 10: Diagram showing the development process that enables iterative qualitative validation

3.9.2 Target groups

The target groups for this service are persons from media organisations such as publishing houses and broadcasters, like journalists, editors, program planners, data analysts as well as communications and marketing managers who need to analyse emerging trends on social media, detect stories, and monitor topics or search in exploratory mode.

3.9.3 Tests

As part of the final evaluation of the InVID dashboard[2] in test cycle 9, participants were invited to: 1) watch a short video tutorial[3] that introduces the main dashboard features, 2) formulate and run several test queries using the public dashboard prototype, and 3) complete the online survey[4] shown in Figure 11 that had been implemented using Google Forms. The participants were informed that their responses would be treated strictly confidential, and that reported results would not contain any personally identifiable information.

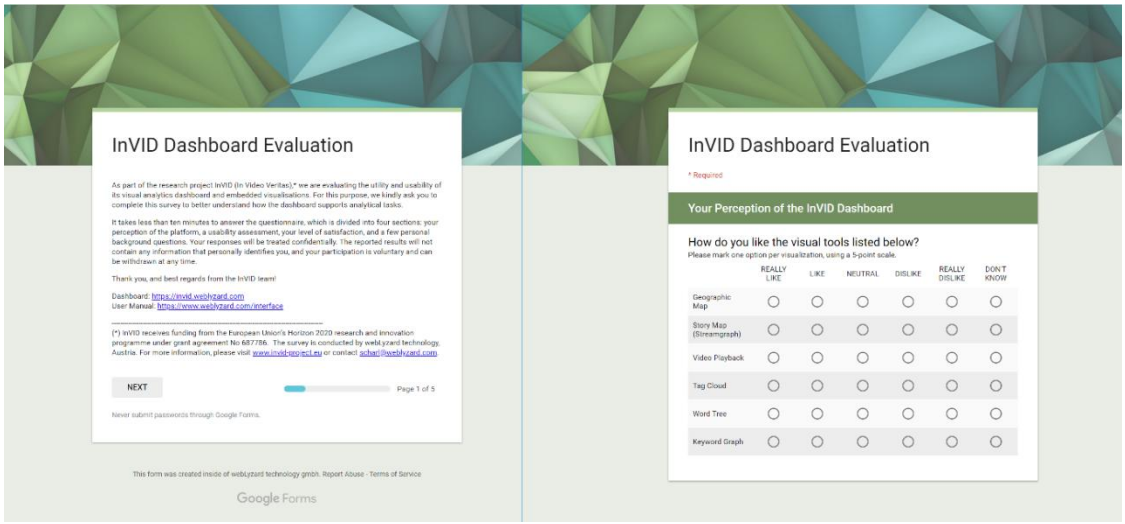

Figure 11: Screenshot of the online questionnaire to conduct the InVID dashboard evaluation

Table 15: Number of received survey responses from external testers

Test cycle	Survey responses
Test cycle 9	18 survey responses

3.9.4 Major outcomes of the test cycles from external users

In total, 18 individuals from Austria, France, Germany and Switzerland completed the survey – see Figure 12 for the specific gender and age distribution. The respondents include employees of the media partners participating in the InVID project, as well as professional contacts from third-party organizations. The respondents had a variety of backgrounds (journalism, media production, law, economics, communications, information systems and computer science) and current positions (managing directors, project and innovation managers, scientists and consultants).

[2] https://invid.weblyzard.com

[3] https://www.youtube.com/watch?v=YhhXqsMotF4

[4] https://bit.ly/invid-survey

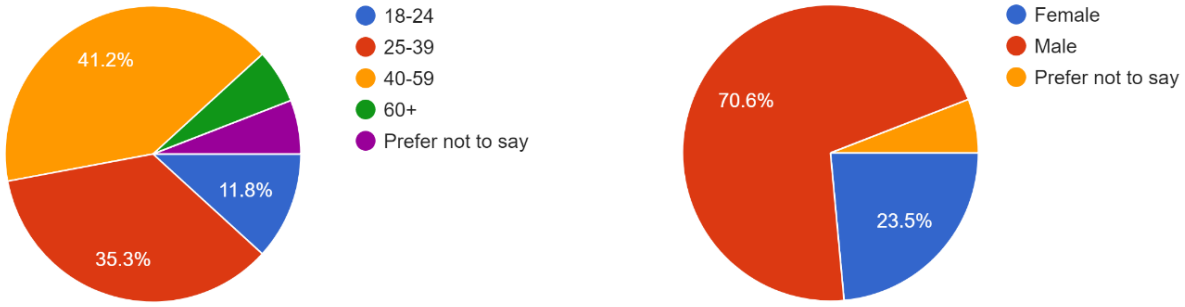

Figure 12: Gender and age distribution of survey respondents

A known limitation of the survey is the fact that a video tutorial of eleven minutes does not suffice to adequately train potential users of the platform. Training workshops that provide a complete overview of the functionality, by contrast, last between three and four hours. But based on earlier discussions within the consortium it was concluded that participants should be able to complete the entire evaluation in about 45 minutes and choose their own timeslot, given the rather limited time resources of the target group especially in the fourth quarter of the year. Therefore, this rather compact format was chosen to shed light on the perceptions of first-time users from media and communication organizations.

Figure 13 summarizes the perceptions of *individual widgets* of the InVID dashboard, including features specifically developed for InVID as well as improved versions of existing components. All widgets received a favourable assessment. The streamgraph-based *Story Map* visualization as well as the video playback functionality, core contributions of WP2 and WP5, received particularly good evaluations. We interpret the “dislikes” of widgets such as the tag cloud as a statement that reflects the comparably low degree of innovation, as compared to more advanced visualizations such as the Story Map or the Word Tree.

In terms of the *integrated dashboard experience* and the synchronisation of its multiple coordinated views, the results were mixed. In the “Satisfaction” section of the survey, users expressed that they like the dashboard’s navigation and particularly its visualizations, and considered the content reliable. In the “Usability” section, users commended the consistency and integration of the dashboard, but also found the interface overly complex and difficult to use. Overall, they were not confident using the dashboard and stated that they needed to learn a lot of things before they could get going with the dashboard. This is reflected in a rather low SUS score of 43.75 and can be partially explained by the survey’s limitation outlined above – i.e., that a 10-minute video tutorial can give a good first impression of the system’s capabilities and features, but not replace a comprehensive training workshop. According to the respondents, the main application scenarios of the integrated dashboard, as shown in Figure 14, are the analysis of emerging trends (88.9%), story detection and visualization (66.7%) and exploratory search (44.4%).

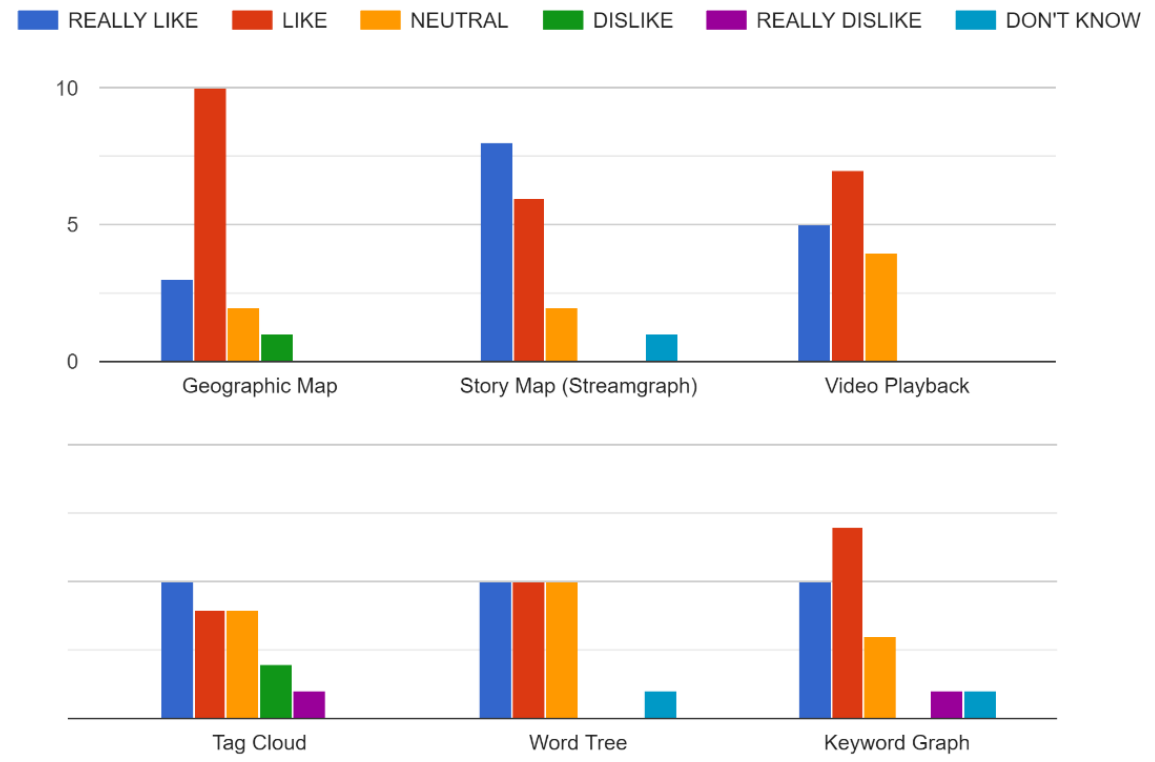

Figure 13: Assessment of the individual widgets offered by the InVID dashboard

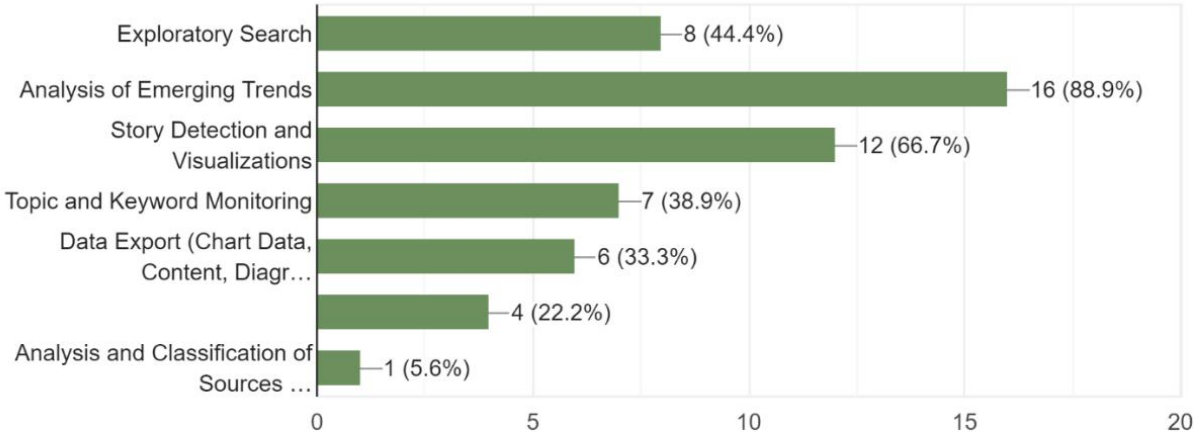

Figure 14: Main usage scenarios of the InVID Dashboard

When asked about ways to improve the dashboard, the survey respondents made the following observations and suggestions (some of them already addressed in the final development cycle of InVID, to be incorporated in the 12-2018 dashboard release):

- **Keyword Extraction.** Test users noticed co-reference variations such as “trump”, “donald trump” and “president trump”, and suggested that they should be merged. This represents a rather fundamental challenge best addressed with named entity recognition and resolution techniques to relate various surface forms encountered in the text to a specific entity (person, organization, location), including disambiguation algorithms to distinguish among multiple entities with the same name. To mitigate this

problem, it is planned to deploy a significantly improved version of the InVID knowledge extraction pipeline in December 2018.

- **Source Management.** It was suggested to segment the social media sources into the accounts of official news sources, and actual user-generated content. While it is not feasible to implement this feature until the end of the InVID project, we will consider this feedback in the exploitation phase and in follow-up projects and have already started with the required classification process.
- **Faceted Search.** Another suggestion related to the use of "facets" to refine queries. The InVID dashboard uses tooltips and an advanced search (not presented in the tutorial) to cover this functionality, as the number and heterogeneity of metadata attributes make it difficult to summarize them in the left sidebar. The availability of vertical space is also limited, given that the left sidebar already contains associations, bookmarks, metadata attributes and the search history. We aim to incorporate this feature in future releases of the advanced search, making it easier to define thresholds and value ranges.
- **Simplification.** Several comments referred to the complexity of multiple coordinated views and recommended to split up the dashboard or omit certain elements. This is already feasible using the embeddable versions of the various widgets, for example, to integrate the streamgraph in a third-party application. There is also a mobile version of the dashboard (mentioned but not shown in the tutorial), which provides a more linear user interface with one visualization per screen. Currently, the mobile version is a separate application, but in future projects we plan to build responsive design capabilities into the main dashboard architecture as well, which will increase flexibility and allow to easily switch between various interface representations and widget combinations.
- **Documentation.** Requests included more general guidance, additional explanations of individual functions, as well as a more detailed communication of the dashboard's use cases. We are currently implementing an extension of the dashboard header that will provide a compact, context-dependent help text to be updated on mouse-over. For advanced users, this help feature will also offer a drop down to show additional metadata instead of the help text; e.g., a list of social media accounts most frequently mentioning a topic.

In addition to the issues listed above, user comments also referred to features that already exist within the InVID dashboard but were omitted due to time constraints (e.g., the analysis of search results over longer time intervals), or features that would be almost impossible to implement without a complete redesign of the platform (e.g., the distinction of dashboard features by colour, which would conflict with the current design principle of mainly using colours to compare topics or metadata attributes). Similarly, a general "back" function would be challenging to implement given browser-based limitations, but the dashboard's optional

"Search History" in the left sidebar (not shown in the tutorial) aims to at least partially address this requirement by repeating previous queries.

3.10 InVID Verification Plugin

3.10.1 Description of the service

The InVID Verification Plugin (see Figure 15) wraps up a number of tools and services and has been developed to help journalists in the verification/fact-checking process. It provides a simple way to use those media verification technologies directly in the browser. Moreover, it enables users to provide their opinion about the usability of the integrated tools and suggest improvements of these technologies via the integrated instant feedback mechanism.

3.10.2 Target groups

The target groups for this service are journalists from media organisations such as publishing houses and broadcasters, members of human rights organisations and other people dealing with videos from social media, with user-generated videos or pictures in general or the verification of videos and pictures in general.

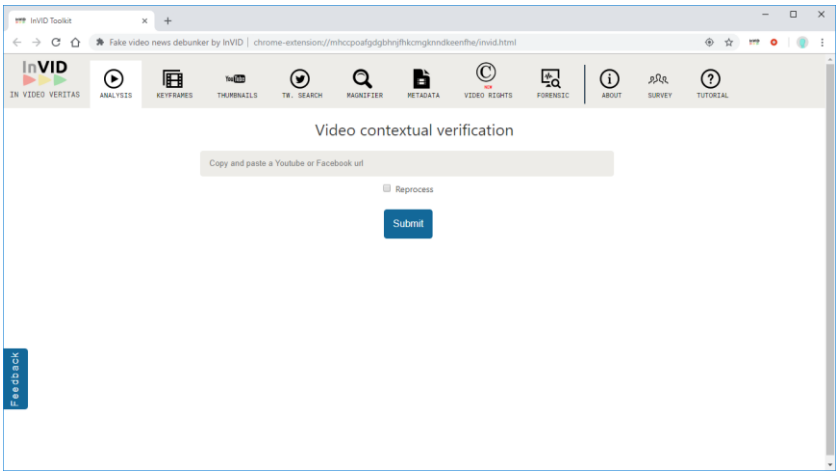

Figure 15: InVID Verification Plugin

3.10.3 Tests

The InVID Verification Plugin was tested in test cycle 7 by members of the companies of the consortium. External testers from outside of the consortium tested the Verification Plugin during the testing period of test cycles 7 to 9. The members of the companies of the consortium tested the improved functionality such as the feedback mechanism and the new video rights section as well as the tool in general. The external testers tested the Verification Plugin as a whole. The results from the external testers were collected with the help of an online survey containing 40 questions regarding the functionality of both the InVID Verification Plugin and each individual component of it. Responses from eight persons were received on the survey during the testing period of test cycles 7 to 9.

Table 16: Number of received feedback comments for the Verification Plugin from members of the companies of the consortium

Test cycle	Feedback comments
Test cycle 7	22 items
Test cycle 8	--
Test cycle 9	--

Table 17: Number of received survey responses from external testers

Test cycle	Survey responses
Test cycle 7 to 9	8 survey responses

The survey responses from the external users were not received during the testing period of the test cycles but during the reporting period for this deliverable. The responses have therefore not been assigned to a specific test cycle.

3.10.4 Major outcomes of the test cycles from members of the companies of the consortium

The members of the companies of the consortium assessed the added value of the InVID Verification Plugin for a journalist as “High” or “Very high” in a five point scale with the following range of answer options: “Very low”, “Low”, “Neutral”, “High”, “Very high”. Additional feedback covered the following subjects:

- Suggestions for small improvements for the different sections of the Verification Plugin.
- Usability suggestions for the new rights section and also for the Verification Plugin in general.
- Reporting of occasional videos which have not been processed correctly by the different tools of the Verification Plugin.

The feedback from a test cycle has been taken into consideration in the following development cycle.

3.10.5 Major outcomes of the test cycles from external users

The data shown in the following diagrams represents the opinion of eight testers regarding several aspects about the usability and functionality of this technology. All testers were external to the consortium and also external to the companies of the consortium. Figure 16 shows the gender and age distribution of the testers and Figure 17 the country distribution.

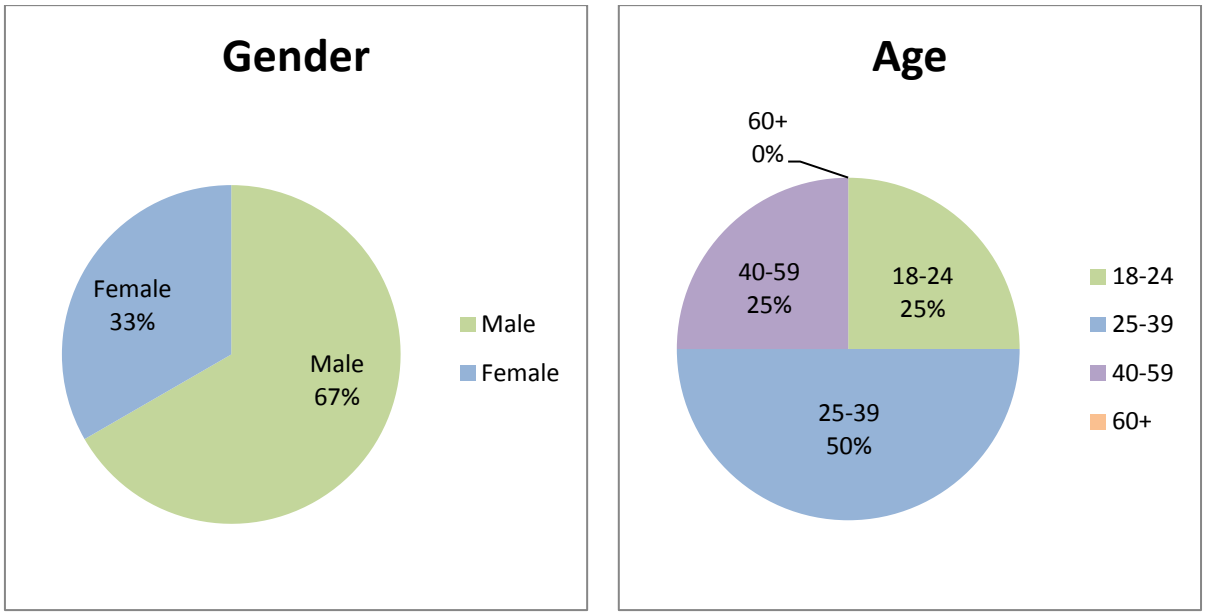

Figure 16: InVID Verification Plugin - external testers: Gender and age of the participants

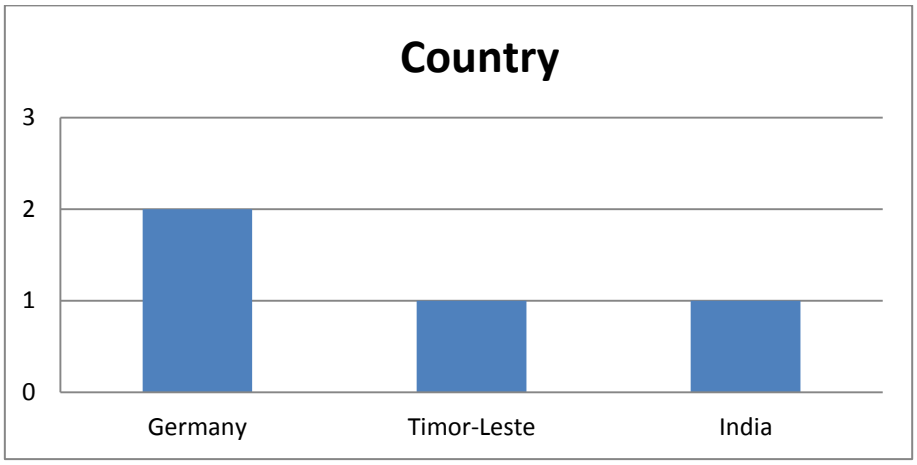

Figure 17: InVID Verification Plugin - external testers: distribution of the testers' countries

The findings of the survey indicate that both the InVID Verification Plugin and the integrated components were very well appreciated by its users. 74% of the users assessed the Verification Plugin as "Very useful" or "Useful", and the same percentage of users found it easy to use. Over 70% of the users were very satisfied or satisfied with almost all features of the Verification Plugin. The Image Magnifier received the lowest rate of satisfaction. Only 37% of the users assessed it with "Very satisfied" or "Satisfied". This is in contrast to the surveys reported in deliverable D7.2 where 76% of the users were very satisfied or satisfied with it. No major changes had been made between the results of the surveys in the Image Magnifier. In the tests reported in this document, some users had problems with the usability (requests for a tutorial). Furthermore, the users requested better filters to enhance the image while keeping its sharpness. Both responses will be considered in the further development of the tool. Nonetheless, 75% of the users would recommend the InVID Plugin to their newsroom and colleagues.

Results of the survey for general questions

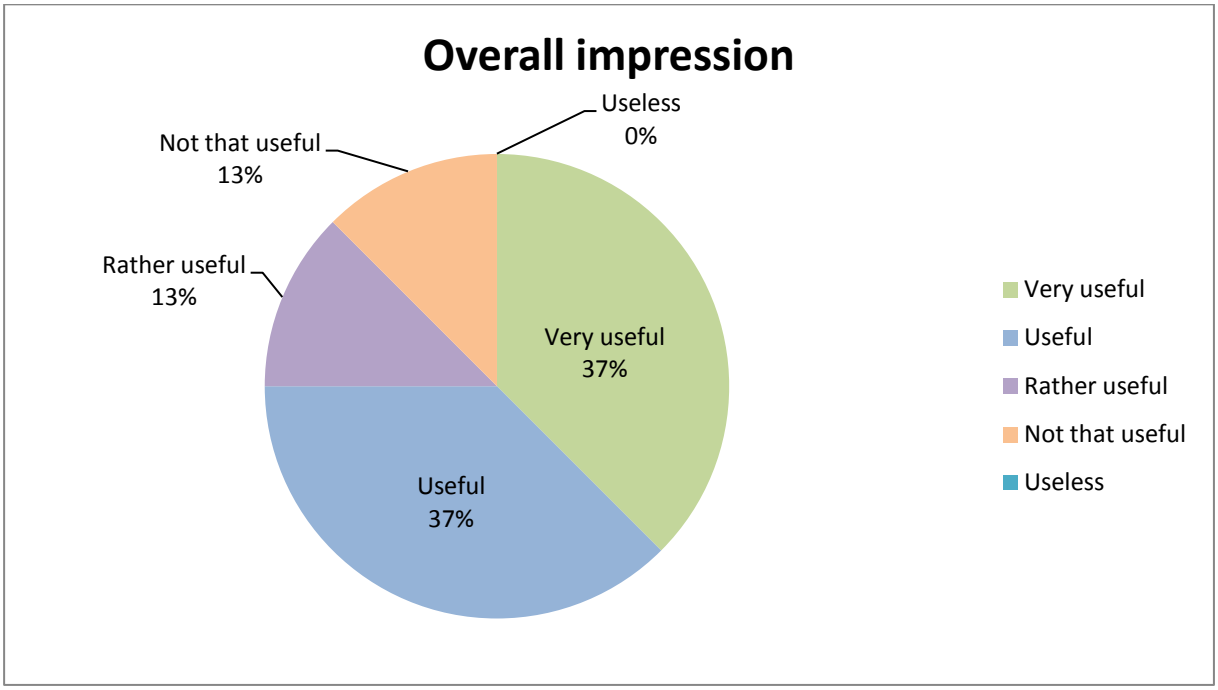

Figure 18: InVID Verification Plugin - external testers: Overall impression

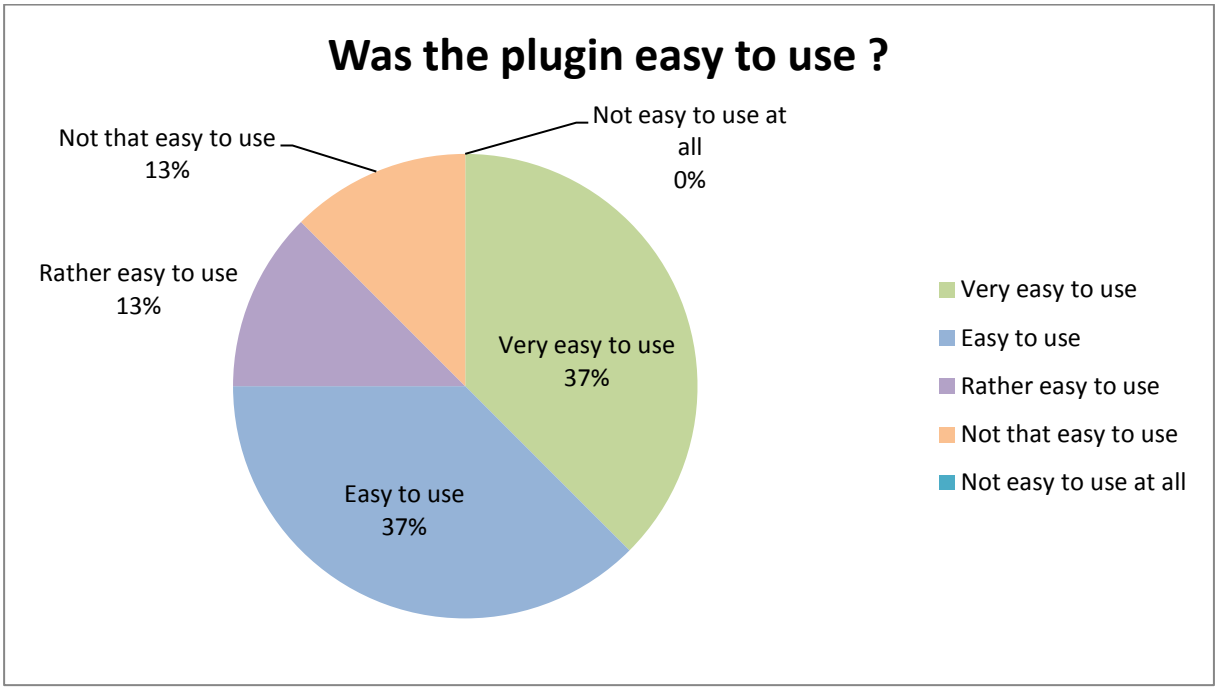

Figure 19: InVID Verification Plugin - external testers: Results for the question “Was the plugin easy to use?”

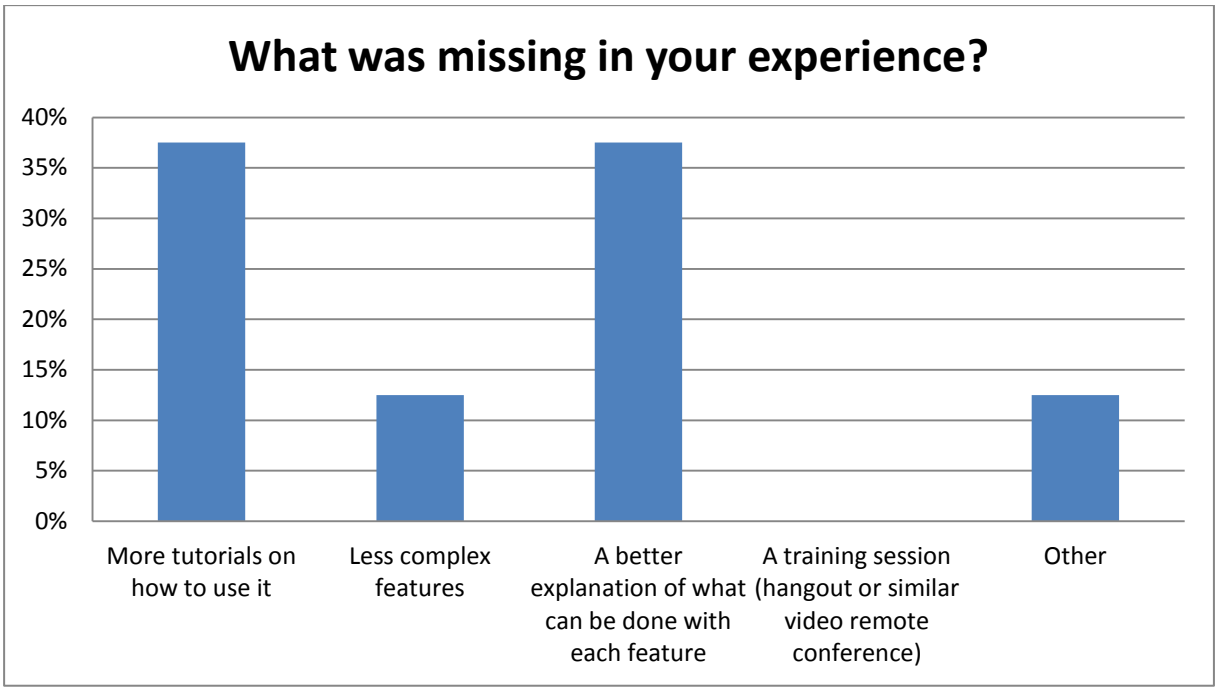

Figure 20: InVID Verification Plugin - external testers: Results for the multiple response question "What was missing in your experience?"

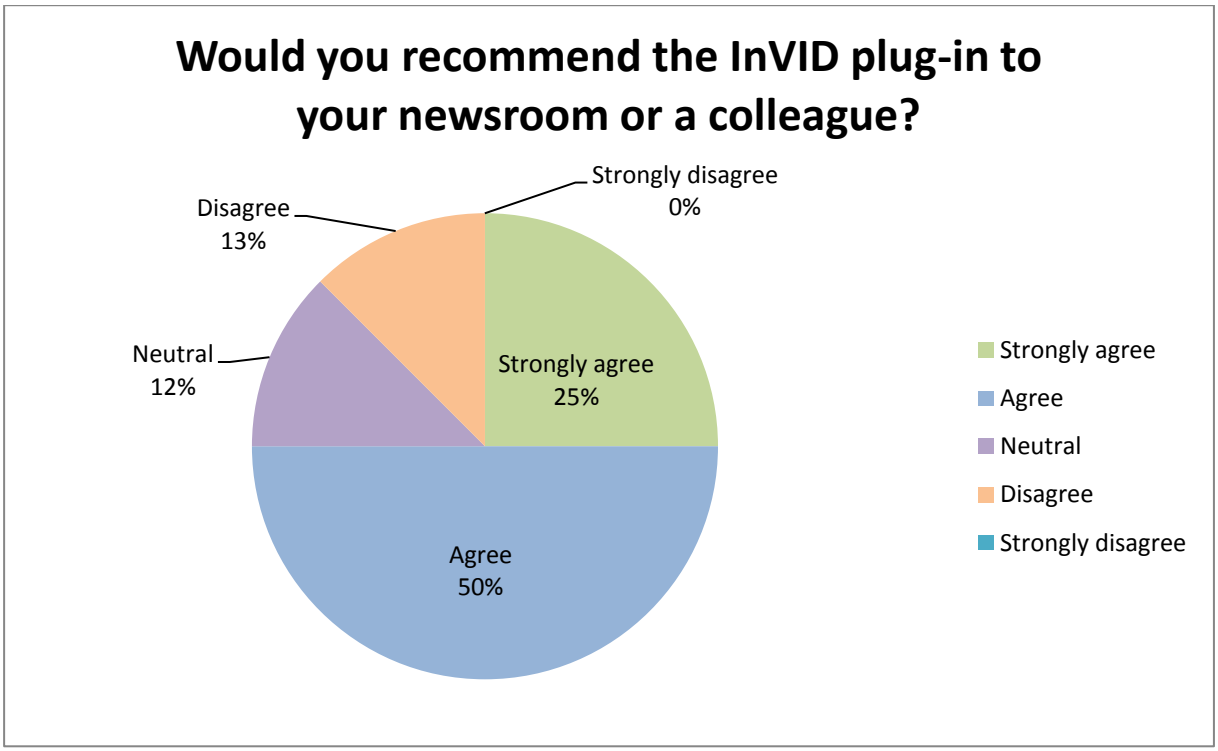

Figure 21: InVID Verification Plugin - external testers: Results for the question "Would you recommend the InVID plug-in to your newsroom or a colleague?"

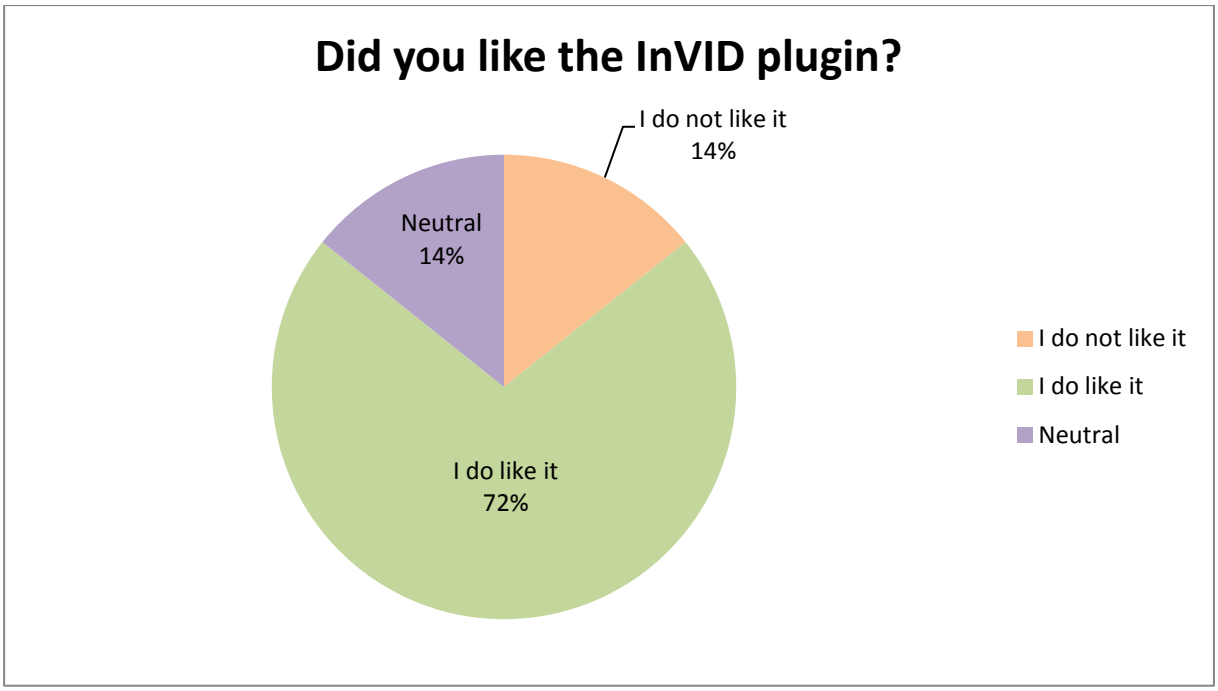

Figure 22: InVID Verification Plugin - external testers: Results for the question “Did you like the InVID plugin?”

Results of the survey for questions regarding the analysis feature (Context Aggregation & Analysis Service)

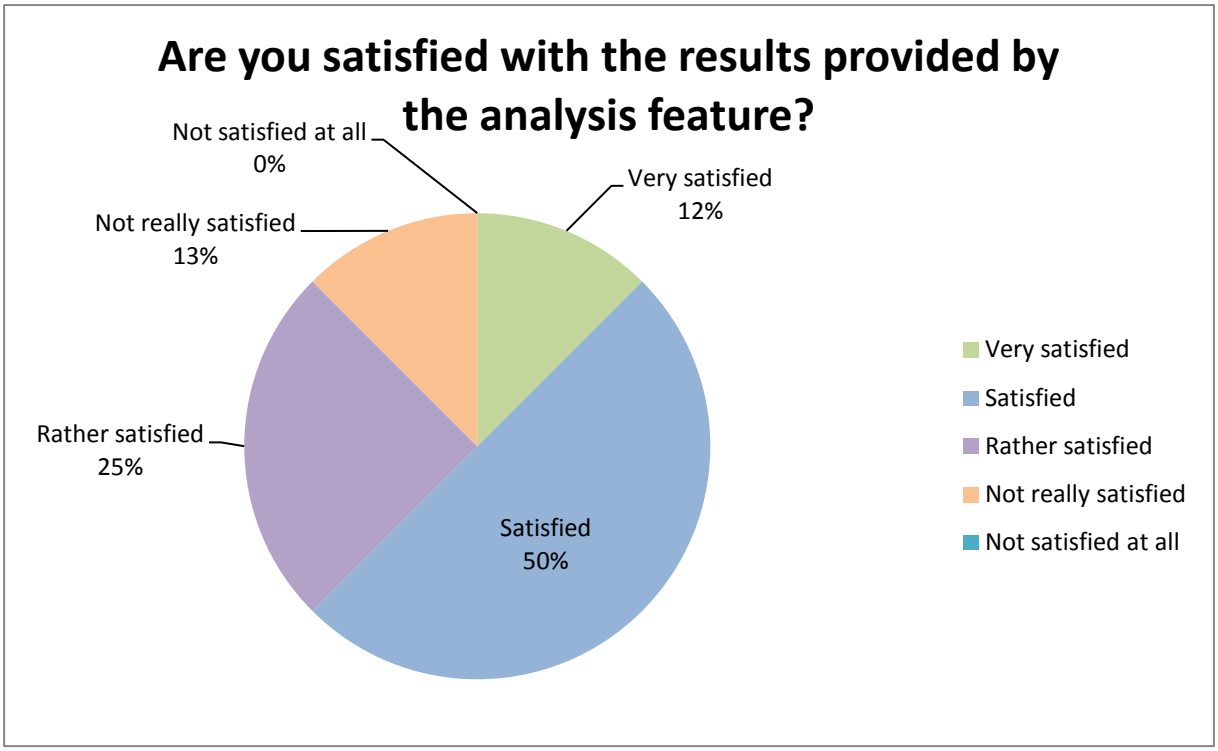

Figure 23: InVID Verification Plugin - external testers: Results for the question “Are you satisfied with the results provided by the analysis feature?”

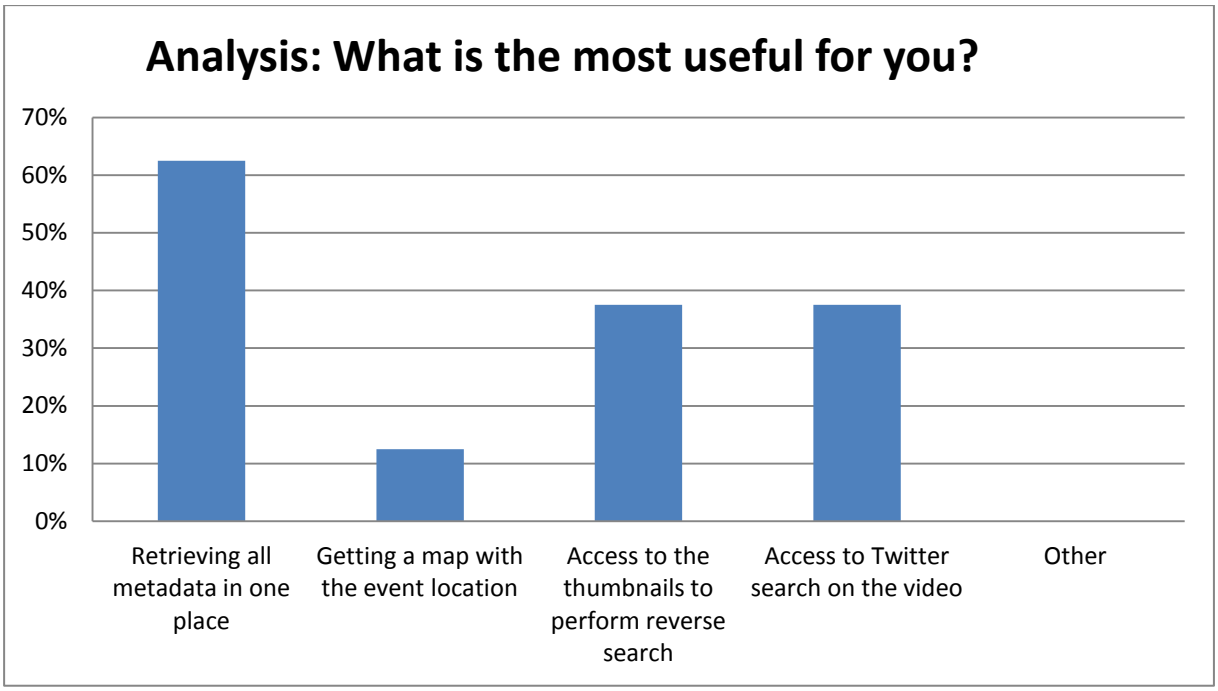

Figure 24: InVID Verification Plugin - external testers - Analysis: Results for the multiple response question “What is most useful for you?”

Results of the survey for questions regarding the keyframe segmentation (Video Fragmentation and Annotation Service)

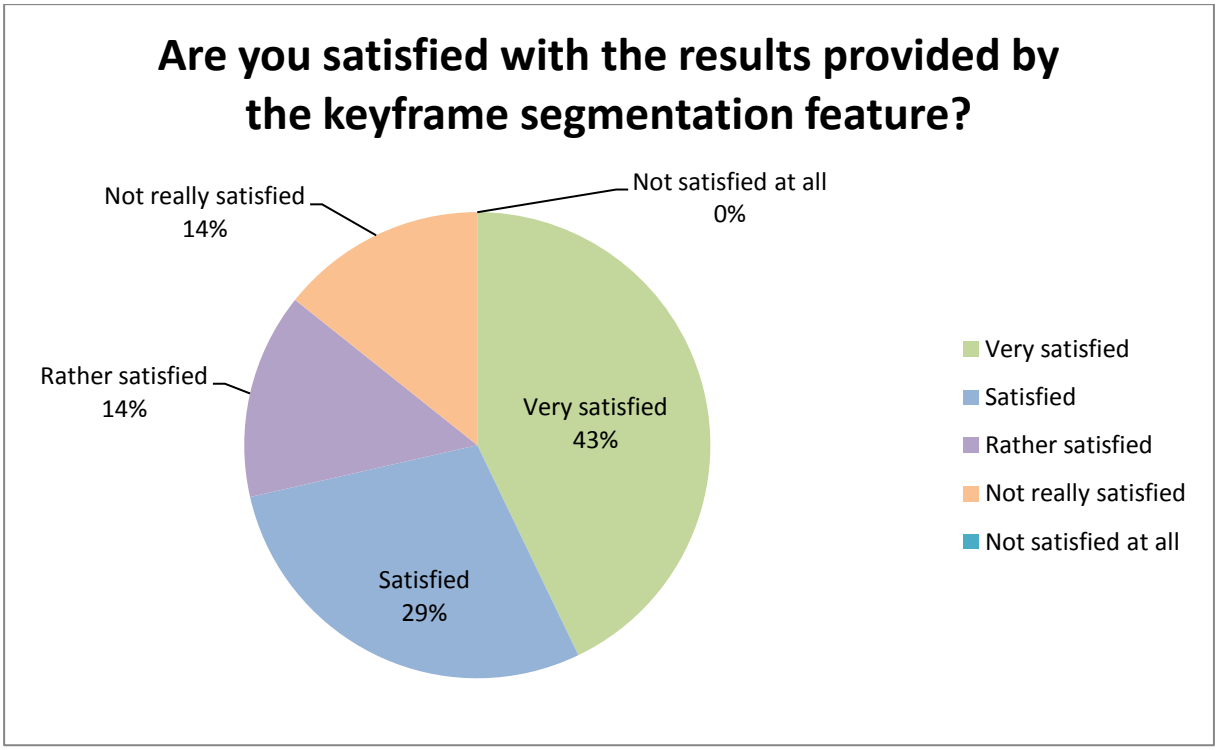

Figure 25: InVID Verification Plugin - external testers - Keyframe segmentation: Results for the question “Are you satisfied with the results provided by the keyframe segmentation feature?”

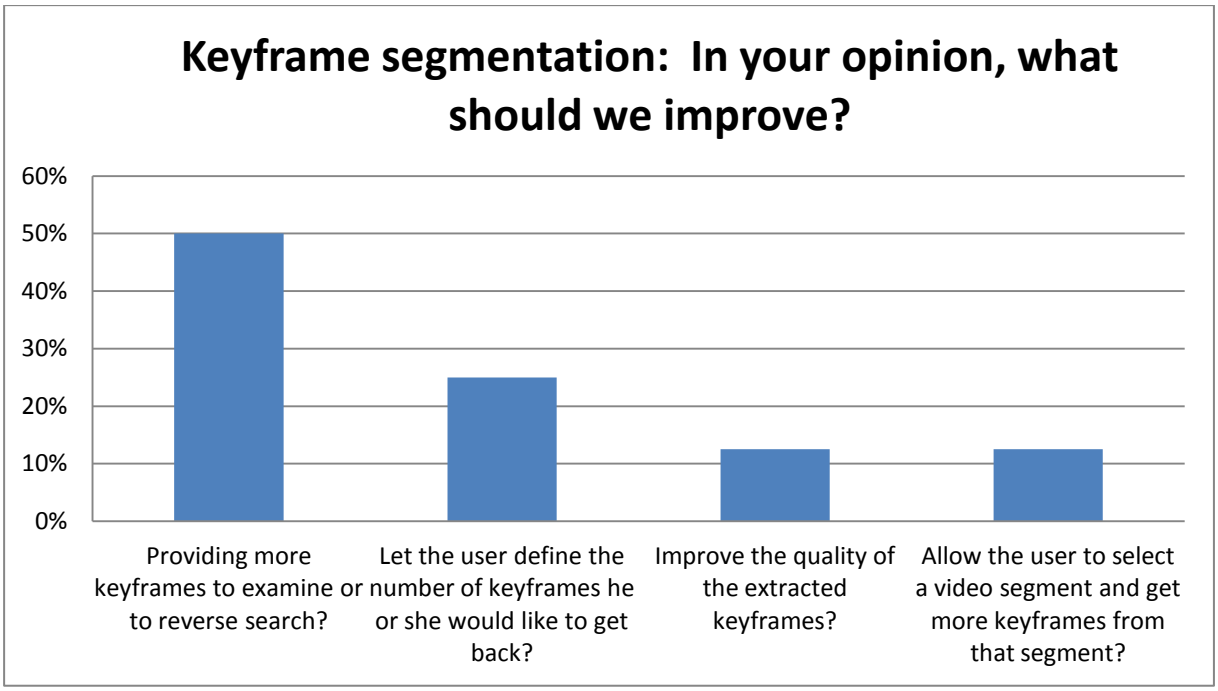

Figure 26: InVID Verification Plugin - external testers - Keyframe segmentation: Results for the multiple response question “In your opinion, what should we improve?”

Results of the survey for questions regarding the YouTube thumbnail feature

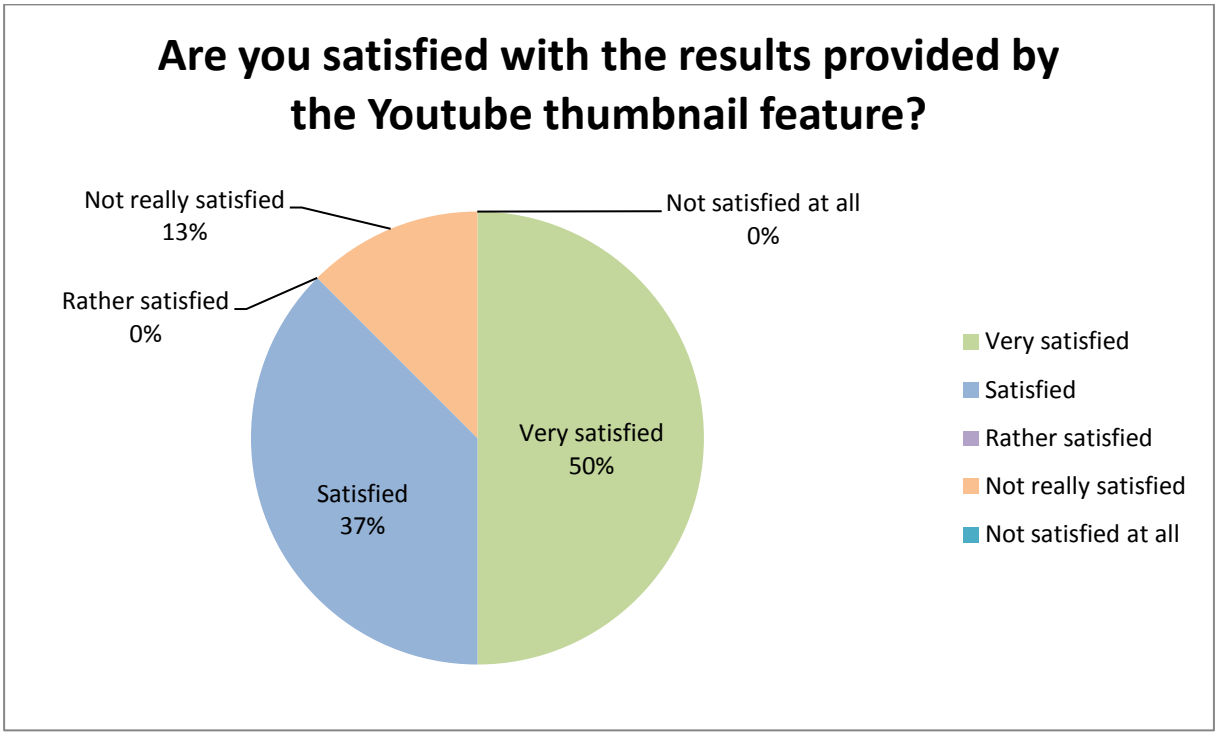

Figure 27: InVID Verification Plugin - external testers - YouTube thumbnail: Results for the question “Are you satisfied with the results provided by the YouTube thumbnail feature?”

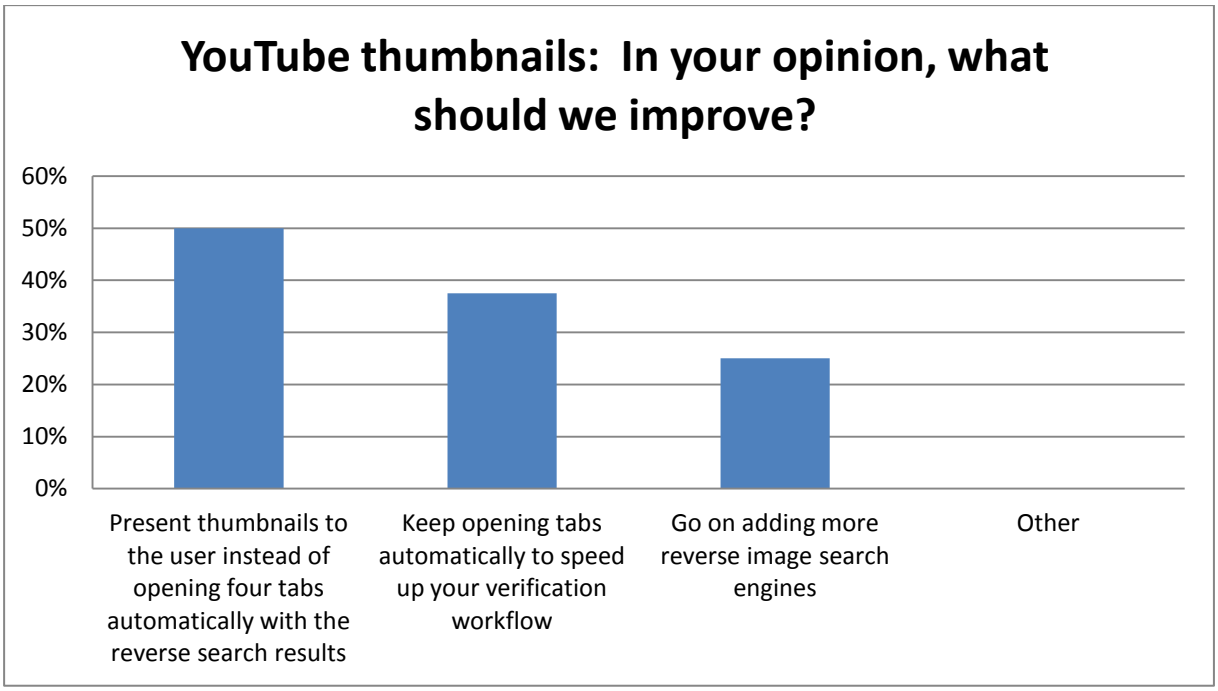

Figure 28: InVID Verification Plugin - external testers - YouTube thumbnails: Results for the multiple response question “In your opinion, what should we improve?”

Results of the survey for questions regarding the Twitter advanced search

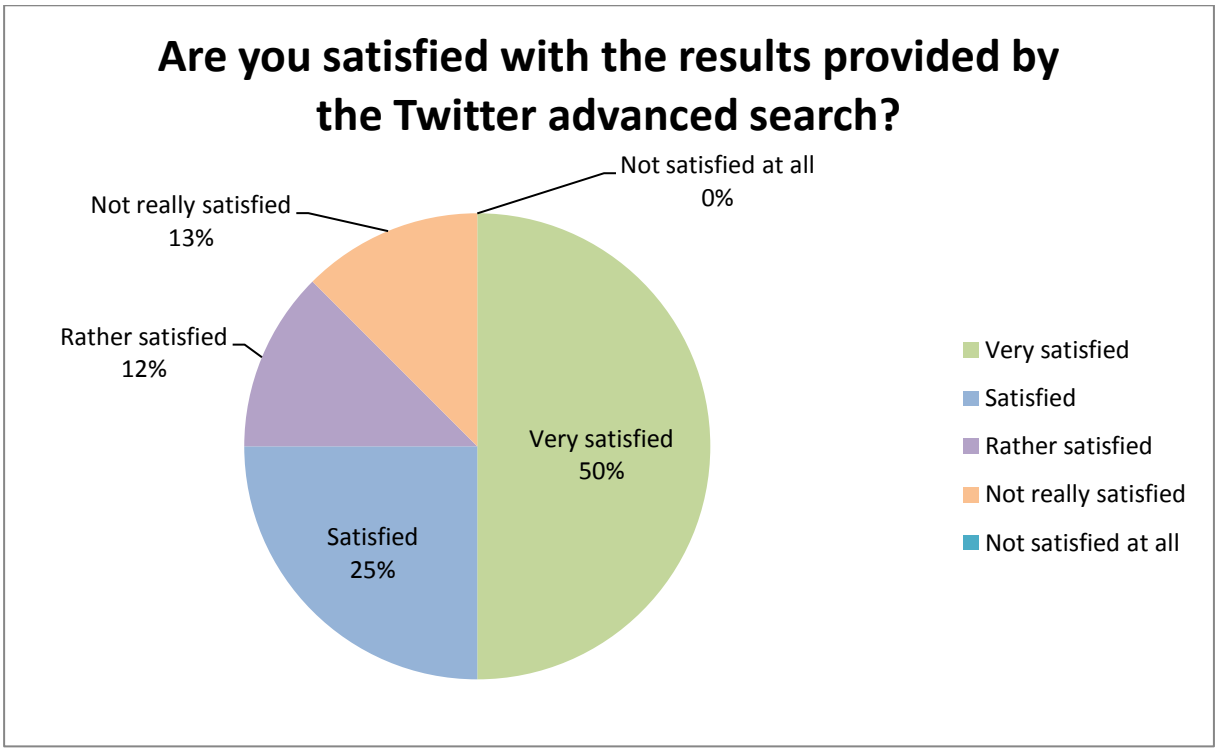

Figure 29: InVID Verification Plugin - external testers - Twitter advanced search: Results for the question “Are you satisfied with the results provided by the Twitter advanced search?”

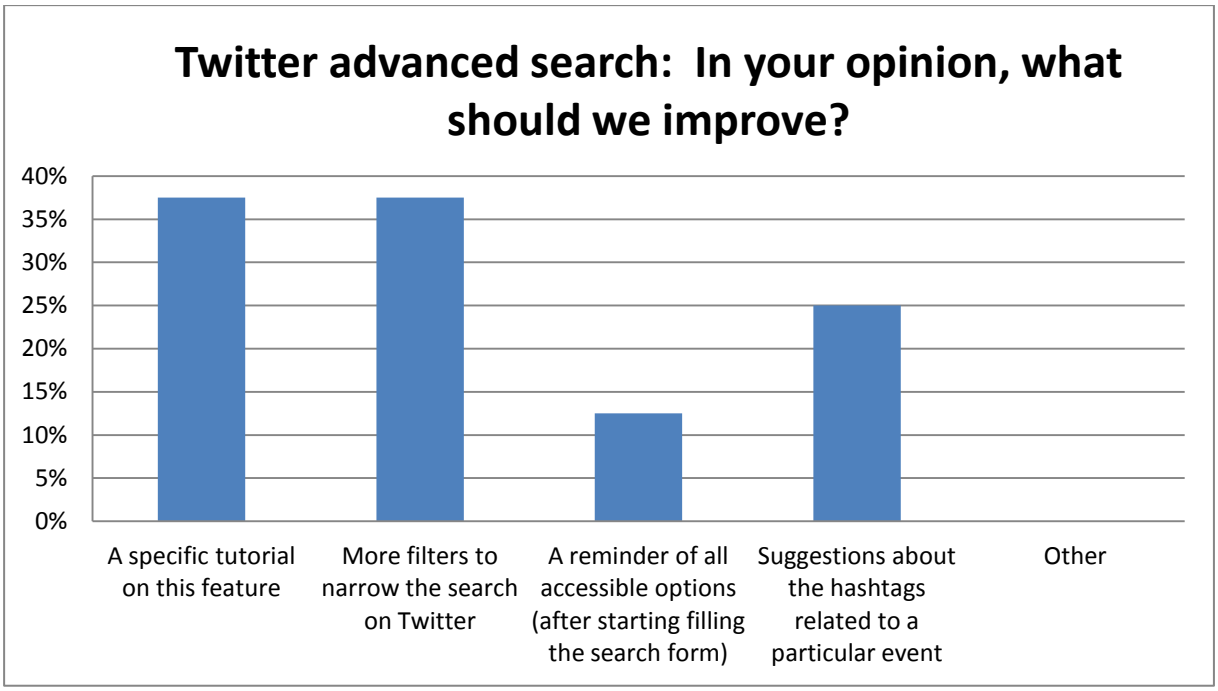

Figure 30: InVID Verification Plugin - external testers - Twitter advanced search: Results for the multiple response question “In your opinion, what should we improve?”

Results of the survey for questions regarding the Image Magnifier

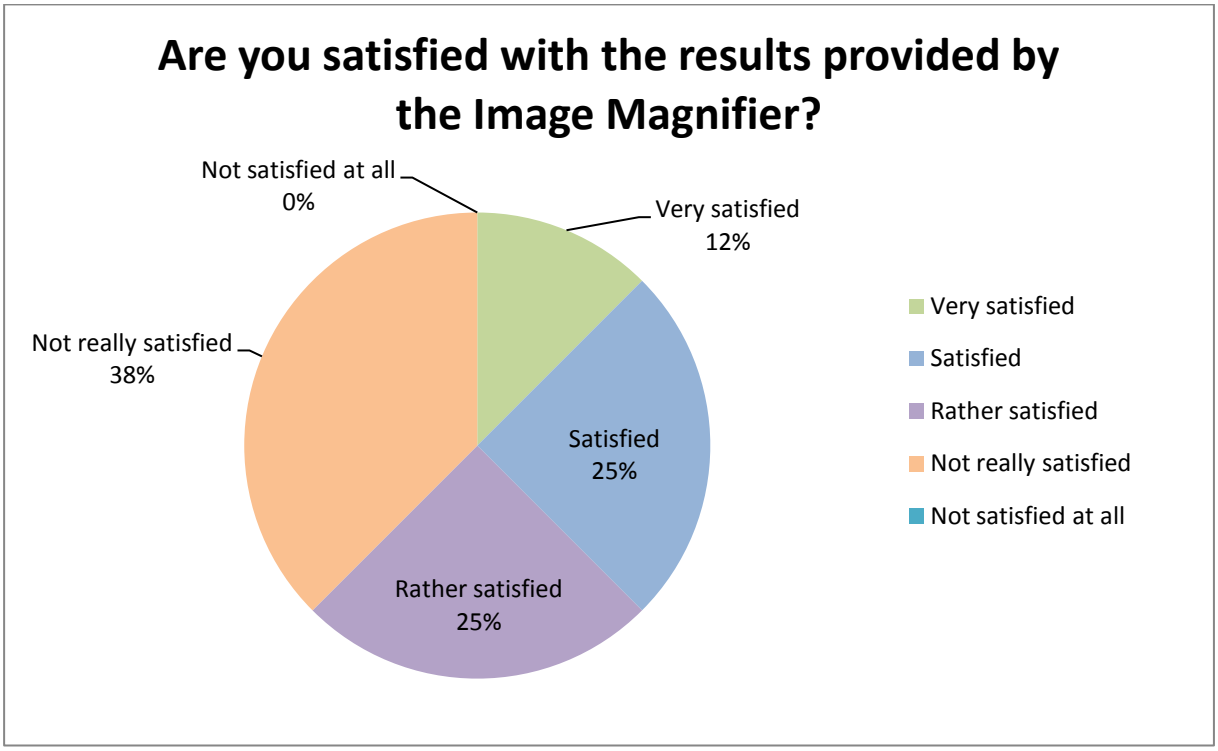

Figure 31: InVID Verification Plugin - external testers - Image Magnifier: Results for the question “Are you satisfied with the results provided by the Image Magnifier?”

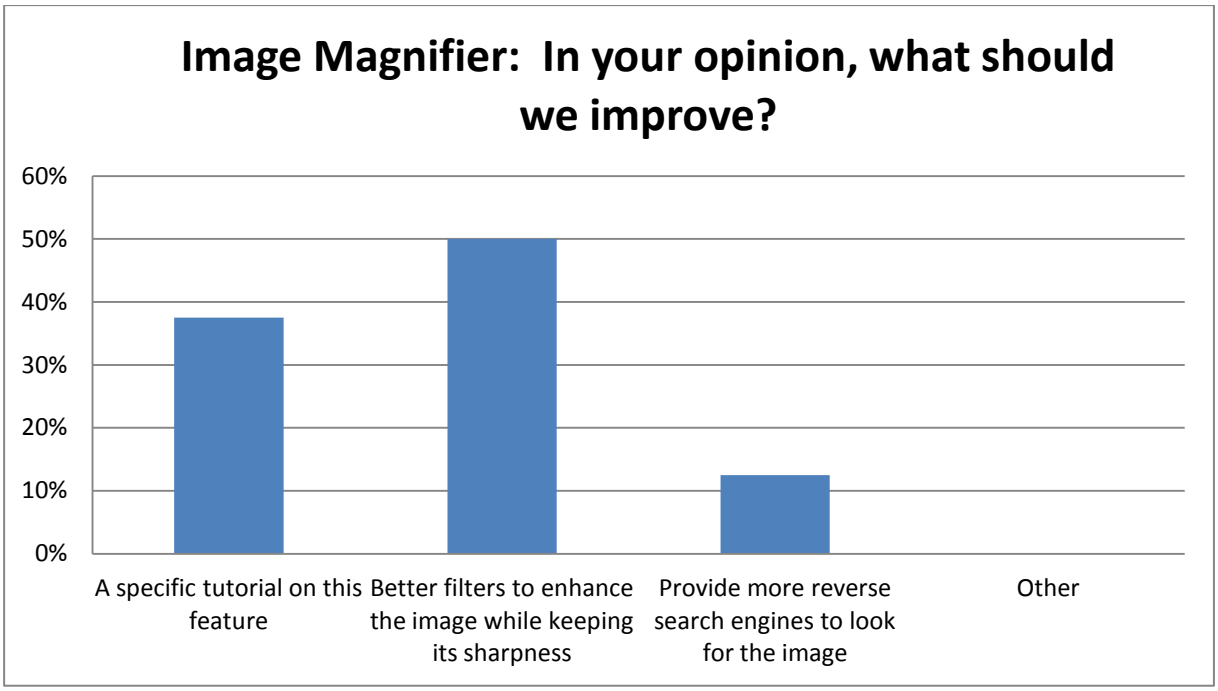

Figure 32: InVID Verification Plugin - external testers - Image Magnifier: Results for the multiple response question “In your opinion, what should we improve?”

Results of the survey for questions regarding the metadata feature

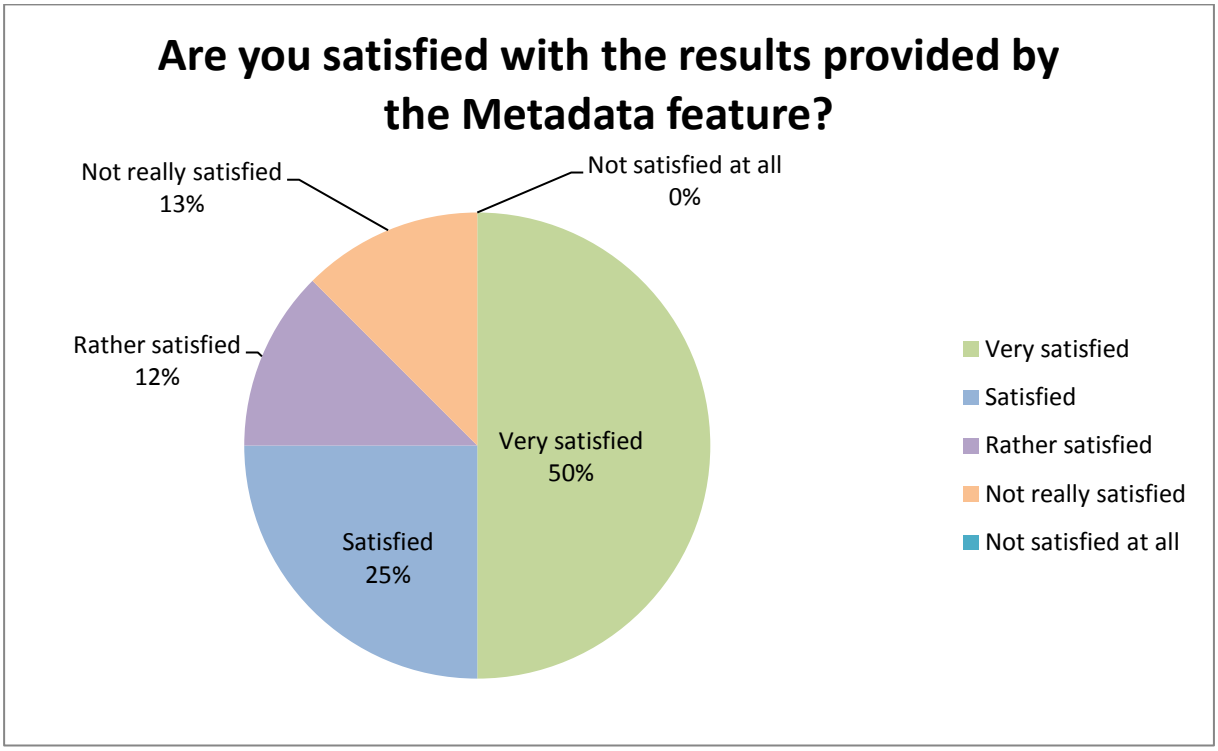

Figure 33: InVID Verification Plugin - external testers - Metadata: Results for the question “Are you satisfied with the results provided by the Metadata feature?”

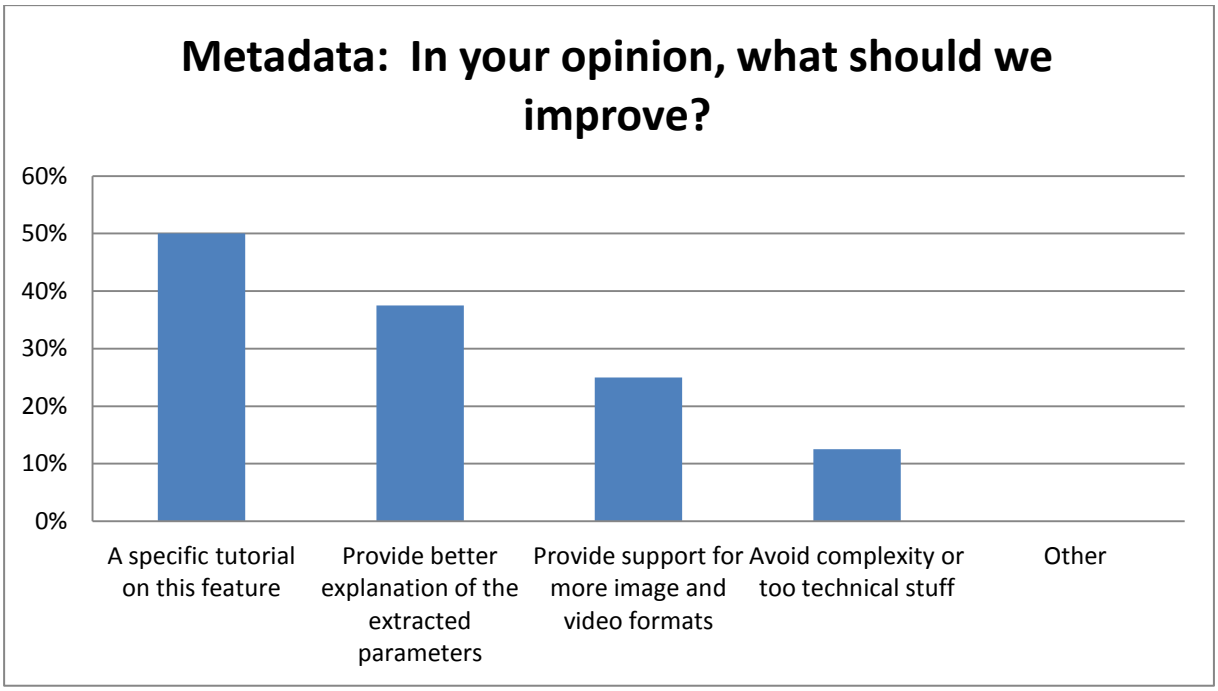

Figure 34: InVID Verification Plugin - external testers - Metadata: Results for the multiple response question “In your opinion, what should we improve?”

Results of the survey for questions regarding the image forensic feature

The InVID Verification plugin integrates the Image Verification Assistant from the project REVEAL – “REVEALing hidden concepts in Social Media”. Therefore the feedback in Figure 35 and Figure 36 refers to the REVEAL project.

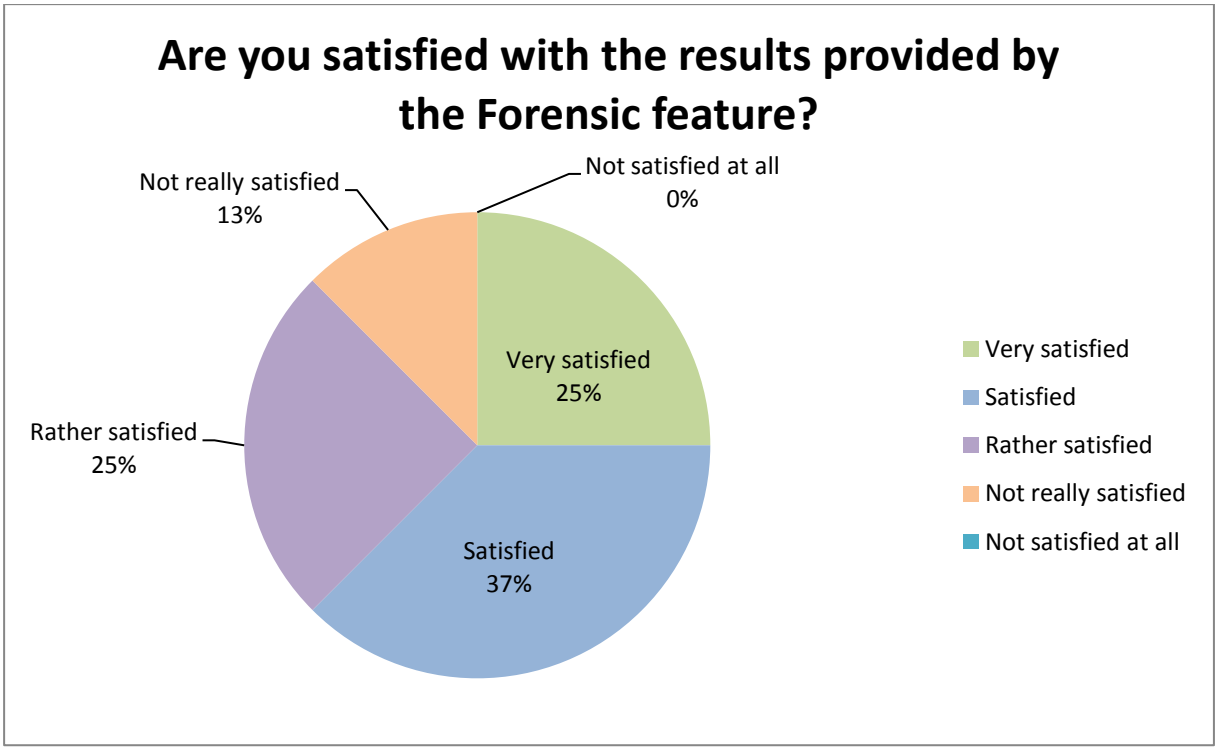

Figure 35: InVID Verification Plugin - external testers - Image forensic feature: Results for the question “Are you satisfied with the results provided by the forensic feature?”

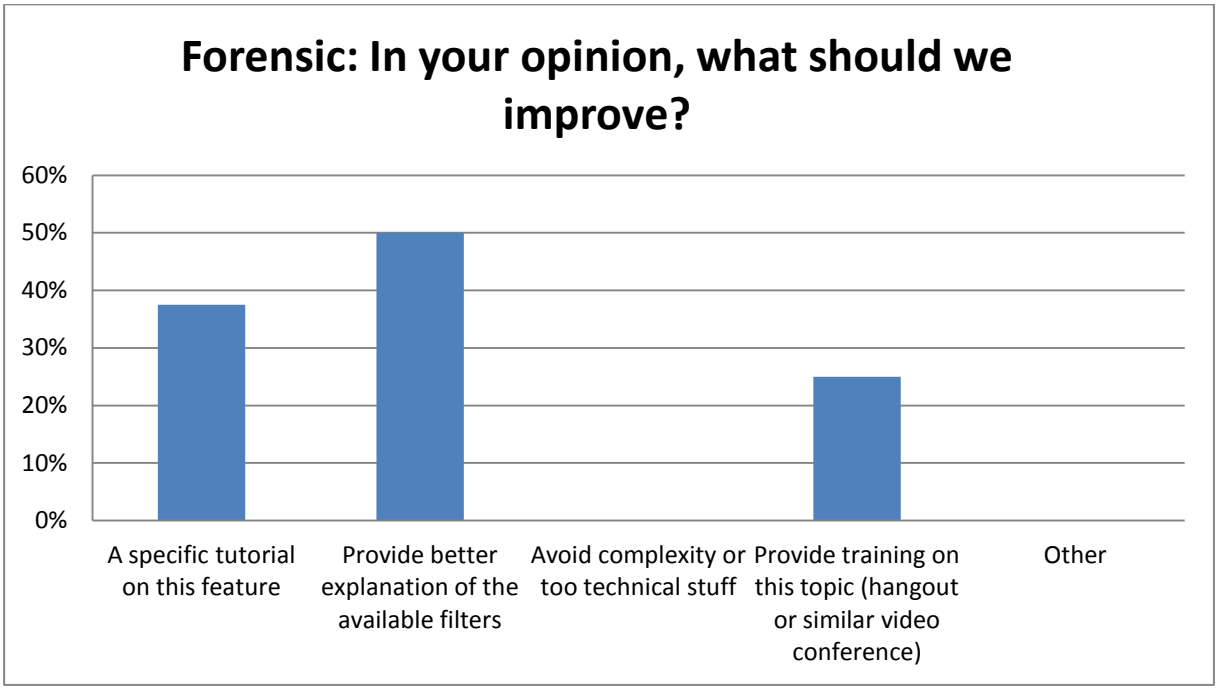

Figure 36: InVID Verification Plugin - external testers - Image forensic: Results for the multiple response question “In your opinion, what should we improve?”

3.11 InVID Verification Application

3.11.1 Description of the service

Shown in Figure 37, the InVID Verification Application is a comprehensive web-based technology that enables journalists to assess the reliability and trustworthiness of user-generated videos with the help of several integrated analysis components.

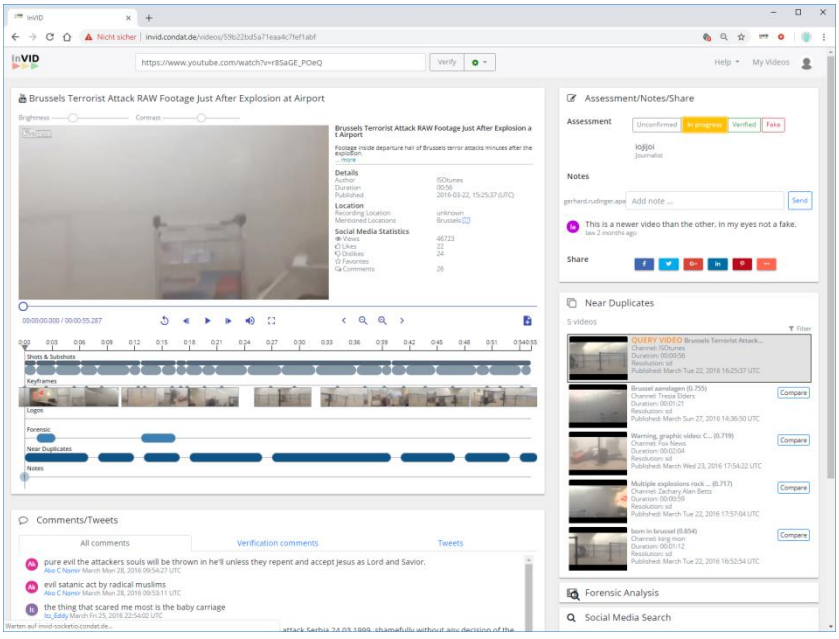

Figure 37: InVID Verification Application

3.11.2 Target groups

The target groups for this service are journalists from media organisations such as publishing houses and broadcasters, members of human rights organisations and other people dealing with videos from social media, with user-generated videos in general or the verification of videos in general.

3.11.3 Tests

The InVID Verification Application was tested by members of the companies of the consortium in test cycle 7. A dedicated testing session by the developers of this integrated technology and the internal testers was also held in test cycle 7 to give structured feedback for improvements on the functionality and usability of the application.

Tests with external users (external to the consortium and external to the companies of the consortium) were done in test cycles 7 to 9. The external testers were lead users from companies such as France24, BBC, Newsy.com, Sky News, Berkeley University, academics from the Journalism Dept. of the Aristotle University of Thessaloniki, and members of the First Draft network. Feedback from the external testers was collected with the help of an online survey that contained 18 questions regarding the usability and functionality of the application. The survey was filled in by 20 testers.

Table 18: Number of received feedback comments for the InVID Verification Application from members of the companies of the consortium

Test cycle	Feedback comments
Test cycle 7	73 items
Test cycle 8	--
Test cycle 9	--

Table 19: Number of received survey responses from external testers

Test cycle	Survey responses
Test cycle 8 to 9	20 survey responses

3.11.4 Major outcomes of the test cycles from members of the companies of the consortium

In test cycle 7 the Verification App was tested extensively by members of the companies of the consortium. The testing and also the dedicated testing session by the developers and the testers resulted in a prioritised list of bugs and improvements that had to be implemented before additional external users could be invited to test the Verification Application.

The list of critical bugs and improvements comprised the following points:

- Certain videos and certain types of videos could not be processed with the Verification Application.
- The spinning wheels showing the status of the video analysis had to be shown more clearly to the user.
- The processing time of the video had to be increased.
- Sometimes the video details showed up only after a long time (>40 seconds) although the “import” task had been finished.
- The keyframes of the segmentations were not visible for certain videos.
- Sometimes only icons for broken images were shown for the keyframes. Reverse image search was also not possible.
- The global forensic filters did not indicate to the user that processing was finished.
- Parts of the verification results were missing in some of the already processed videos due to an error at the time of processing. They had to be reprocessed.
- Videos longer than five minutes had to be allowed.

Additionally, a comprehensive list of bugs and suggestions for improvements with normal and low priority had been provided, e.g. the need to refresh the browser page in certain situations to see new information, icons showing the processing status when there is no processing, suggestions to fill out the author of a new comment with the username of the user logged in, and more.

All major bugs and suggestions were fixed in the following development cycle, as well as many issues with priority “normal” or “minor”.

3.11.5 Major outcomes of the test cycles from external users

The following diagrams give an overview of the findings of the survey from the testers of the application. In total we received responses from 20 external testers. The testers were external to the consortium and also external to the companies of the consortium. Figure 38 shows the gender and age distribution, Figure 39 the occupation and Figure 40 the country distribution.

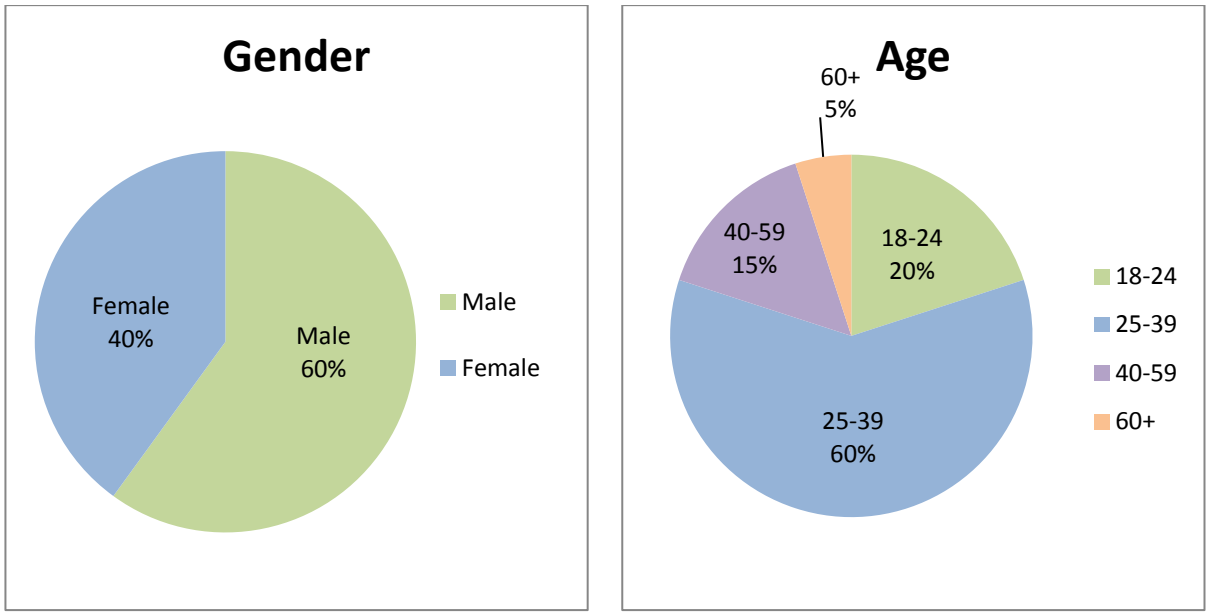

Figure 38: InVID Verification Application – external testers: Gender and age of the participants

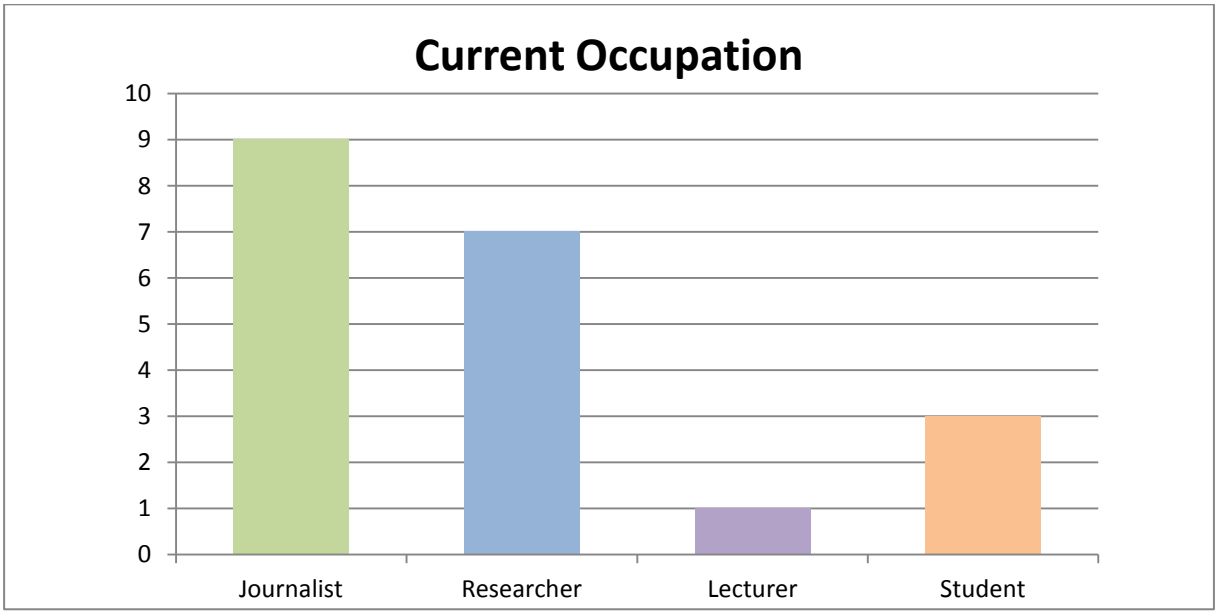

Figure 39: InVID Verification Application – external testers: Occupation of the participants

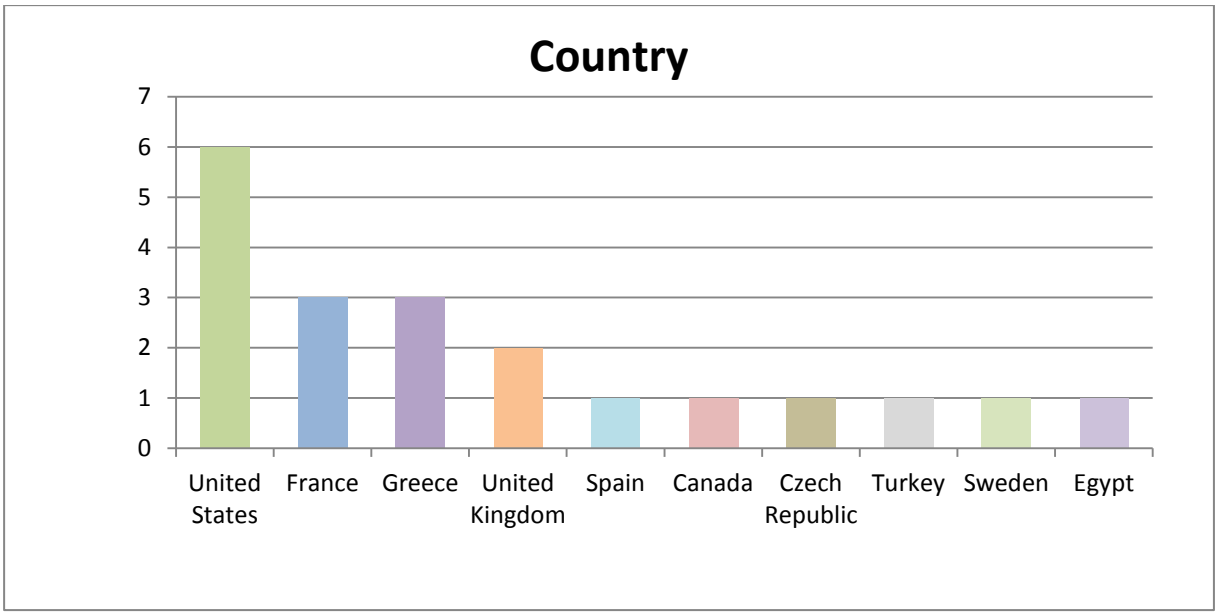

Figure 40: InVID Verification Application – external testers: Countries in which the participants live

The findings of the survey indicate that both the InVID Verification Application and the integrated components were very well appreciated by its users. 90% of the users assessed the Verification Application as “Very useful” or “Useful” and 95% of the users found it easy to use. The other questions regarding usability received positive feedback from 70 to 95% of the users. The responses regarding the helpfulness of the different features have been also very good. 65 to 90% of the users found the different features very helpful or helpful. In the case of the near duplicate component, a somewhat lower percent of users (44%) responded they were satisfied with this component. This not so good, assessment was due to the fact that there was still a relatively small index of videos used for near duplicate search (as compared to the vast amount of videos available on YouTube and other video sharing platforms). The additional feature that automatically performs a video-title-based search on YouTube and indexes similar videos for further processing and assessment by the near duplicate component is expected to work well in a real-life scenario, where several people are interested in the same viral video. But in a testing scenario where the testers are verifying different videos this feature is not fully effective.

Besides the results shown below, the testers also suggested improvements in usability and additional functionality, e.g. for the registration process, the enlarging of images, a translation function, the possibility to upload local videos for analysing, utilities for analysing images and audio, and the suggestion of an online tutorial.

Results of the survey for general questions

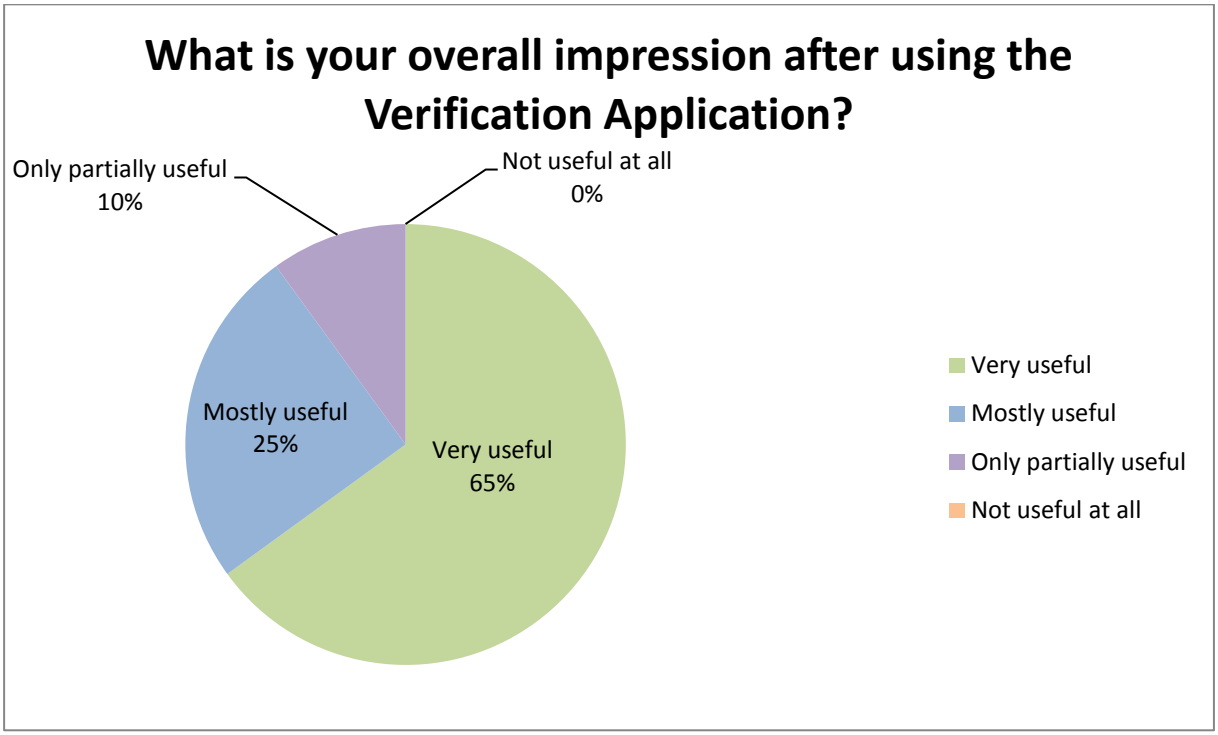

Figure 41: InVID Verification Application – external testers: Overall impression

The testers assessed the InVID Verification Application in general with the following words:

- Very useful for assisting the verification of a video. Somewhat of an all-in-one platform to add to a Digital Sherlock's tool kit.
- It's wonderful! I really appreciate how comprehensive it is and how much effort has clearly gone into making it user friendly.
- I really like the format; it's easy to see an overview of the verification process.
- It's a great, great tool. The key shots and video information alone make it very useful. Brightness and contrast tools seem great.
- It has and will continue to be one of the most valuable tools in my personal toolbox and a critical component of any and all trainings I conduct with journalists. The verification components as a stand-alone product are impressive in their own right. However, the abilities to do advanced social searches sets InVID apart from any other existing application that I am aware of. I have not fully taken advantage of the personal archive of videos made available in the new version of InVID; however, in doing retrospective research and building long-term research questions, it is a highly valuable bookmarking function. I think that the incorporation of video rights is very strategic as far as an adoption strategy goes - offers a free alternative to rights-clearing for-profit business models such as Storyful and other social media clearinghouses.
- I really like the application. I would just hope that it would be able to fill out the different categories and features with information.

- It is good, but it requires some training in order to leverage every feature of the tool. Such as the forensics, even after reading the description you might still be a bit puzzled if you haven't used it several times before.
- This tool is useful for our field because I don't need to individually search YouTube, then Twitter, then Google, then Bing, then Yandex etc. by going to each individual website. Everything is contained within one location and makes the process of verification quicker because I can write the title of the video and simply click YouTube or Bing as needed. In addition, I like the overall structure of organizing by category and I like that the colours are very neutral because it is less distracting. I'm a bit confused about what the button next to "verify" is supposed to do.
- Clearest than the previous version, but as well maybe too "expert", seems very time-consuming
- Friendly, flexible, and responsive.
- When the bugs get sorted out, this is going to be amazingly helpful. I hope the near duplicates function will work well. And I really like that we can search all the reverse search engines so easily - that's going to save us a lot of time!
- A really great tool we will be using most weeks
- Very useful
- Useful
- Quite strong.
- Good.
- Not during every day work routine, but sometimes for sure.

Results of the survey for questions regarding usability

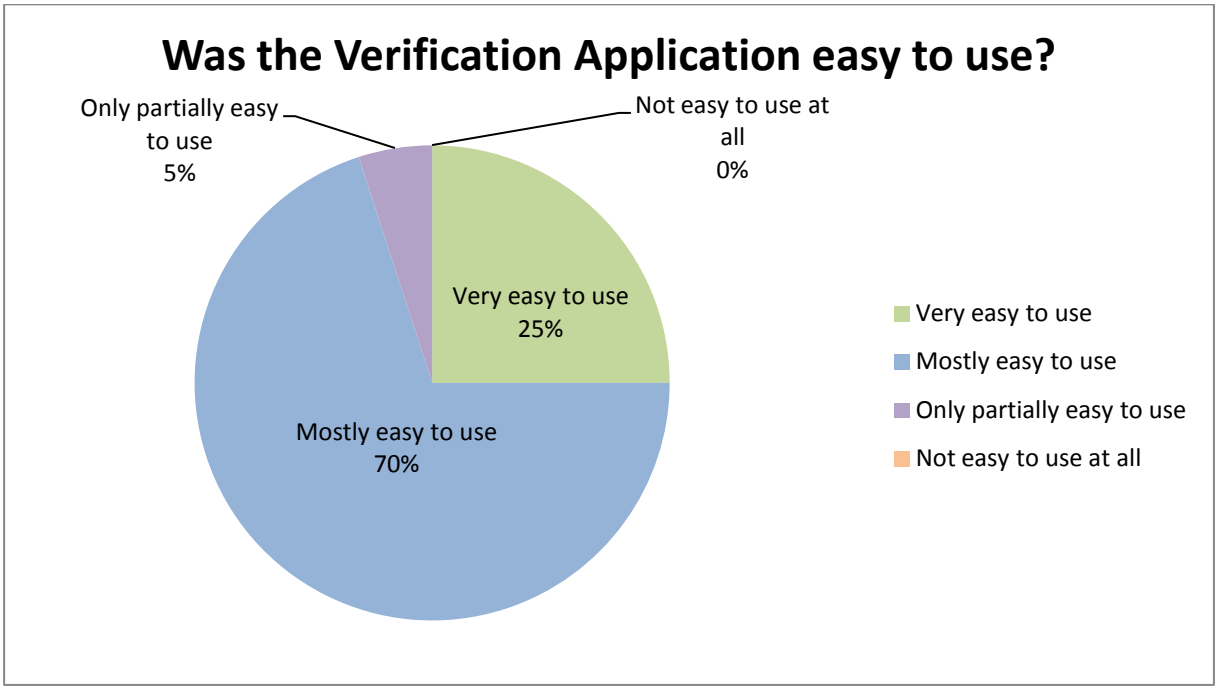

Figure 42: InVID Verification Application – external testers: Results for the question "Was the Verification Application easy to use?"

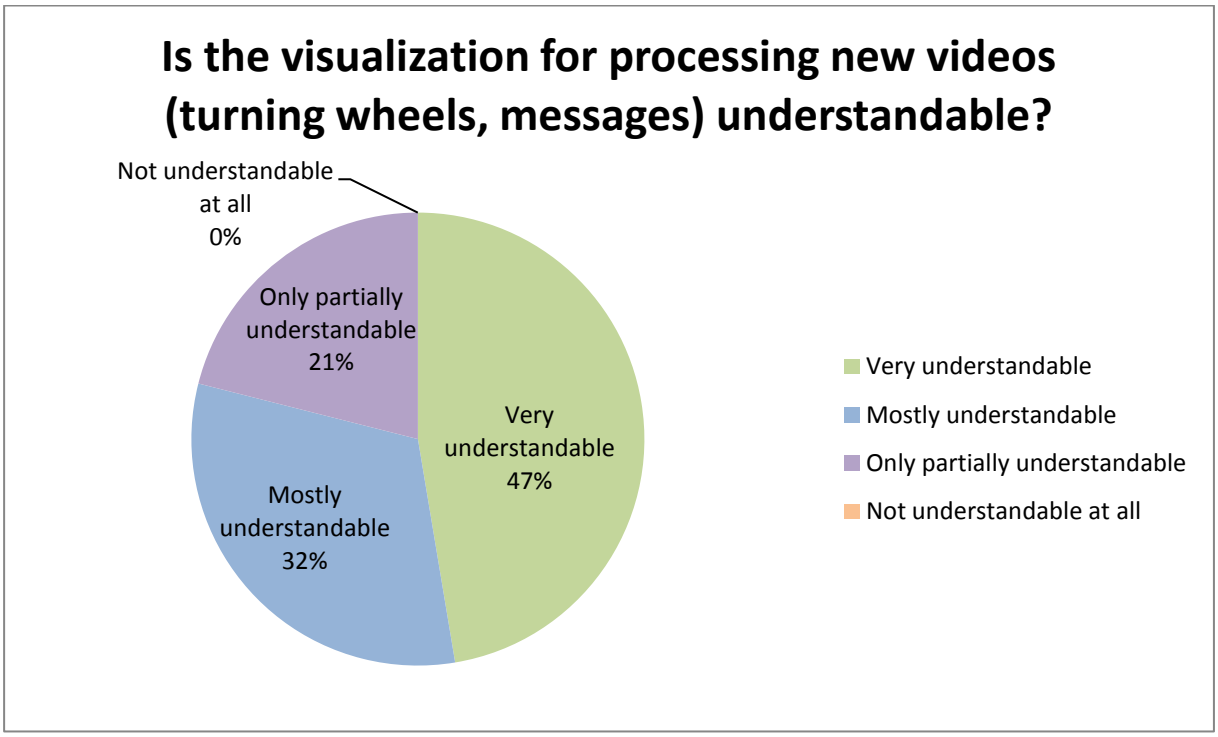

Figure 43: InVID Verification Application – external testers: Results for the question "Is the visualization for processing new videos (turning wheels, messages) understandable?"

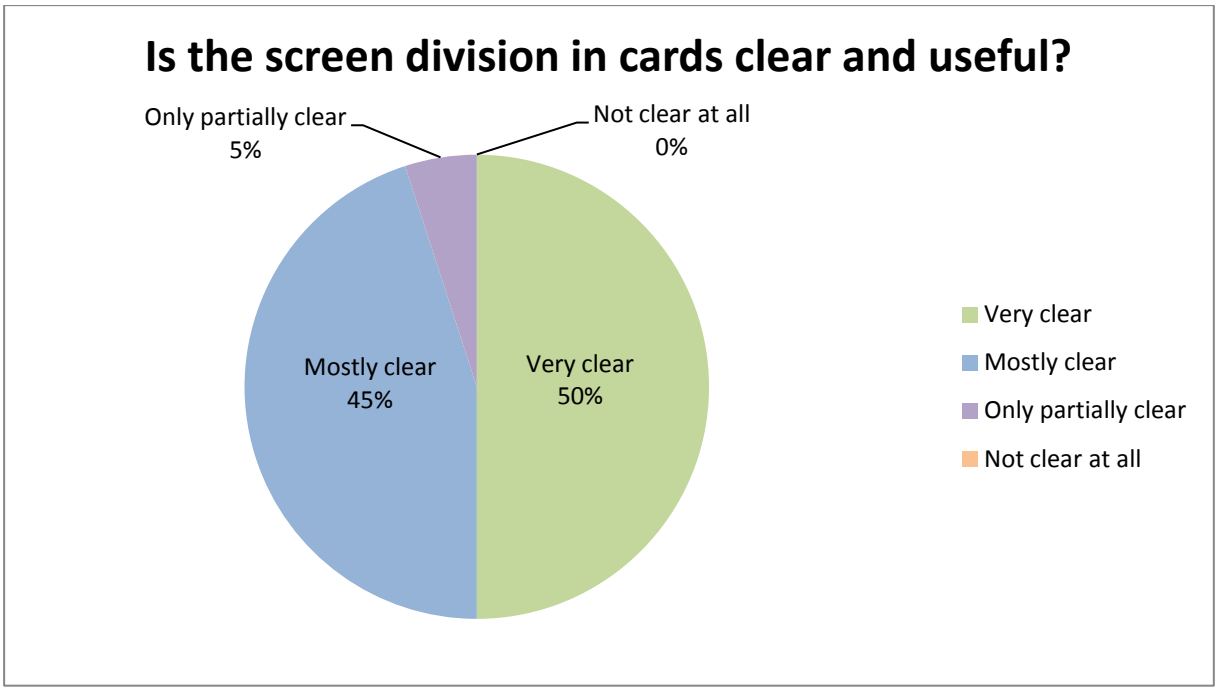

Figure 44: InVID Verification Application – external testers: Results for the question "Is the screen division in cards clear and useful?"

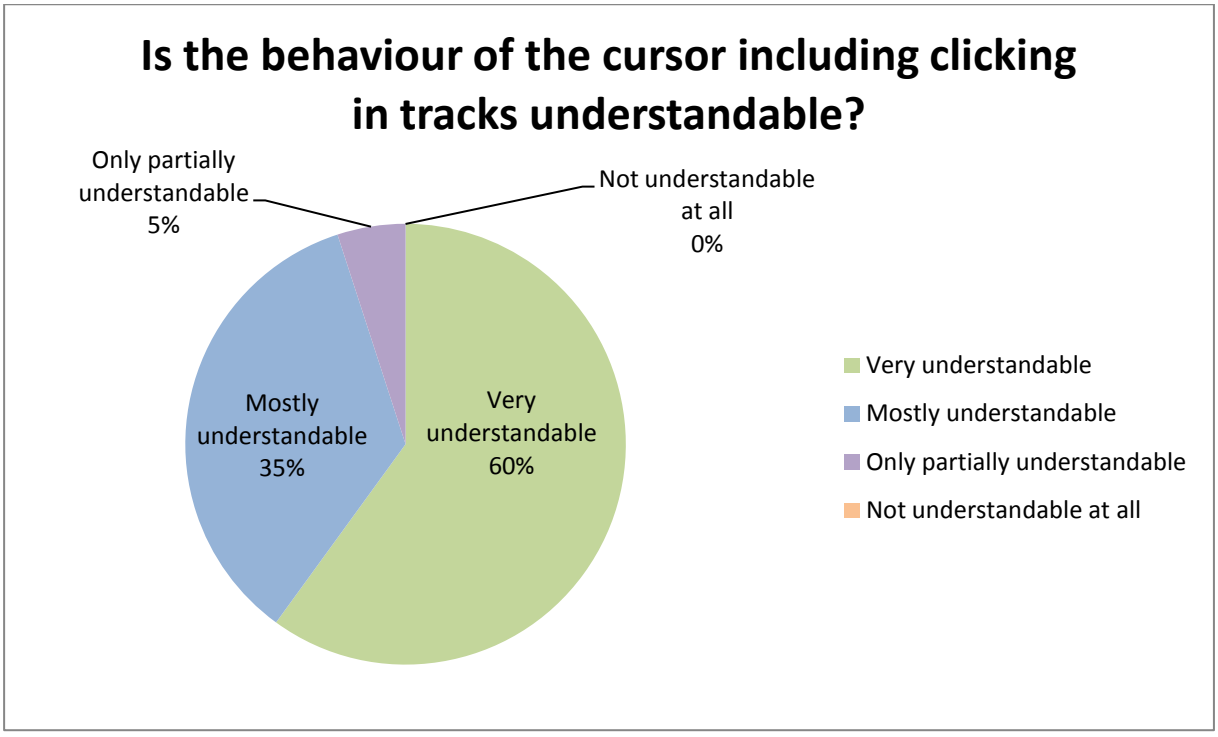

Figure 45: InVID Verification Application – external testers: Results for the question "Is the behaviour of the cursor including clicking in tracks understandable?"

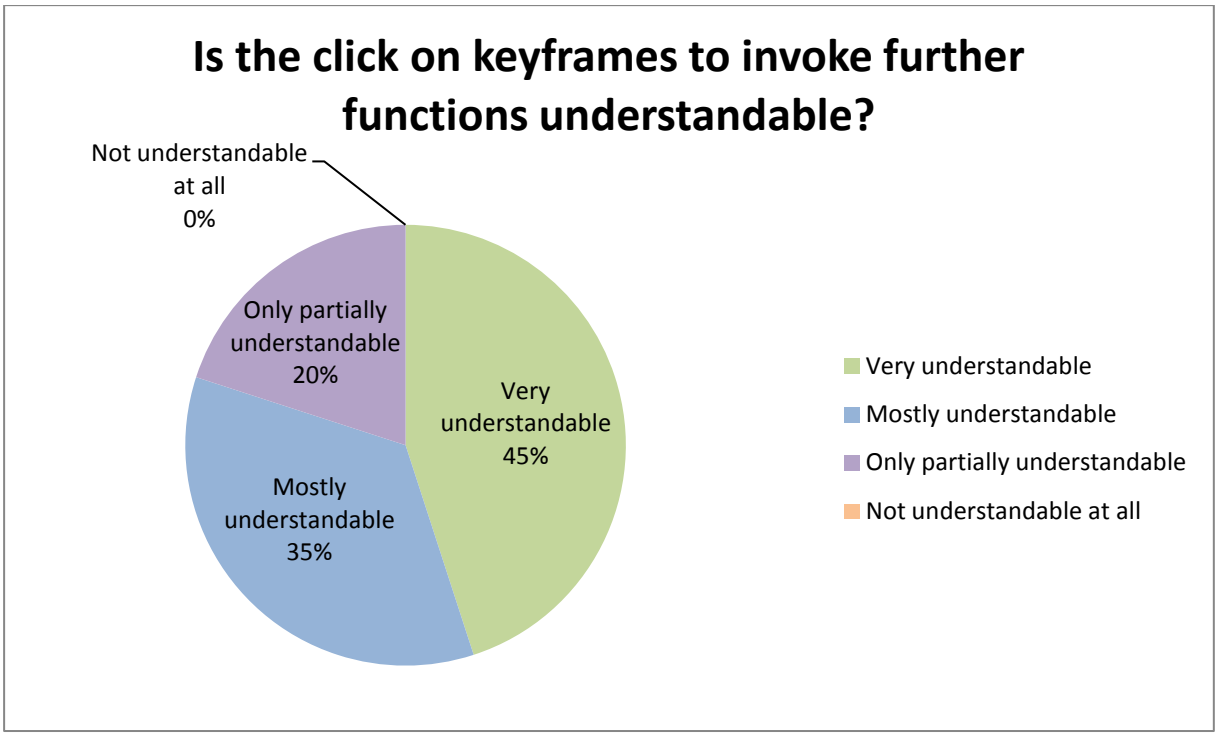

Figure 46: InVID Verification Application – external testers: Results for the question “Is the click on keyframes to invoke further functions understandable?”

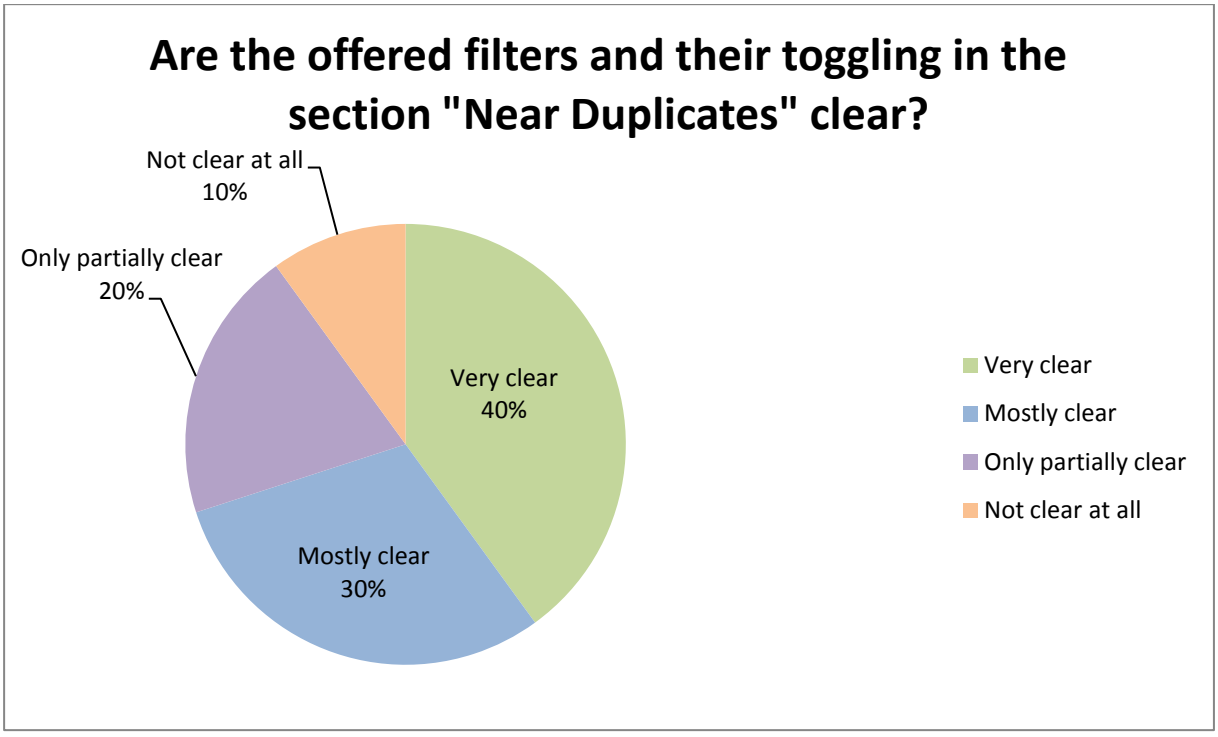

Figure 47: InVID Verification Application – external testers: Results for the question “Are the offered filters and their toggling in the section "Near Duplicates" clear?”

Results of the survey regarding the different features of the Verification Application

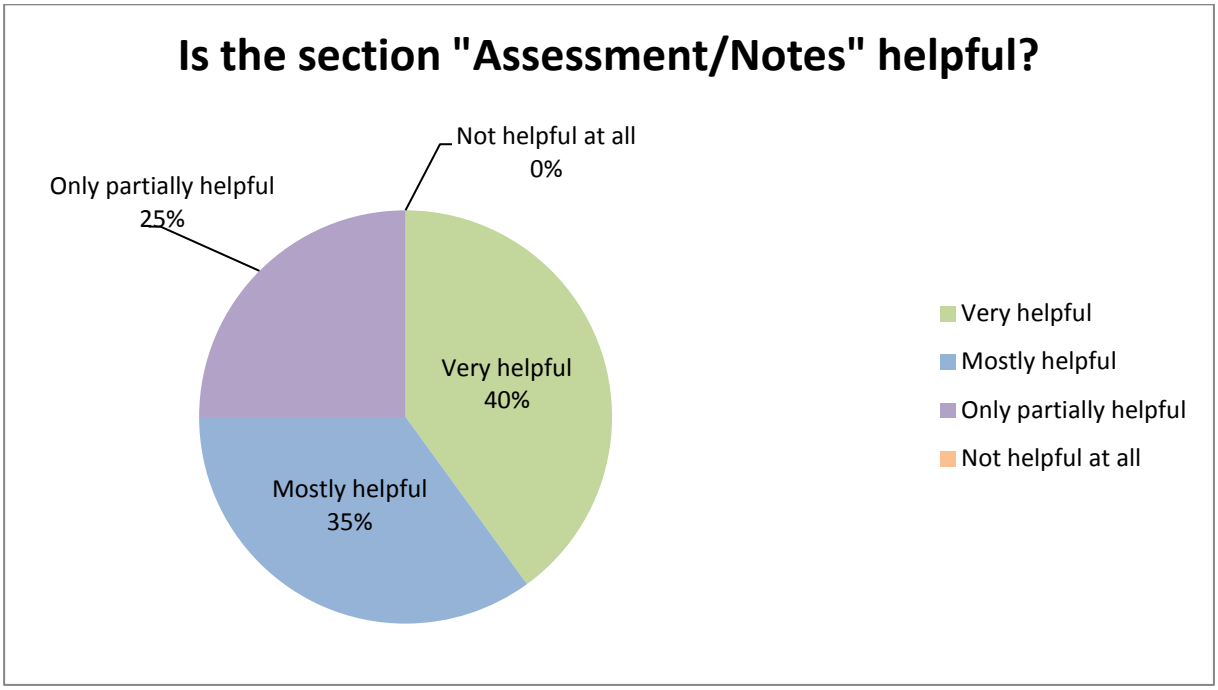

Figure 48: InVID Verification Application – external testers: Results for the question “Is the section "Assessment/Notes" helpful?

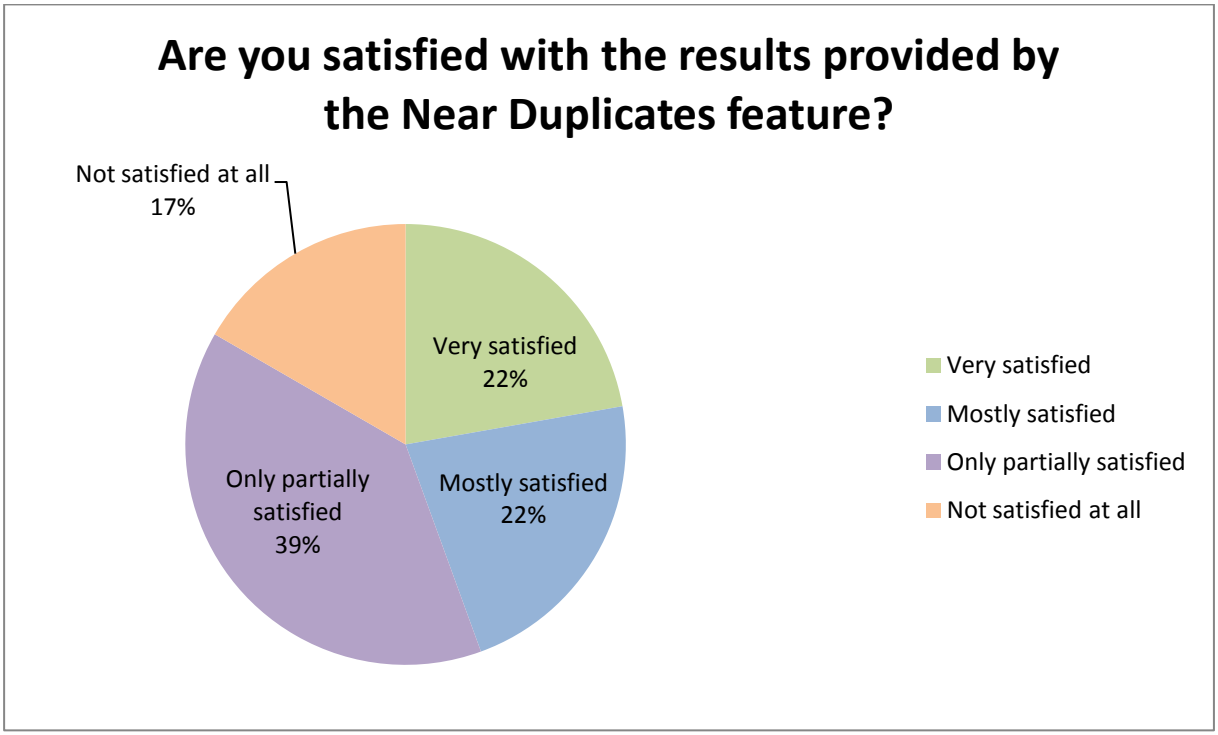

Figure 49: InVID Verification Application – external testers: Results for the question “Are you satisfied with the results provided by the Near Duplicates feature?”

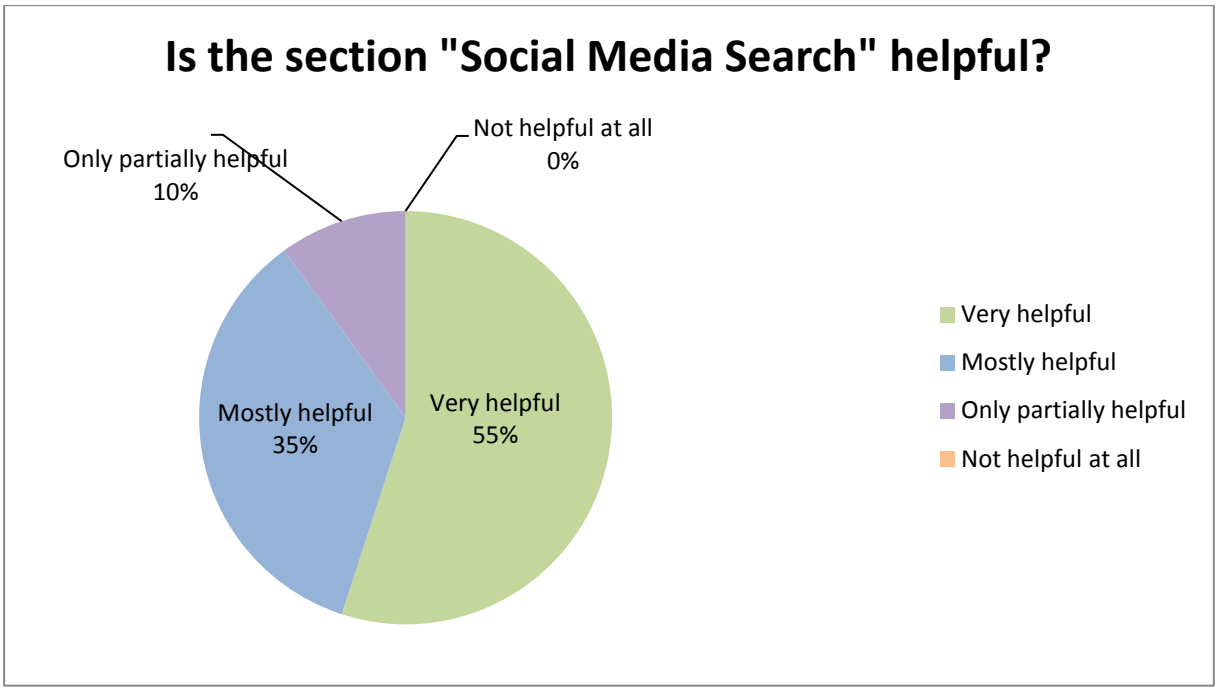

Figure 50: InVID Verification Application – external testers: Results for the question “Is the section "Social Media Search" helpful?”

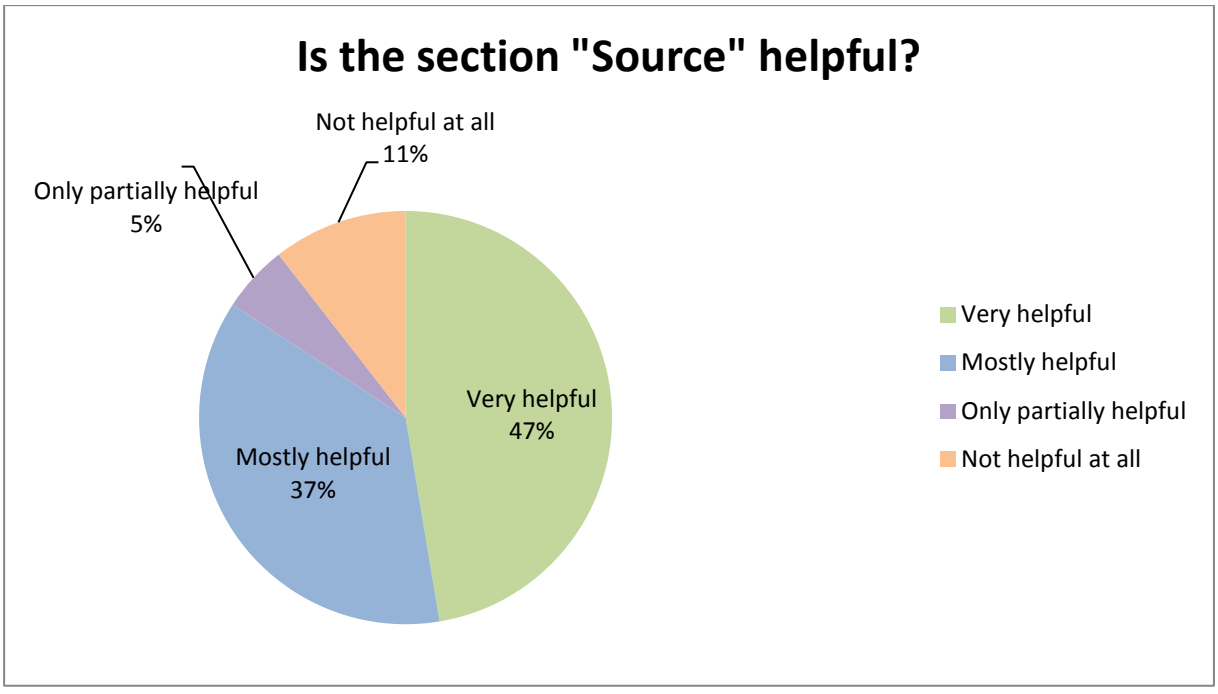

Figure 51: InVID Verification Application – external testers: Results for the question “Is the section "Source" helpful?”

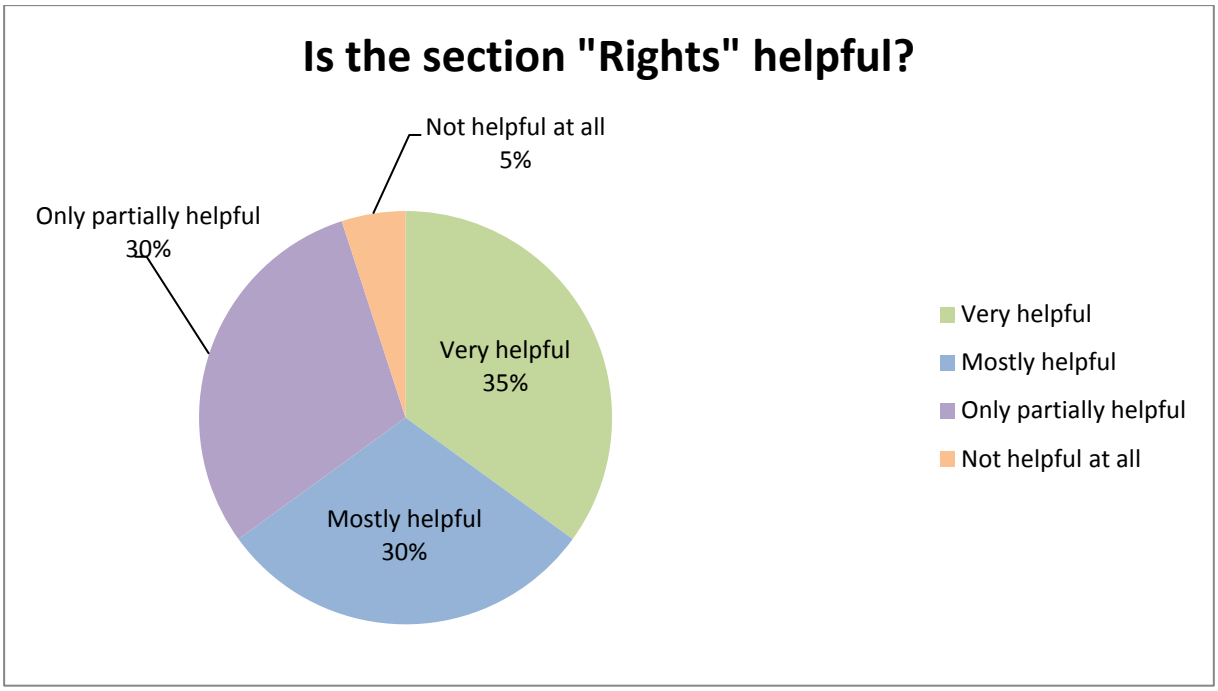

Figure 52: InVID Verification Application – external testers: Results for the question “Is the section "Rights" helpful?”

3.12 InVID Mobile Application

3.12.1 Description of the service

The Mobile Application (see Figure 53) has been developed to allow communities of non-journalist users of a journalistic service (e.g. registered users of a news web portal or the web edition of a newspaper or members of an emergency service) to contribute user-generated videos directly to the news organisation that provides this service. Uploaded videos are forwarded to the editorial UGC management system which also provides integration with the InVID Verification Application.

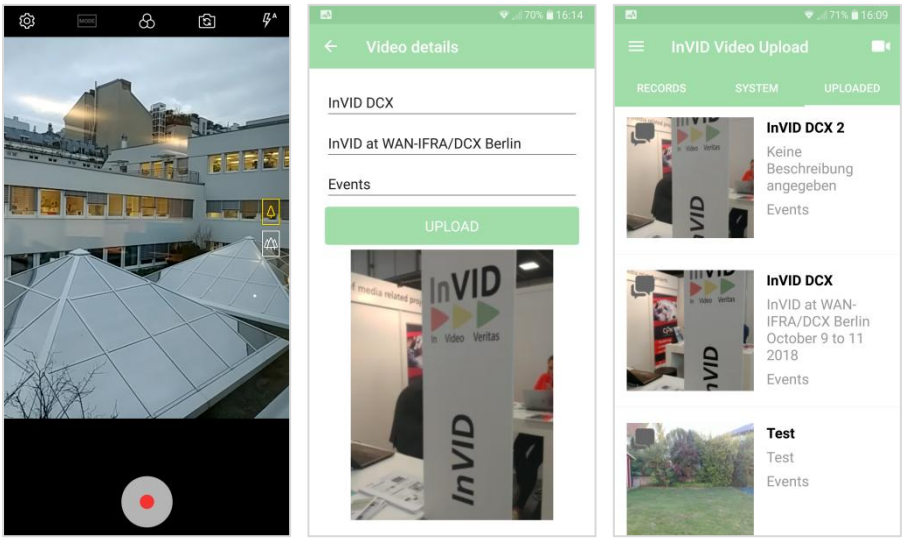

Figure 53: InVID Mobile Application

3.12.2 Target groups

The target groups for the InVID Mobile Application are media organisations, such as publishing houses that want to collaborate with their user community on UGV - either with their entirety or just with specific user groups (e.g. members of emergency services).

3.12.3 Tests

The InVID Mobile Application was tested inside the consortium in all three test cycles covered in this document (test cycles 7 to 9). Tests with external users were done in test cycles 7 and 9. The external testers were editors of the Tiroler Tageszeitung (local newspaper in Austria) and members of the emergency service Österreichische Wasserrettung Vorarlberg (Austrian water rescue service Vorarlberg). An online survey was used to gather the assessment and suggestions for improvements from the external testers.

Table 20: Number of received feedback comments for the InVID Mobile Application from testers inside the consortium

Test cycle	Feedback comments
Test cycle 7	12 items
Test cycle 8	9 items
Test cycle 9	2 items

Table 21: Number of received feedback comments for the InVID Mobile Application

Test cycle	Survey responses
Test cycle 7 and 9	7 survey responses

3.12.4 Major outcomes of the test cycles from members of the consortium

The testing of the InVID Mobile Application by members of the consortium in test cycles 7 to 9 focused on the functionality of the InVID Mobile App, the Administration Client, the integration of the InVID Mobile Application with the APA CMS (APA-OnlineManager) and the integration of the APA CMS with the InVID Verification Application.

Major bugs that were found concerned crashes of the Mobile Application in certain situations, errors in displaying character sets like Arabic, as well as problems in transferring videos with Arabic letters in the description to the Verification Application. Several errors were found in the communication process between the editor and the UGV providing user especially with notifications and the display of messages. In the administration client, errors were found concerning the search functions and the display of the results list.

The feedback from the test cycles has been taken into consideration in the following development cycle and has been retested by the users in the next test cycle.

3.12.5 Major outcomes of test cycles from external users

The external tests were done by editors of the Tiroler Tageszeitung (local newspaper in Austria) and members of the emergency service Österreichische Wasserrettung Vorarlberg (Austrian water rescue service Vorarlberg). In total, seven survey responses were submitted. Figure 54 shows the gender and age distribution of the testers and Figure 55 the functions of the testers.

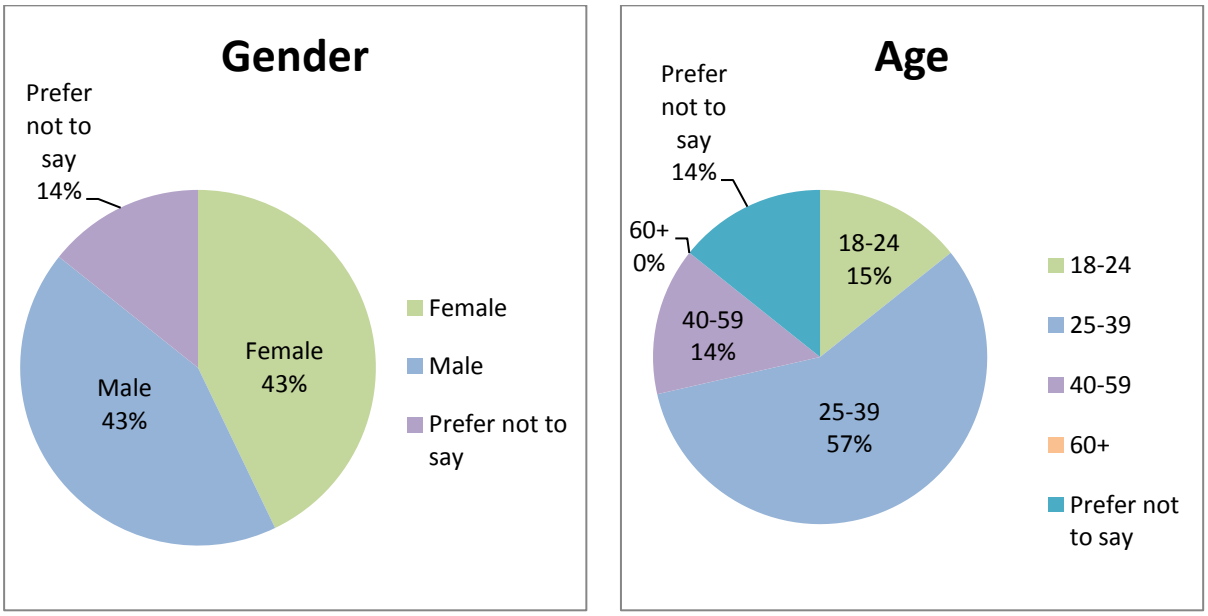

Figure 54: InVID Mobile Application– external testers: Gender and age of the participants

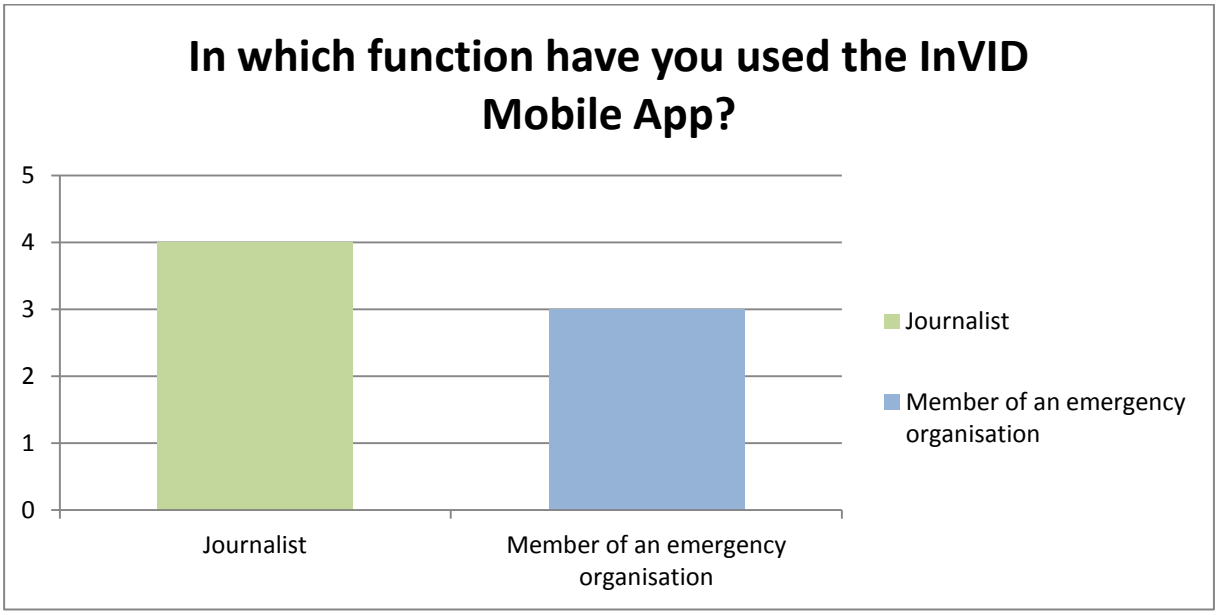

Figure 55: InVID Mobile Application– external testers: Results for the question “In which function have you used the InVID Mobile App?”

The findings of the survey indicate that the InVID Mobile Application was very well appreciated by its users. In a free text question, the testers assessed the InVID Mobile Application as “Good”, “Very good” and “Useful”. The Mobile Application provides direct access to editors of a newspaper for UGV-providing users via an uploading of a video. This was rated as very helpful or helpful by all users. The Mobile Application was rated as very easy to use or easy to use by all testers, and the recording and uploading of a video was also easy and understandable. None of the testers had problems using the Mobile Application. In a free text question the testers suggested to add an editing possibility for the recorded videos, which will be considered for future development. Also the testers suggested minor improvements in usability and reported a single video that was uploaded with the Mobile Application and then shown upside-down in the APA-CMS (APA-Online Manager).

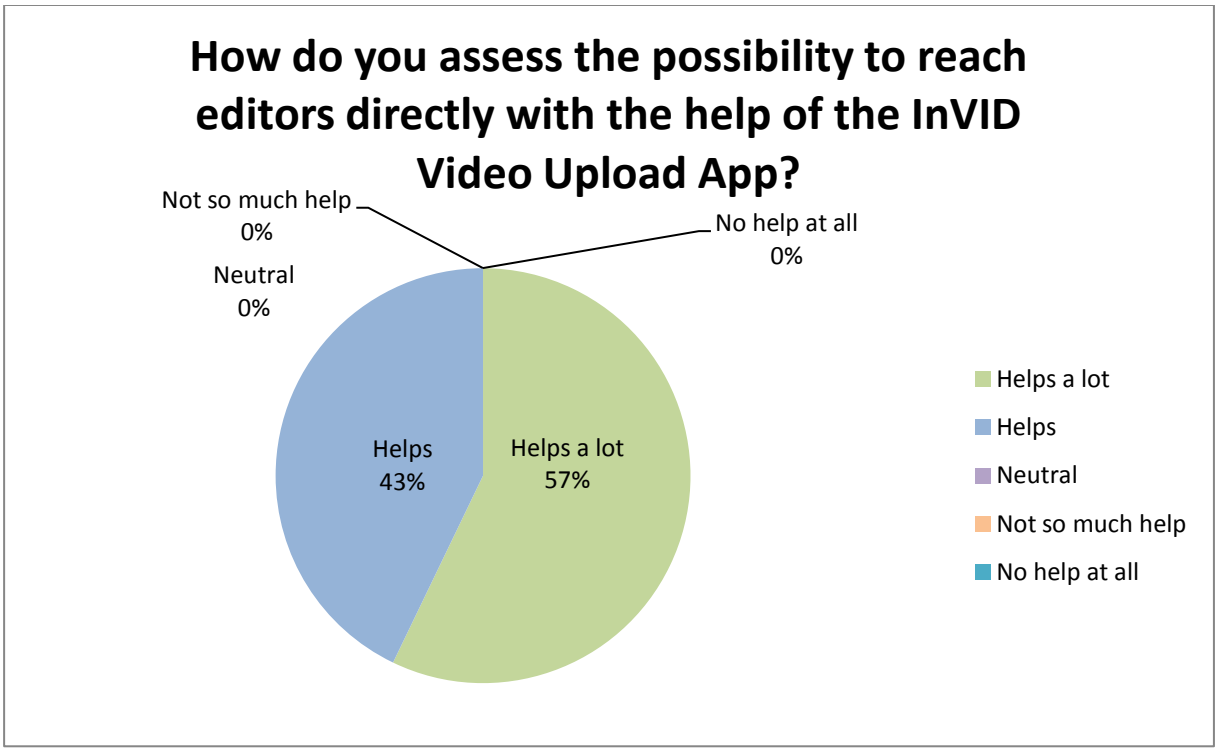

Figure 56: InVID Mobile Application – external testers: Results for the question “How do you assess the possibility to reach editors directly with the help of the InVID Video Upload App?”

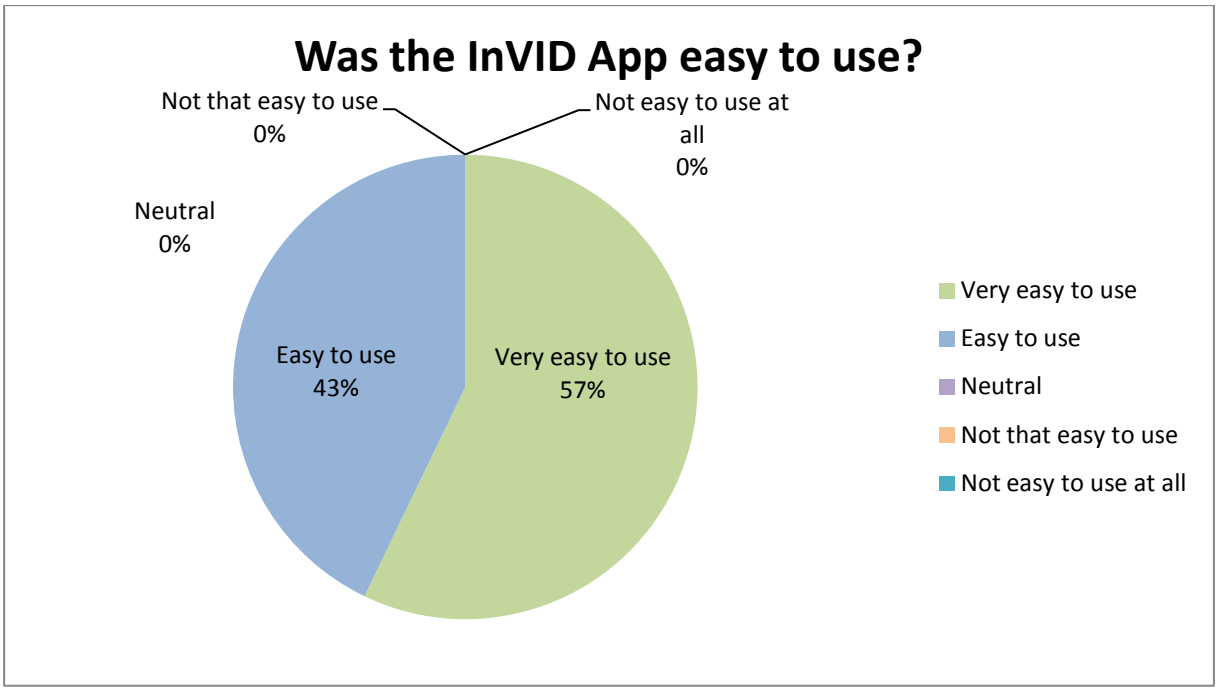

Figure 57: InVID Mobile Application – external testers: Results for the question "Was the InVID App easy to use?"

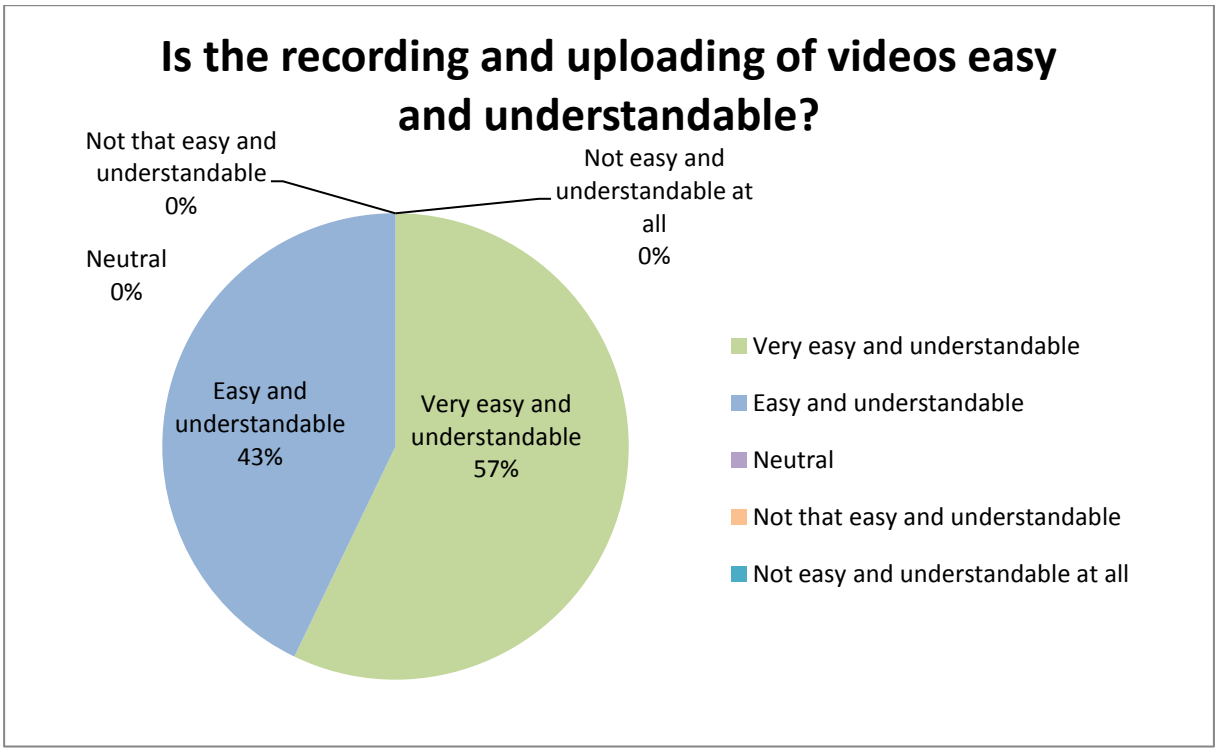

Figure 58: InVID Mobile Application– external testers: Results for the question "Is the recording and uploading of the video easy and understandable?"

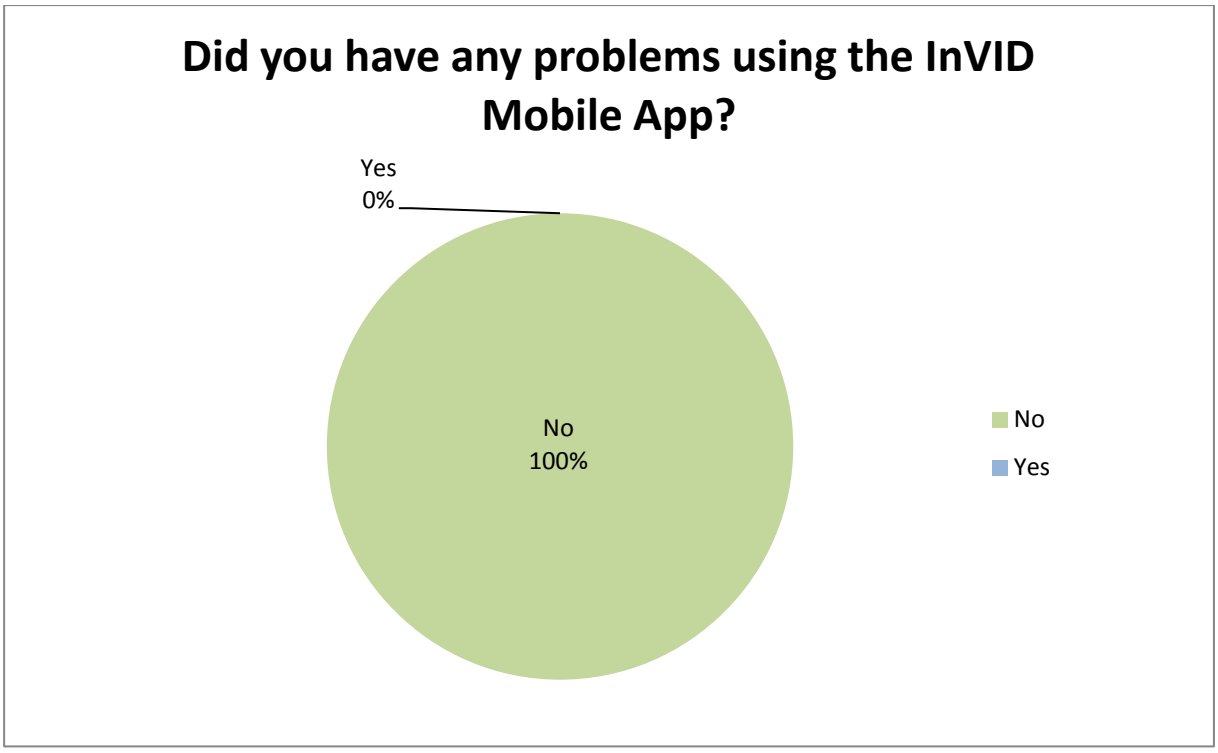

Figure 59: InVID Mobile Application– external testers: Results for the question “Did you have any problems using the InVID Mobile App?”

3.13 InVID Core Platform API

3.13.1 Description of the service

The InVID Core Platform is the central metadata repository that acts as a data exchange platform for all InVID applications and components. It also provides analysis services such as the ones supported by the Video Fragmentation and Annotation Service and the Recognyze module via an API.

3.13.2 Tests

The InVID Core Platform API had reached a mature state prior to test cycle 7. Therefore no explicit tests were done in the test cycles 7 to 9. As the InVID Core Platform provides major services to many InVID components and applications, the service has been tested indirectly by tests of these applications, such as the InVID Verification Application.

4 Summary of the pilot testing throughout the project's lifetime

4.1 Integration of the pilot testing in the development cycles

Pilot testing was an essential part of the InVID development in the whole project's lifetime. InVID adopted an agile development approach in order to promote the speedy development of both InVID individual technologies and applications that optimally addressed the industry needs and satisfied the innovation goals set within the project. The partners worked in successive development cycles, which included (re-)prioritizing the industrial requirements; assessing, selecting and adapting individual technologies that best addressed these requirements; exposing and integrating them to the platform and applications. The platform and applications were then tested and validated in the pilots. Based on the results obtained from the pilot-testing, the industry requirements were redefined. The results from the pilot testing and the updated industrial requirements provided the input for the next cycle of platform and application development. The duration of a complete development cycle was three months. Nine development cycles were realized in the project beginning in month 10 of the project.

4.2 Overview of the conducted pilot testing

The primary aim of the test cycles was to collect user feedback both on the tools and components themselves, their usability and appropriateness for performing various tasks, and the overall effectiveness of the InVID system in terms of reliability and accuracy. The testing was focused on video breaking news emerging from social networks and media web sites, as well as on the exploration of news stories, such as images coming from conflicts or disasters.

The tests were targeted at API and user interface level (integration testing) as well on the integration of systems (system testing) to ensure the interoperability of the different InVID components and applications. The alpha testing of the components was done inside of the consortium, whereas beta testing was done by users outside of the consortium. In each test cycle, functional tests, non-functional tests and performance tests, as well as regression tests, were executed. Exploratory testing, fault based techniques, scenario testing and walkthrough testing were used as testing techniques. For tests on API-level, automatic tests were also implemented to ensure the functionality of the tested APIs and reduce time and effort for re-testing the APIs after a change.

Each test cycle consisted of a preparation phase, the testing itself and a follow-up phase.

In the preparation phase, the consortium decided which applications and components would be included in the test cycle. A stable version of each included component was then provided by the developers. All the data required to gain access to the exposed technologies, such as URLs and access credentials were gathered. The documentation for the testers was updated

with the changes that had been made since the last tested version. Testing guidelines were established and the report templates for recording the outcomes of the evaluations set up.

In the testing phase, the fixed bugs and implemented suggestions from an earlier test cycle were re-assessed. The different applications and components were tested based on the guidelines and evaluation scenarios that had been determined by the developer(s) of each technology. The evaluation outcome was reported with the help of the provided testing templates.

In the follow-up phase, the results collected from the different testers were merged and provided to all technology partners of InVID. A conference call among the technology providers and the technology testers (involving only members of the InVID consortium but not external testers) was set up in four test cycles to clarify potential questions and comments regarding the test results. In another two test cycles the project meetings were used to clarify questions regarding the test results.

The InVID applications and systems were pilot-tested in real-world environments with different user groups. First, by members of the consortium with a journalistic background (AFP and DW), second, by journalists of the companies from the consortium that were not connected with the project (also AFP and DW) and third, by external testers such as journalists, researchers and data scientists, not connected to the companies from the consortium. The external testers were lead users from companies such as France24, BBC, Storyful, Newsy.com, Sky News, Berkeley University, academics from the Aristotle University of Thessaloniki, and members of the First Draft network. Tests were also done with regional newspapers in Austria (Tiroler Tageszeitung, Vorarlberger Nachrichten) and emergency organisations (Vorarlberger Wasserrettung). IT specialists of the companies from the consortium tested the components of the project on API level. The diversity of the pilot testers (news agencies, broadcaster, newspapers and researchers), the diversity of channels that these industry institutions use for communicating News content (broadcast, Web, print, B2B) and the diversity of News topics that they deal with (world, national, regional) ensure the thorough validation of the InVID solutions.

The results of the evaluations in the test cycles were gathered with different methods like direct observations, interviews, questionnaires and online surveys and direct sessions between testers and developers. The evaluations followed established criteria and were made “fit for purpose” for the respective evaluation tasks.

The feedback from the test cycles was summarized in reports with a special focus on fit-for-purpose, functionality, performance and usability of InVID components as well as the Verification Application and other InVID applications/systems as a whole. The results were used for re-prioritizing the InVID development work in subsequent iterations of the project cycles. As the testing was started very early in the project’s lifetime, it allowed us to continuously monitor and re-assess the effectiveness and maturity of each involved technology and to take corrective actions (including the adoption of a different technology, if necessary) in a timely and efficient manner.

4.3 Focus of the pilot tests in the different test cycles

The work in WP7 already started before the first test cycle with the definition of the testing process in the consortium. This included the breaking down of the testing into different phases (preparation phase, the testing itself and a follow-up phase) and the definition of the different roles and tasks in this process. The testing environment was defined and set up. Further improvements both in the testing process and in the testing environment were enhanced during the test cycles.

Already in test cycles 1 to 3, the tests were done from a journalistic point of view. But the main focus in test cycles 1 to 3 was on the functionality of the different applications and components. In test cycle 3, a dedicated testing session between testers and developers provided a first structured feedback on the usefulness of the different tools for the workflow of a journalist. Shortcomings of the application at that time were pointed out. In test cycles 4 to 7, the main focus of the tests shifted to the assessment of how the tools could help a journalist in the task of verification and rights management. Feedback on a detailed level was also provided in the test cycles 4 to 7 for functional and non-functional (e.g. performance, usability) features of the evaluated technologies. First external testers were included in test cycle 4 and 5 for the Verification Plugin and in test cycle 5 for the Verification Application. In test cycles 6 and 7, the focus was also on getting all tools and all of their functions ready for external testers. In the final test cycles 8 to 9, the focus was on testing with external users. Detailed tests on functional and non-functional features were also made and reported test cycles 8 to 9. In all test cycles, extended evaluations of the exposed APIs ensured a stable connection between the different components and services of the InVID platform.

4.4 Summary of the tests of the different components and applications

4.4.1 General remarks

The results of the individual test cycles were considered in the following development cycles and then re-tested in the next test cycles for all components and applications.

4.4.2 Testing of the Video Fragmentation & Annotation Service

The Video Fragmentation & Annotation Service was tested directly in the test cycles 1 to 4 as well as 6 to 7. Further tests were done indirectly through the integration of the service into the Verification Application. The tests in the earlier part of the project pointed out shortcomings in the documentation and suggested major improvements in the structure of the API. The following tests focused on functionality, reliability and performance. The compatibility with the supported video platforms was examined and problems with different video types e.g. Facebook videos were reported. A delay in the processing was reported, and a major bug regarding the retrieved thumbnails was found. In addition, many suggestions for the improvement of the interface structure, the documentation and the

functionality were made. The cycles of development, testing and improvements lead to a mature, reliable and responsive API.

4.4.3 Testing of the Web Application for Video Fragmentation & Reverse Image Search

The Web Application for Video Fragmentation & Reverse Image Search was tested directly in the test cycles 1 to 4, 6 and 8. Further tests were done indirectly through the integration of the service into the Verification Plugin. In the earlier test cycles of the project, the testers pointed out that the application workflow was too complex and that the performance was low. The processing of the video files took too long to be used in a journalistic workflow. Another feedback concerned the number, quality and coverage of the extracted thumbnails. Later tests in the project gave feedback on the keyframe selection effectiveness and on the compatibility of the web application with the supported video platforms. The testers pointed out that Vimeo and DailyMotion took longer to process than the videos from YouTube and Twitter. Also, some Vimeo clips were not processed correctly. The collected feedback and the actions made by the developers of this technology resulted in a reliable, user-friendly and time-efficient tool for fragment-level reverse search of videos on the Web. In test cycle 8, video sub-shot fragmentation and keyframe extraction were evaluated comparatively against two alternative baseline approaches. The findings of this testing indicated the efficiency of the InVID video fragmentation and keyframe selection approach in producing concise and complete keyframe-based summaries of the video content, that effectively assist journalists when trying to assess the originality of a given video by searching for prior occurrences of it on the Web.

4.4.4 Testing of the Near Duplicate Detection Service

The Near Duplicate Detection Service (API) was tested directly in the test cycles 2 to 4, 6 and 8. Further tests were done indirectly through the integration of the service into the Verification Application. In the first cycles, the testers suggested improvements in the documentation and reported shortcomings in reliability (timeouts). Additional feedback covered improvements in functionality and error handling of the API. Tests in the later test cycles showed the improvements of the documentation and reliability, but suggested additional improvements in error handling and error messages. The tests in test cycle 8 showed the mature state of the interface and only minor bugs were reported. A low response time for a single function of the API was still pointed out.

4.4.5 Logo Detection Service

The Near Duplicate Detection Service was tested directly in test cycles 1 to 4. Further tests were done indirectly through the integration of the service into the Verification Application. In the first test cycles, the testers reported slow processing of the videos, errors in processing with different videos and suggested improvements in the user interface and the workflow. Logos that were not recognised were also reported. Later tests showed the improvements of

both the user interface and the API. Minor improvements were suggested for the user interface and for the documentation of the API in the later tests cycles.

4.4.6 Forensic Analysis Service

The Forensic Analysis Service was tested directly in test cycles 6 and 7. Further tests were done indirectly through the integration of the service into the Verification Application. The service was well documented and reliable. No major bugs were found in the test cycles. The main feedback focused on improvements in error handling.

4.4.7 Context Aggregation & Analysis Service

The Context Aggregation & Analysis Service was tested directly in the test cycles 1 to 4 as well as 6 to 8. Further tests were done indirectly through the integration of the service into the Verification Application and into the Verification Plugin. The user interface received a lot of positive feedback from the testers. This included positive feedback on the functionality and also that the results were appropriate and accurate enough for journalistic needs. The tool helped the testers in solving video verification problems. Various suggestions for improvements of the data displayed in the user interface were made by the testers. Regarding the API, the change of the processing call from synchronous to asynchronous was appreciated. Suggestions for improvements on the API level concerned missing error messages for different situations. In test cycle 4, the testers reported problems regarding the robustness of the application which were fixed and successfully retested in test cycle 6. At the API level, the service was assessed as very good and complete in test cycle 6. The testers assessed that this tool helps a journalist get more insights into a video based on the contextual data. The added value of this tool for a journalist was rated as "High".

4.4.8 Rights Management Service

The Rights Management Service was tested on user interface level in the test cycles 3, 4 as well as 6 to 9. On the API level, the tests were conducted in test cycles 1 to 4. Further tests were done indirectly through the integration of the service into the Verification Application and into the Verification Plugin. The results from the API tests were very good in all test cycles. The API was very well documented, reliable and no bugs were found for this API in any of the test cycles. The feedback for the user interface concerned improvements such as additional options in the reuse request, wording and adoption of the display of the different rights data. Problems with some Twitter and Facebook videos were pointed out and fixed in the following development cycles. The testers stated that this tool helps a journalist a lot to clear the rights situation about a newsworthy video. In test cycles 8 and 9, the user interface was in a very stable and usable state.

4.4.9 InVID Multimodal Analytics Dashboard and Tool for Social Media Retrieval and Topic Detection

The InVID Multimodal Analytics Dashboard and the integrated Tool for Social Media Retrieval and Topic Detection were tested by users of the consortium and users from the companies of the consortium with a journalistic background in test cycles 1 to 4. In test cycle 9, the Multimodal Analytics Dashboard was tested by users external to the consortium. The major feedback in test cycles 1 to 4 covered the presentation of the stories from the Tool for Social Media Retrieval and Topic Detection. This concerned both the visualisation and the textual representation of the stories. Further feedback pointed out that the InVID Multimodal Analytics Dashboard was a powerful but complex tool. Proper training for this tool would be necessary to successfully use the existing possibilities and understand all results. The tests with external users in test cycle 9 gave positive feedback on the different widgets of the InVID Dashboard. Another result of the testing with external user was that a comprehensive training workshop would be necessary to work confidently with the tool.

4.4.10 InVID Verification Plugin

The InVID Verification Plugin was tested by users of the consortium with a journalistic background in test cycles 3 to 9. In test cycles 4 to 9, users from the companies of the consortium and external users tested the Verification Plugin. The test results from testers of the consortium suggested improvements in the interface and reported minor bugs. The results provided by users from the companies of the consortium and external users indicated that both the InVID Verification Plugin and the integrated components were very well appreciated by its users. 90% of all survey responses assessed the Verification Plugin as “Very useful” or “Useful”, 72% of the users found it easy to use and 93% of the users would recommend the InVID Plugin to their newsroom or a colleague.

4.4.11 InVID Verification Application

The InVID Verification was tested by users of the consortium with a journalistic background in test cycle 1 and in test cycles 3 to 7. In test cycles 5, 8 and 9, the Verification Application was tested by users from the companies of the consortium and external users. Based on the responses from the testers of the consortium with a journalistic background in the early test cycles, the workflow and the interface of the Verification Application were redesigned. The new design was then verified with external testers and users from the companies of the consortium not connected with the project in test cycle 5. The responses suggested further improvements of the interface and the functionality. The internal testers reported different bugs and suggestions for improvements in all test cycles. The improved Verification Application was retested with external users in test cycles 8 and 9. The Verification Application and the integrated components were very well appreciated by the testers. 90% of the users assessed the Verification Application as “Very useful” or “Useful” and 95% of the users found it easy to use.

4.4.12 InVID Mobile Application

The InVID Mobile Application was tested inside of the consortium in test cycles 4 to 9. External testers from regional newspapers in Austria and emergency services in Austria tested the Mobile Applications in test cycles 5, 7 and 9. The tests inside of the consortium found a wide range of bugs on different smartphone devices and also suggested improvements for the registration process. The external users tested the Mobile Application in a real-world environment for journalists and emergency services. They assessed the InVID Mobile Application as “Good”, “Very good” and “Useful” and easy to use.

4.4.13 InVID Core Platform API

The InVID Core Platform API was tested by IT specialists of the companies from the consortium in test cycles 1 to 4. Major feedback for this API covered bugs in the management of documents and missing details in the documentation such as which functions are available. Temporary proxy problems were reported and suggestions for improvements in error handling and error reporting were made. The feedback from the test cycles was implemented in the following development cycles and the API reached a mature state in an early stage of the project.

5 Conclusions

In test cycles 7 to 9, the focus was to get all tools ready for external testers and testing with external users. This allowed to gather feedback from people that were not connected to the project, and therefore had an independent opinion on the usefulness and stability of the applications. The external testers were lead users from companies such as France24, BBC, Newsy.com, Sky News, Berkeley University, regional newspapers from Austria, and members of the First Draft network. Additional detailed tests on functional and non-functional features were made in test cycles 7 to 9 by users of the consortium with a journalistic background and by users from the companies of the consortium. Extended tests of the exposed APIs were done by IT specialists from the companies of the consortium.

The different InVID applications received very good feedback from the external testers. 90% of the users assessed the Verification Application as “Very useful” or “Useful” and 95% of the users found it easy to use. The InVID Mobile Application was rated as “Good”, “Very good” and “Useful” by the external testers and all external testers found it easy to use. 74% of the users assessed the Verification Plugin as “Very useful” or “Useful” in test cycles 7 to 9, and the same percentage of users found it easy to use. For the Dashboard, all assessed features received positive feedback from the majority of the external testers.

The API tests in the test cycles ensured a stable connection between the different components and services of the InVID platform. The APIs of the InVID components Video Fragmentation & Annotation Service, Near Duplicate Detection Service, Logo Detection Service, Forensic Analysis Service, Context Aggregation & Analysis Service, Rights Management Service and Core Platform API were assessed as reliable and responsive. No major bugs were found in the last version of the different API services.

The results of test cycles 7 to 9 have shown the success of the InVID project to develop tools for journalists that help them in the task of verification and rights management for user-generated videos and have also proven that the InVID tools are fully compatible with the market needs for effective video verification and rights clearance.

References

Hubmann-Haidvogel, A., Scharl, A. and Weichselbraun, A. (2009). “Multiple Coordinated Views for Searching and Navigating Web Content Repositories”, Information Sciences, 179(12): 1813-1821.

Matera, M., Rizzo, F. and Carughi, G.T. (2006). "Web Usability: Principles and Evaluation Methods", Web Engineering. Eds. E. Mendes and N. Mosley. Berlin: Springer. 143-180.