This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement n: 761999

EasyTV: Easing the access of Europeans with disabilities to converging media and content.

EasyTV self-learning system for improving personalisation capabilities

EasyTV Project

H2020. ICT-19-2017 Media and content convergence. – IA Innovation action.

Grant Agreement n°: 761999

Start date of project: 1 Oct. 2017

Duration: 30 months

Document. ref.:

Disclaimer

The EasyTV Consortium consists of the following partners:

	Partner Name	Short name	Country
1	Universidad Politécnica de Madrid	UPM	ES
2	Engineering Ingegneria Informatica S.P.A.	ENG	IT
3	Centre for Research and Technology Hellas/Information Technologies Institute	CERTH	GR
4	Mediavoice SRL	MV	IT
5	Universitat Autònoma Barcelona	UAB	ES
6	Corporació Catalana de Mitjans Audiovisuals SA	CCMA	ES
7	ARX.NET SA	ARX	GR
8	Fundación Confederación Nacional Sordos España para la supresión de barreras de comunicación	FCNSE	ES
9	Unione Italiana dei ciechi e degli ipovedenti	UICI	IT

PROGRAMME NAME:	**H2020. ICT-19-2017 Media and Content Convergence – IA Innovation Action**
PROJECT NUMBER:	761999
PROJECT TITLE:	EASYTV
RESPONSIBLE UNIT:	CERTH
INVOLVED UNITS:	UPM, CCMA
DOCUMENT NUMBER:	D4.5
DOCUMENT TITLE:	EasyTV user interface adaptation framework
WORK-PACKAGE:	WP4
DELIVERABLE TYPE:	Prototype
CONTRACTUAL DATE OF DELIVERY:	30-1-2020
LAST UPDATE:	31-1-2020
DISTRIBUTION LEVEL:	PU

Distribution level:

PU = *Public,*

RE = *Restricted to a group of the specified Consortium,*

PP = *Restricted to other program participants (including Commission Services),*

CO = *Confidential, only for members of the LASIE Consortium (including the Commission Services)*

Document History

VERSION	DATE	STATUS	AUTHORS, REVIEWER	DESCRIPTION
v.0.1	9-12-2019	Draft	Salim Gannoum (CERTH/ITI) Nikolaos Kaklanis (CERTH/ITI) Dimosthenis Elmas (CERTH/ITI) Georgios Gerovasilis (CERTH/ITI) Konstantinos Votis (CERTH/ITI) Dimitrios Tzovaras (CERTH/ITI)	Table of Contents definition and document structure
v.0.2	20-12-2019	Draft	Salim Gannoum (CERTH/ITI) Nikolaos Kaklanis (CERTH/ITI) Dimosthenis Elmas (CERTH/ITI) Georgios Gerovasilis (CERTH/ITI) Konstantinos Votis (CERTH/ITI)	Chapter 3
v.0.3	08-01-2020	Draft	Salim Gannoum (CERTH/ITI) Nikolaos Kaklanis (CERTH/ITI) Dimosthenis Elmas (CERTH/ITI) Georgios Gerovasilis (CERTH/ITI) Konstantinos Votis (CERTH/ITI)	Chapter 4
v.0.4	13-01-2020	Draft	Salim Gannoum (CERTH/ITI) Nikolaos Kaklanis (CERTH/ITI) Dimosthenis Elmas (CERTH/ITI) Georgios Gerovasilis (CERTH/ITI) Konstantinos Votis (CERTH/ITI)	Chapter 5 & 6
v.0.5	16-01-2020	Draft	Josep Maria Carmona (CCMA) Francesc Mas (CCMA)	Added Chapter 6

			Jordi Fabregat (CCMA)	
V0.6	21-01-2020	Peer-review/revision	Salim Gannoum (CERTH/ITI) Nikolaos Kaklanis (CERTH/ITI) Dimosthenis Elmas (CERTH/ITI) Georgios Gerovasilis (CERTH/ITI) Konstantinos Votis (CERTH/ITI) Dimitrios Tzovaras (CERTH/ITI)	Complete version to be sent for peer review
V0.7	30-01-2020	Final ready for submission	Salim Gannoum (CERTH/ITI) Nikolaos Kaklanis (CERTH/ITI) Dimosthenis Elmas (CERTH/ITI) Georgios Gerovasilis (CERTH/ITI) Konstantinos Votis (CERTH/ITI) Dimitrios Tzovaras (CERTH/ITI)	Final version

Definitions, Acronyms and Abbreviations

ACRONYMS / ABBREVIATIONS	DESCRIPTION
RBMM	Ruled based matchmaker
STMM	Statistical matchmaker
JSON	JavaScript Object Notation
DB	Data Base
SR	Screen Reader
TTS	Text To Speech
HBBTV	Hybrid Broadcast Broadband TV
ARIA	Accessible Rich Internet Applications
W3C	World Wide Web Consortium
UI	User Interface
HTML	HyperText Markup language

Table of Contents

List of Figures

List of Tables

Executive Summary

This document corresponds to the deliverable D4.5 "*EasyTV self-learning system for improving personalisation capabilities*" that is related to WP4. It describes the work done in T4.3 "*Personalised services for people with disabilities, including self-adaptive and tailored services, which can learn from users' actions to improve the accuracy of the personalisation*". The deliverable describes how the personalisation framework improves the personalisation accuracy by taking into account user's history of interaction and their profiles. The document consists of following chapters:

Chapter 1 An overview of the refinements process and used data.
Chapter 2 An architectural overview of hyper-personalisation framework and its engaged components.
Chapter 3 Addresses profiles refinement process, and how users' history of interaction is mined for reoccurring patterns.
Chapter 4 Describes rules refinement process, and how association analysis is used for rule refinement.
Chapter 5 Presents two indicative cases of the workings of the user profile refinement and rule refinement processes.
Chapter 6 Presents HbbTV accessibility services, namely HbbTV screen-reader and HbbTV subtitles functionality and intermediate tests results.
Chapter 7 Conclusions and future work.
Chapter 8 List of references

1. INTRODUCTION

The personalisation process is the process of suggesting and adapting the user settings and configuration to fulfil the user needs. At the heart of this process is the inference of suggestions that meet the user preferences. The hyper-personalization framework is the component that implements the personalization. Our implementation follows a hybrid matchmaking approach that dynamically combines the suggestions inferred by two matchmaking approaches, a knowledge approach and a collaborative one. The collaborative approach is based on the fact that similar users have similar preferences The two major factors that play a role in the quality of the inference suggestions are, the similarity function and the quality and amount of available data. The knowledge base approach uses knowledge and rules to infer suggestions. From technical point of view, the knowledge based approach is implemented using semantic web technologies and the collaborative one using statistical analysis method.

To improve the accuracy of the personalization process we mainly focus on improving the input of the collaboration approach and the semantic rules of the knowledge-based one. Two major factors mainly affect the statistical analysis: a) the similarity function used and b) the quality of the data in terms of how much representative it is. A good similarity function is enough to guarantee good matching results. Where, improving the data quality is expected to positively affect the statistical matchmaking suggestions. Similarly, knowledge-based suggestions are mainly based semantic rules, thus, making the rules more accurate will lead to better suggestions.

In addition, two new accessibility services have been developed to include personalization features of HbbTV. The first is the HbbTV screen reader solution to enhance the access to blind and low vision users. The second is the HbbTV subtitles service, where a new color palette for color subtitles was proposed. These services will enrich the personalization process with more data and allow it to infer suggestions for HbbTV settings.

This deliverable describes the processes of improving the accuracy of both matchmaking approaches, namely the collaborative and knowledge-based one, in addition to the presentation of the two new accessibility services that expand the personalization features into HbbTV. We start first by providing an architectural overview of the engaged components, their functionalities and communications. The next section describes the profiles refinement process and how user history of interaction is stored and mined for identifying reoccurring patterns of interactions and finally how these patterns are converted and proposed as profile modification suggestions. Next is the description of the rule refinement section, which explains how association analysis is used to extract association rules, which consequently is used to refine knowledge based rules. A set of explanatory use cases are presented in indicative case section that explain with paradigm the workings of both refinement approaches. The HbbTV screen reader and subtitles services are addressed in section HbbTV accessibility services. The document ends with a conclusion and future work section.

2. SYSTEM ARCHITECTURE

2.1. Overview

Both refinement processes, namely profile refinement and RBMM sematic rules refinement, take place in the statistical matchmaker. The statistical matchmaker is a component of the hyper-personalization framework that stores users' data and implements the statistical analysis approach. Figure 1 shows an overview of the statistical matchmaker components and its place within the hyper-personalization framework. The relevant components for the personalization accuracy refinement process are:

1. **User profile repository**: A relational database that contains all available users' profiles. Profiles are stored in JSON format in the database and the STMM has direct access to this data.
2. **User history of interaction repository:** A relational database that contains a time line of the user interaction data with the platform. Although this repository is depicted as separated one, interaction data are stored in the user profile repository DB.
3. **EasyTV ontology:** A repository that contains the set of rules supported by the rule-based matchmaker (RBMM), where the STMM has direct access for the purpose of the automatic rule refinement process.
4. **RBMM refiner:** This component performs the rules refinement process. Its main functionality is to extract association rules by performing association analysis of user profiles and use these rules to refine semantic rules of the RBMM.
5. **User model refiner**: This component performs the profile refinement process. Its main functionality is to extract interaction patterns from the user history of interaction and then suggest profile modifications to the user. Suggestions accepted by the user are then reflected in the corresponding user profile.

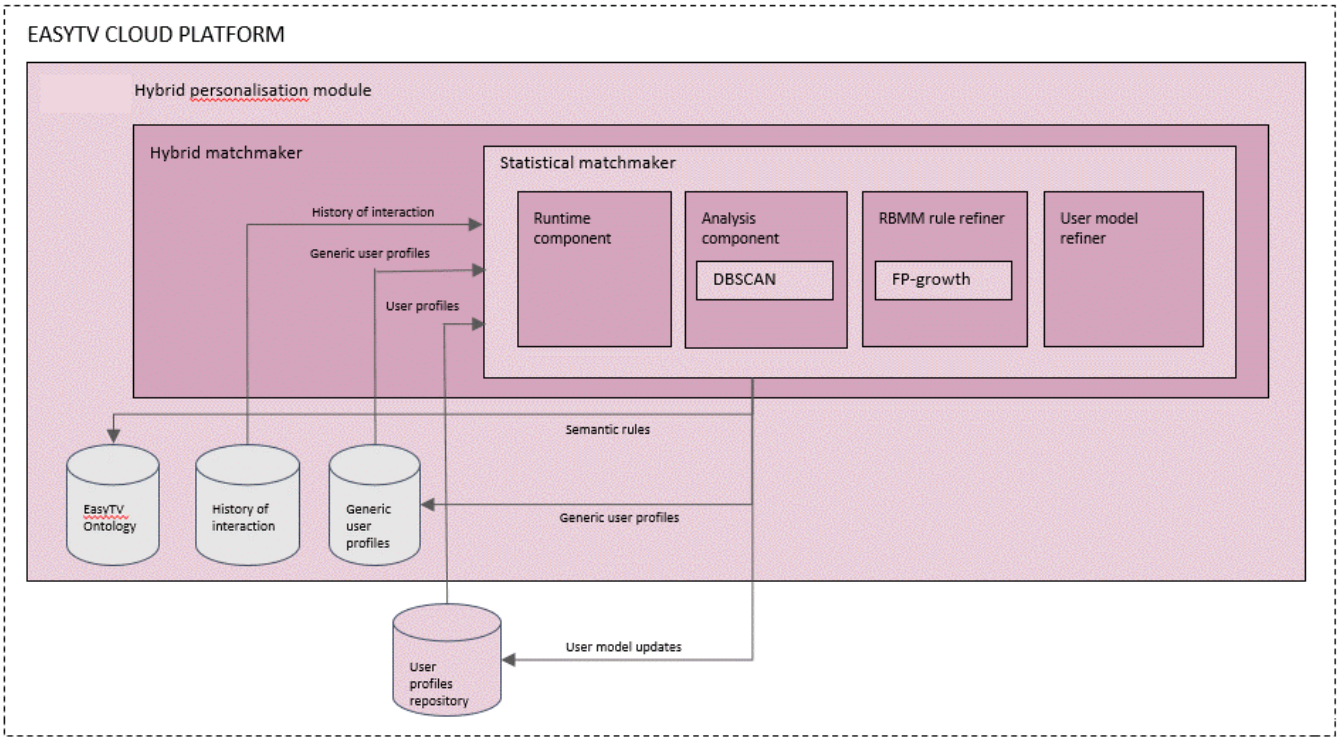

Figure 1 Statistical matchmaker architectural overview

3. PROFILES REFINEMENT

3.1. Overview

Profiles refinement is the process of extracting interaction patterns from the user history of interaction, which are converted into suggestions for profile refinement. The profile refinement includes adding a new preferences or modifying the value of existing one. The following diagram (Figure 2) illustrates the refinement process flow. The user interaction with the device produces interaction events, which are stored in the users' interaction history. On regular bases (i.e. once a day), the repository is analyzed for reoccurring actions and presented to the user for confirmation. Confirmed refinements are saved in the user profile. The process aims at gradually improving profiles quality by making them more representative of users' needs, improving in this way the quality of the statistical analysis suggestions..

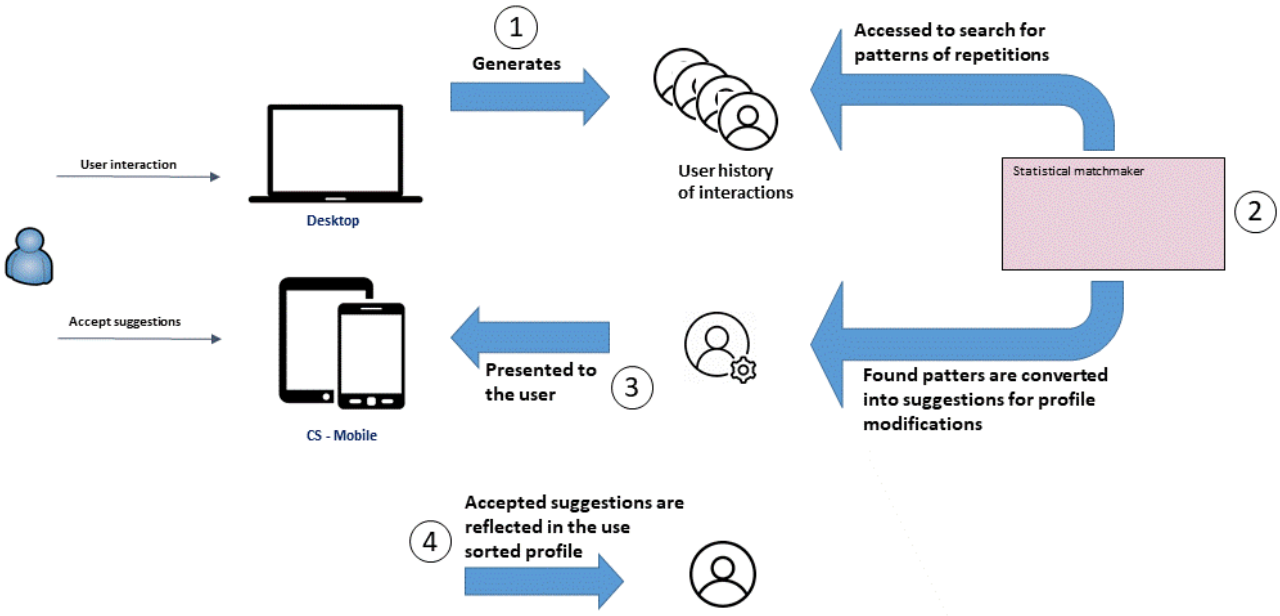

Figure 2 Profiles refinements process

3.2. User interaction tracking

The user interacts with EasyTV platform through the CSapp, which is an application that can be installed on mobile phone or tablet. The user through this application can control the platform settings Table 1. Each user of the EasyTV platform declares his/her preferences in a user profile. To enable personalization the user is requested to login to the platform. After login, a profile is activated and all user interactions are related to this activated profile. Figure 3 shows a user profile example.


```
{
    "user_preferences": {
        "default": {
            "preferences": {
                "http://registry.easytv.eu/gazeControl": false,
                "http://registry.easytv.eu/zoomTextField": false,
                "http://registry.easytv.eu/application/cs/eq/bass": 0,
                "http://registry.easytv.eu/application/cs/eq/mids": 0,
                "http://registry.easytv.eu/touchVibration": false,
                "http://registry.easytv.eu/application/cs/eq/highs": 0,
                "http://registry.easytv.eu/common/contrast": 100,
```

```
                "http://registry.easytv.eu/application/voiceControl": true,
                "http://registry.easytv.eu/application/cs/GUILanguage": "en",
                "http://registry.easytv.eu/application/cs/audioVolume": 50,
                "http://registry.easytv.eu/application/cs/audioLanguage": "en",
                "http://registry.easytv.eu/application/cs/textDetection": false,
                "http://registry.easytv.eu/application/cs/audioDescription": false,
                "http://registry.easytv.eu/application/cs/csSoundDetection": false,
                "http://registry.easytv.eu/application/cs/characterRecognition": false,
                "http://registry.easytv.eu/common/content/audio/volume": 90,
                "http://registry.easytv.eu/application/cs/audioAssistanceBasedOnTTS": false,
                "http://registry.easytv.eu/common/display/screen/enhancement/font/size": 12,
                "http://registry.easytv.eu/common/display/screen/enhancement/font/type": "Roboto",
                "http://registry.easytv.eu/common/display/screen/enhancement/background": "#18d4dc",
                "http://registry.easytv.eu/common/display/screen/enhancement/font/color": "#b23b41"
            }
        }
    }
}
```


Figure 3 User profile in JSON format

Table 1 CS app application specific preferences

Preference	Type	Value range	Short description
Text-to-speech rate	Integer	range (-10…+10) (default 0)	Speech speed
Text-to-speech volume	Integer	0 - 100 (default 90)	Speech volume level
Text-to-speech language	String	[“en”,“ca”, “it”, “es”]	Language
Text-to-speech voice	String	-	The voice of the speaker
Text-to-speech quality	Integer	range (1…8)	1 (best quality, more storage) to 8 (worst quality, less storage).
Touch vibration	Boolean	[True, false]	Touch vibration
Text size	Integer	range (1…3)	UI text size
Cursor size	Integer	range (1…3)	Cursor size
Cursor color	String	range (#000000…#FFFFFF)	Cursor font color
Zoom text	Boolean	[True, false]	Zoom text enable/disabled

Audio assistance	Boolean	[True, false]	Audio assistance based on TTS technology
Language	String	[“en”,“ca”, “it”, “el”, “es”]	The language of the GUI interface
Image magnification scale	Numeric	[1.5 - 3.5]	Image magnification scale
Text detection	Boolean	[True, false]	Text detection service
Face detection	Boolean	[True, false]	Face detection service
Audio description	Boolean	[True, false]	Audio description service
Sound detection	Boolean	[True, false]	Sound detection service
Character recognition	Boolean	[True, false]	Character recognition service
Audio equalizer bass	String	[-15, 15]	Equalizer bass band
Audio equalizer mid-range	String	[-15, 15]	Equalizer mid band range
Audio equalizer high end	String	[-15, 15]	Equalizer high band range
Audio volume	Integer	0 – 100	Audio volume of audio subtitle
Audio language	String	[“en”,“ca”, “it”, “el”, “es”]	The language of the description
Audio subtitles	Boolean	[True, false]	Audio subtitles service
Subtitle language	String	[“en”,“ca”, “it”, “el”, “es”]	Subtitle language
Subtitle font size	Integer	[1 – 50]	Subtitles font size
Subtitle font color	String	range (#000000…#FFFFFF)	Subtitle font color
Subtitle background color	String	range (#000000…#FFFFFF)	Subtitle background color

Gesture control	Boolean	[True, false]	Gesture control of the application
Voice control	Boolean	[True, false]	Use voice command to control the application
Gesture control	Boolean	[True, false]	User hand gesture to control of the application
Sign avatar	Boolean	[True, false]	Sign avatar
Sign avatar language	String	[“ca”, “it”, “el”, “es”]	Sign avatar language
Read aloud text detection	Boolean	[True, false]	Enable reading text detection results
HBBTV screen reader	Boolean	[True, false]	Enable HBBTV screen reader
HBBTV reader speed	String	“slow”, "normal", "fast"	HBBTV screen reader speed
HBBTV subtitles	Boolean	[True, false]	Enable HBBTV on screen subtitles
HBBTV subtitles position	String	up,"down"	HBBTV subtitles position on screen
HBBTV subtitles language	String	ca,"en","ar", "es"	HBBTV subtitles language
HBBTV subtitles font size	Integer	[1, 50]	HBBTV subtitles font size
HBBTV subtitles background	Boolean	[True, false]	Enable HBBTV subtitles background

After activating a profile, each change made by the user to any application setting, it generates an event. The event is in JSON format (Figure 4) and contains the action made by the user in addition to any change in the corresponding context (Table 2).

```
{
    "user_action": {
        "http://registry.easytv.eu /application/cs/ui/zoomTextField": true
    },
    "user_context": {
        "http://registry.easytv.eu/context/device": "tablet",
        "http://registry.easytv.eu/context/light": 10,
        "http://registry.easytv.eu/context/proximity": 20,
```

```
        "http://registry.easytv.eu/context/location": "es",
        "http://registry.easytv.eu/context/time": "09:47:00"
    }
}
```

Figure 4 User interaction event

Table 2 Contextual information

URL	Data Type	Range	Short description
Device	String	[“Mobile”, “tablet”, “PC”]	Device type, one of mobile, tablet or PC
Brightness	Numeric	range (0.1…*)	The ambient light level (illumination) in lx
Proximity	Integer	range (0…100)	Measures the proximity of an object in cm relative to the view screen of a device
Location	String	-	The device's location, country code
Time	String	-	Current time in the form HH:MM:SS

3.3. Mining user history of interaction

Mining user history of interactions is the process of mining reoccurring modification actions. The user history of interaction repository is a sequence of user modifications over time over the user profile initial values in relation to contextual information. Figure 5 depicts the modifications actions of a user in relation to its initial profile, a dot over preference axes illustrates a modification event.

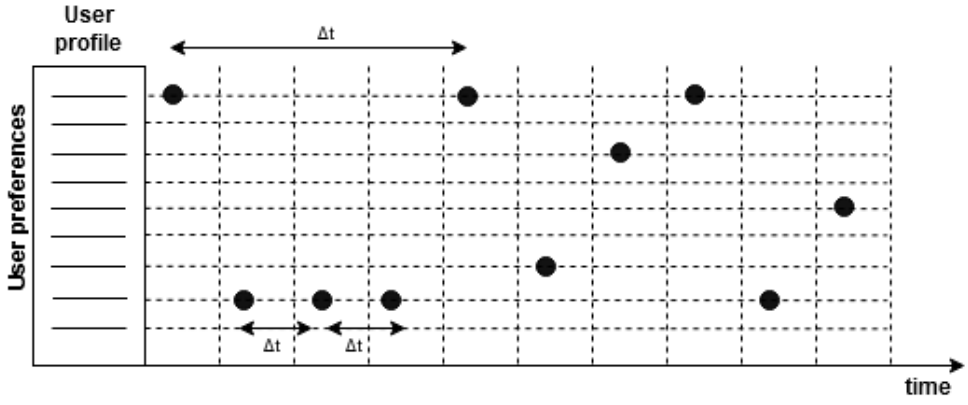

Figure 5 User history of interactions

These interaction events are stored in the history of interaction database in Figure 4 format and processed on regular bases to extract reoccurred pattern of interactions. Figure 6 illustrates the process steps. The whole process mines these patterns by sorting the events in descended order relevant to their occurrence time and then locating any action sequence that occurs N or more times.

The number of occurrences of an action in the user history is an indication of the significance of the action. The more an action occurs, the more it is considered proper to be included in the user profile. A proper value for N is four occurrences, one and two occurrences are too low and three is not large enough.

The next step is to merge all events, which results in a merged event that contains each preference assigned values over time. Identifying repeated pattern of interaction is done by looking over each preference for N repeated values. The process ends by presenting these modification suggestions to the user through a drop down list in the CSapp and only those suggestions accepted by him/her are reflected in the user profile. Assuming the list of events shown in the following table, then the merging step will produce the file in Table 4 .

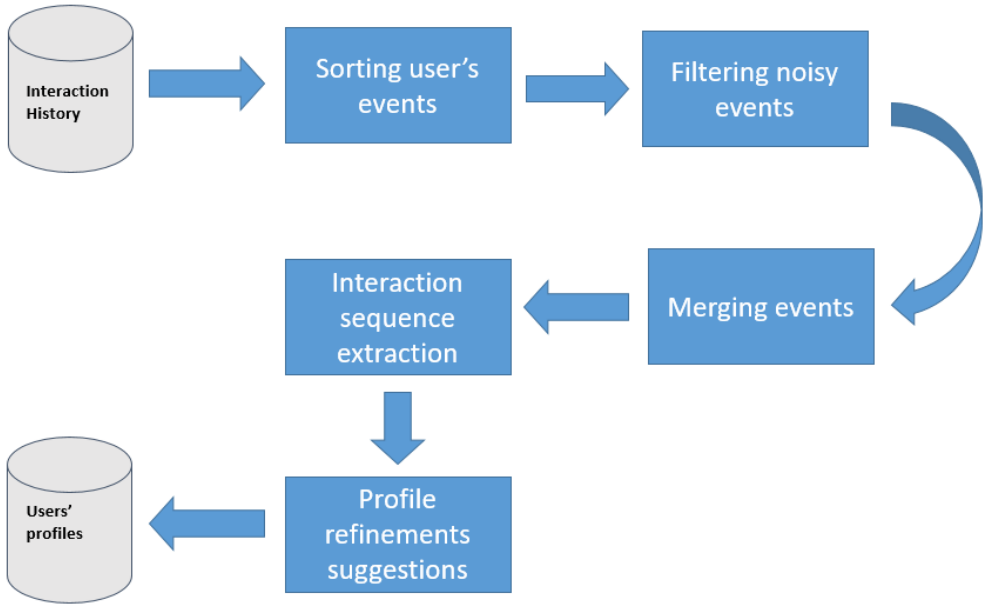

Figure 6 Profile refinements pipeline

Table 3 Interaction events


```
{{
    "user_action": {
         "http://registry.easytv.eu/application/cs/ui/zoomTextField": true
    },
    "user_context": {
         "http://registry.easytv.eu/context/device": "tablet",
         "http://registry.easytv.eu/context/light": 10,
         "http://registry.easytv.eu/context/proximity": 20,
         "http://registry.easytv.eu/context/location": "es",
         "http://registry.easytv.eu/context/time": "09:00:00"
    }
},
{
    "user_action": {
        "http://registry.easytv.eu/application/cs/accessibility/detection/sound": true
    },
    "user_context": {
         "http://registry.easytv.eu/context/device": "tablet",
         "http://registry.easytv.eu/context/light": 10,
         "http://registry.easytv.eu/context/proximity": 20,
         "http://registry.easytv.eu/context/location": "es",
         "http://registry.easytv.eu/context/time": "11:00:00"
    }
},{    "user_action": {
        "http://registry.easytv.eu/application/cs/ui/zoomTextField": true
    },
```

```
    "user_context": {
        "http://registry.easytv.eu/context/device": "tablet",
        "http://registry.easytv.eu/context/light": 10,
        "http://registry.easytv.eu/context/proximity": 20,
        "http://registry.easytv.eu/context/location": "es",
        "http://registry.easytv.eu/context/time": "20:00:00"
    }
}}
```

Table 4 Result of merging step

```
{
    "user_action": {
        "http://registry.easytv.eu/application/cs/ui/zoomTextField": [true, null, true]
        "http://registry.easytv.eu/application/cs/accessibility/detection/sound": [null, true, null]
    },
    "user_context": {
         "http://registry.easytv.eu/context/device": ["tablet","tablet","tablet"]
         "http://registry.easytv.eu/context/light": [10, 10, 10]
         "http://registry.easytv.eu/context/proximity":  [20, 20, 20]
         "http://registry.easytv.eu/context/location": ["es","es","es"]
         "http://registry.easytv.eu/context/time": ["09:00:00” ,"11:00:00","20:00:00"]
    }

}
```

3.4. Profile’s refinement

After extracting refinement suggestions from user history of interaction (section 3.3), these suggestions are presented to the user in the form of a drop list that the user can select from which suggestions to accept. The accepted suggestions are then reflected in the corresponding user profile.

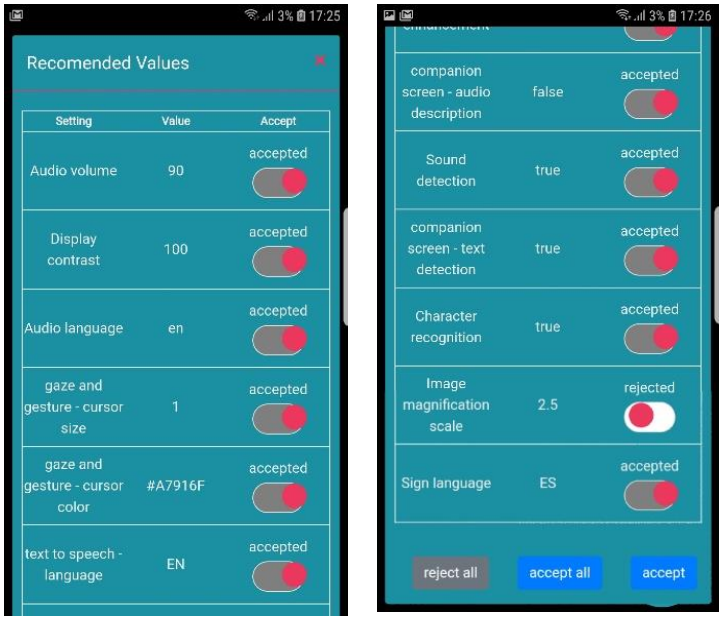

Figure 7 Profile refinements suggestions

4. RULES REFINEMENT

4.1. Overview

Rules refinement is the process where user profiles are analyzed in order to automatically refine the semantic rules supported by the RBMM. The rules can be refined by removing/updating existing rules or by adding new rules. The input of the process is the set of user profiles stored in the user profile repository. The output of the process is a set of suggestions rules that associate system settings and contextual information with other system settings. Lastly, these extracted rules are used to remove, update or extend the rules supported by the RBMM. Figure 8 depicts this flow.

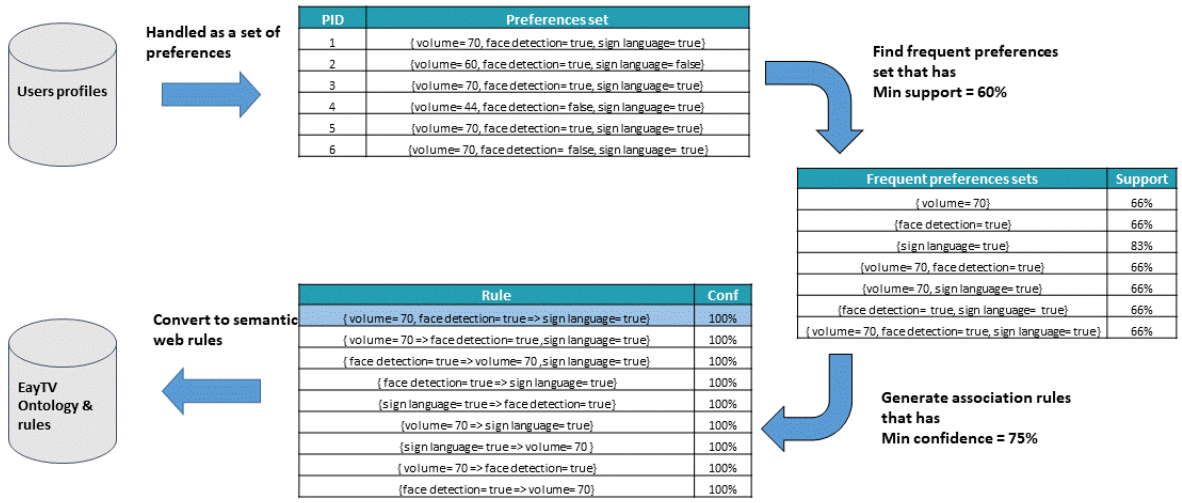

Figure 8 Rules refinements flow

We use data mining techniques to mine users' profiles for preferences patterns. Following the data mining terminology, the set of user profiles stored in the repository is our data set. The data set consists of data objects that represent entities of the domain problem and are described by a set of attributes. The process flow starts by association analysis, which mines for frequent patterns (in our case frequent patterns of user preferences). These frequent patterns are used for mining association rules [1]. The final step is to use these association rules for refinement of the rules supported by the RBMM.

4.2. Association analysis of user profiles

Association analysis [2] is a data mining method that extracts frequently patterns from a data set. A frequent pattern is a set of items that appear frequently, or in other words, items that are highly correlated. In the context of our implementation, our data set is the repository of user profiles and an itemset is a set of preferences and their values. For example, a set of items, such as *face detection=enabled* and *text detection=enabled* that appear frequently in user profiles together is a *frequent itemset,* which means that these values are correlated or associated.

Two metrics are used to extract association rules: support and confidence [3]. Support, is an indication of how frequently an item appear in a dataset, the support of A with respect to database entries T is the number of entries in T that contains A. In other words, this metric measures the probability of A. In the content of our implementation, A is a preference item set A = $\{P_1, P_2, P_3 \ldots \ P_n\}$ where P is a preference with its value.

$$support(A) = P(A) = \frac{|\{t \in T;\ \ A \in t\}|}{|T|}$$

T: Transactions set, A: Itemsets

For rule mining, in addition to support, confidence is measured. Confidence is an indication of how often a rule is found to be true. These metrics measure the importance of a rule.

$$support(A \rightarrow B) = P(A \cup B) = \frac{|\{t \in T;\ \ A \cup B \in t\}|}{|T|}$$

$$confidence(A \rightarrow B) = P(B|A) = \frac{support(A \cup B)}{support(A)}$$

T: Transaction set, A, B: Itemsets

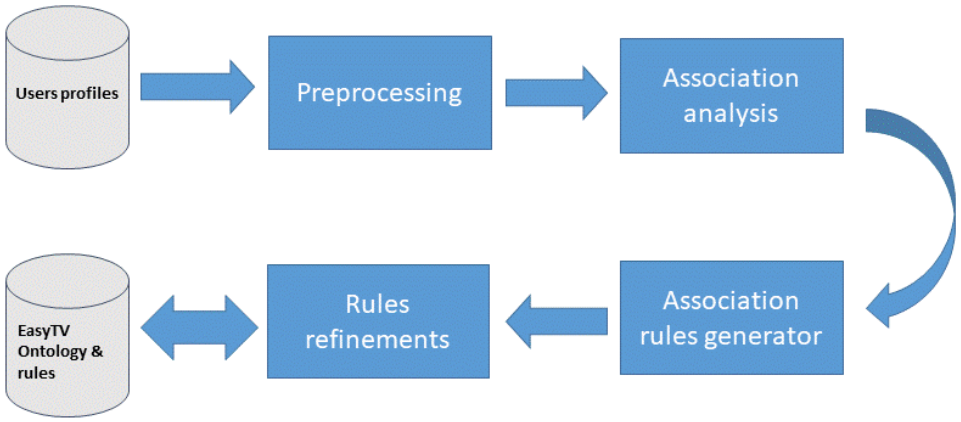

Figure 9 Rule refinements flow

The following sections address the different technical issues of association mining process, such as preferences discretization, preferences transformation and rule filtering and refining.

4.2.1. Data preprocessing

As described in the sections above, the set of user profiles included in the repository is our data set. A user profile consists of a set of preferences and their values. A preference is an attribute that is characterized by its data type and range of values. Data type can be one of String, Integer, Double or Boolean. Data range can be one of the following types:

1. **Nominal**: The values of a nominal attribute are symbols or *names of things*. Each value represents some kind of category, code, or state, and so nominal attributes are also referred to as categorical**.** The values do not have any meaningful order
2. **Binary:** Binary attributes are referred to as Boolean if the two states correspond to *true* and *false*. Binary attributes can be of two kind:
    a. **Symmetric Binary:** if both of its states are equally valuable and carry the same weight; that is, there is no preference on which outcome should be coded as 0 or 1
    b. **Asymmetric Binary:** if the outcomes of the states are not equally important, such as the positive and negative outcomes of a medical test for HIV.
3. **Ordinal:** An ordinal attribute is an attribute with possible values that have a meaningful order or ranking among them, but the magnitude between successive values is not known.
4. **Numeric**: The values of numeric attribute represent a quantitative values, it can be integer or real values.

4.2.1.1 Numeric Preference discretization

All attribute types can be handled properly by the association analysis method except for the numeric attributes. A numeric attribute is a dimension that has a continuous values in a specific range. When the range is long enough it will contain many values. That affects negatively the probability of finding associations, especially when adjacent values are not considered to differ significantly. For instance, a volume preference has a range of [0, 100] with 100 discrete values.

To increase processing time and the probability of finding meaningful patterns, we discretize numeric preferences that meet the following two conditions: a) to have long value range b) and their adjacent value does not differ too much and thus can be grouped without significantly affecting the results. For instance, changing volume size by one degree does not have a huge impact on the volume level. In this case we can group every two values such as [0, 1], [2, 3] … together. Discretization or binning is a clustering method where neighboring values are grouped and represented by the group value. The group value can be the average value of the group elements or the middle value of the group. For example, volume bin that has the values 10, 12, 13 are represented by the value 12. The effect of discretization is to reduce the range values into that of bin's values. Rather than having 100 values in the range [0, 100], after discretization into 50 bins each with two values, we have only 50 values. That equals to reducing the available values into half. Following are the formulas for calculating bin size, reduction factor and grouping factor:

$$Bin\ \ size = \frac{values\ \ number}{bin\ \ numbers}$$

$$Grouping\ \ factor = \frac{Bin\ \ size}{values\ \ number} \qquad Reduction\ \ factor = \frac{bin\ \ numbers}{values\ \ number}$$

Bin size is calculated by dividing the values number with bin numbers. As a division formula, we can specify two cases related to the relation between the divider and the divide:

$$\frac{values\ \ number}{bin\ \ numbers} = (Bin\ \ size\ \ \times\ \ bin\ \ numbers) + r$$

a) **r = 0, perfect division**: bin numbers can preference divide values number into equal parts, which means an even distribution of values between all bins.
b) **r != 0, remainder**: Bin numbers can't equally divides values number, which means an uneven distribution between bins. Because (r != 0) and (r < bin size) some bins will have one additional element, starting from the first bin.

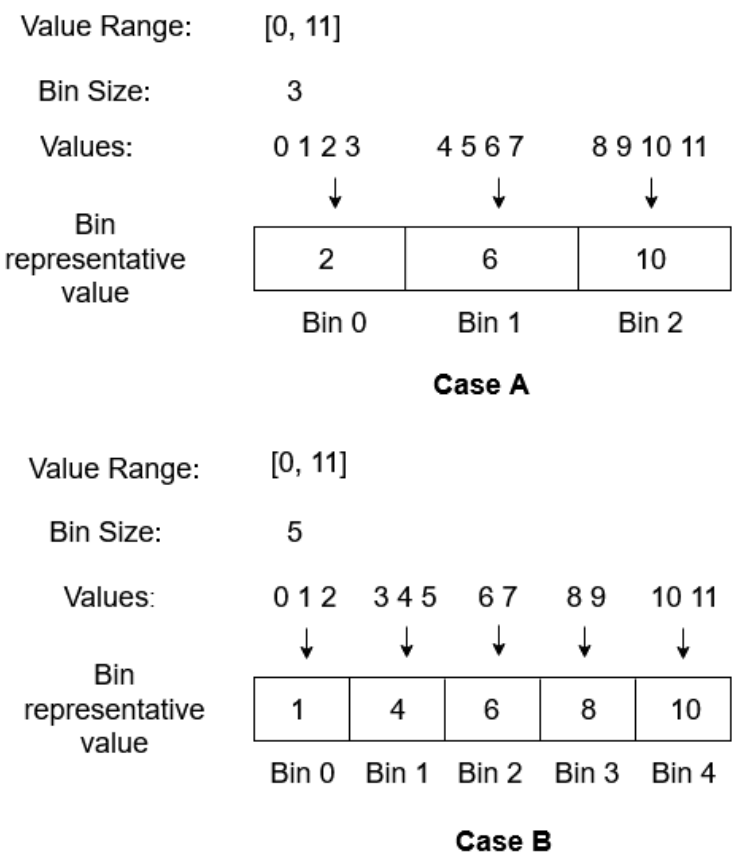

Figure 10 Binning cases

Figure 10 illustrates the two binning cases, in the first case (A) the requested bin size can be fulfilled with even distributed number of elements in all three bins, which is four elements. On the other hand, the second case where twelve numbers (values number) that we need to distribute on five bins cannot be done in an even matter. The results of diving the $\frac{12}{5} = 2 \times 5 + 2$ so this remaining value (r = 2) will be distributed evenly on bins starting from the first one. That means that two bins will have one element additionally (bins 0 and 1 have three elements). Of course, when it is possible we preferred to follow the first case by choosing bin numbers that can divide the values numbers. Following are the program output of different preferences types. The output includes statistical information, discretization or binning information and values histogram.

http://registry.easytv.eu/application/cs/audio/eq/highs

Attribute properties	
Range	Missing Value
[-15.0 , 15.0]	-16.0

Bins histogram										
Id	0	1	2	3	4	5	6	7	8	9
Range	-15, -12	-11, -9	-8, -6	-5, -3	-2, 0	1, 3	4, 6	7, 9	10, 12	13, 15
Center	-14	-10	-7	-4	-1	2	5	8	11	14
Counts	132	108	108	93	106	103	95	93	96	66

Statistical data					
Total	sum	Stand dev	Mean	Min	Max
1000	-737.0	8.6	-0.7	-15.0	14.0

Discretization properties			
Bins number	Bin Size	Remaining	Step
10	3	1	1.0

Values histogram	
Value	Frequency
2.0	34
-2.0	29
0.0	35
-3.0	33
3.0	39
4.0	36
-4.0	25
-5.0	35
5.0	29
-6.0	37
6.0	30
-7.0	33
7.0	36
8.0	26
-8.0	38
9.0	31
-9.0	34
-10.0	42
10.0	31
-11.0	32
11.0	34
12.0	31
-12.0	32
13.0	35
-13.0	27
-14.0	34
14.0	31

Figure 11 Numeric preference with 10 bins

http://registry.easytv.eu/application/cs/audio/eq/highs

Attribute properties	
Range	Missing Value
[-15.0 , 15.0]	-16.0

Bins histogram

Id	0	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
Range	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Center	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Counts	39	34	27	32	32	42	34	38	33	37	35	25	33	29	42	35	30	34	39	36	29	30	36	26	31	31	34	31	35	31	0

Statistical data

Total	sum	Stand dev	Mean	Min	Max
1000	-737.0	8.6	-0.7	-15.0	14.0

Discretization properties

Bins number	Bin Size	Remaining	Step
31	1	0	1.0

Values histogram

Value	Frequency
2.0	34
-2.0	29
0.0	35
-3.0	33
3.0	39
4.0	36
-4.0	25
-5.0	35
5.0	29
-6.0	37
6.0	30
-7.0	33
7.0	36
8.0	26
-8.0	38
9.0	31
-9.0	34
-10.0	42
10.0	31
-11.0	32
11.0	34
12.0	31
-12.0	32
13.0	36
-13.0	27
-14.0	34
14.0	31
-15.0	39
-1.0	42
1.0	30

Figure 12 Numeric preference with no bins

http://registry.easytv.eu/application/cs/cc/subtitles/language

Attribute properties	
Range	Missing Value
[0.0 , 4.0]	-1.0

Bins histogram					
Id	0	1	2	3	4
Range	EN	ES	CA	GR	IT
Center	EN	ES	CA	GR	IT
Counts	290	331	344	299	330

Figure 13 Nominal preference information

http://registry.easytv.eu/application/control/voice

Attribute properties	
Range	Missing Value
[0.0 , 1.0]	-1.0

Bins histogram		
Id	0	1
Range	false	true
Center	false	true
Counts	3209	1

Figure 14 Boolean preference information

http://registry.easytv.eu/application/cs/accessibility/magnification/scale

Attribute properties	
Range	Missing Value
[1.5 , 3.5]	0.0

Bins histogram					
Id	0	1	2	3	4
Range	1.5, 1.5	2.0, 2.0	2.5, 2.5	3.0, 3.0	3.5, 3.5
Center	1.5	2.0	2.5	3.0	3.5
Counts	798	0	796	0	0

Statistical data					
Total	sum	Stand dev	Mean	Min	Max
1594	3187.0	0.5	2.0	1.5	2.5

Discretization properties			
Bins number	Bin Size	Remaining	Step
5	1	0	0.5

Values histogram	
Value	Frequency
2.5	796
1.5	798

Figure 15 Numeric preference with 0.5 step

Figure 11 is the program output of numeric preference dimension with 10 bins, namely audio equalization highs with preference url `http://registry.easytv.eu/application/cs/audio/eq/highs`, where it shows statistical information and discretization information. The first table *Attribute properties* lists the dimension *range* [-15, 15] and that -16 indicates a *missing value*. Discretization information is shown in *discretization properties* table. The dimension has been discretized into 10 bins, where each bin has size of three except for the first one that has size four (due to the remaining value of one). The *step* value indicates the difference between any two values, which is equals to one. In addition, Bin's information is shown in *bins histogram* table, where the first bin has value range of [-15, -12] and the last of [13, 15]. The bin representative value of the first is -14 and 14 of the last one. The counts row list how many values of this bin has been measured. The total number of all profiles, the dimension's sum value, their standard deviation value and their mean are in *statistical data* table. The values of Min and Max are for the minimum and maximum located values. The last table *values histogram* shows the frequency of each value. On the other hand, Figure 12 shows the same preference information with no binning (each bin holds one number only), in this case the frequencies number shown in *values histogram* and *bin histogram* are identical. Figure 13 shows the program output of nominal preference, namely `http://registry.easytv.eu/application/cs/cc/subtitles/language`. The preference is for subtitle language and can be one of the five listed values (EN, ES, CA, GR, IT), thus the integer range of the attribute is [0, 4] and -1 indicates missing value. Similarly, Figure 14 shows the program output of boolean preference, namely http://registry.easytv.eu/application/control/voice. The preference is to enable for control and can be (true, false), thus the integer range of the attribute is [0, 1] and -1 indicates missing value. Lastly, Figure 15 illustrates a numeric preference where values step is 0.5 for a range of values [1.0, 2.0] we have three bins each representing only one value.

4.2.1.2 Preference integer coding

Towards further reducing processing time and having a unified processing method, we code all preference values as integer representations. Processing integer values is more efficient than processing other data types (e.g. string), as we need only one operation to compare two integers, where the number of comparison needed for string values is equal to the string length. As mentioned in the previous section, nominal and ordinal attributes many be of any type, e.g. string, which requires to be converted into a corresponding integer representation. The process is straightforward in both cases (nominal and ordinal): each value is converted into an integer value that corresponds to its position in the value range. For example, user preference for language type has a value range of {"EN", "ES ", "CA", "GR:, :IT:}. Then a user profile with value "CA" is converted into an integer value "2" as shown in Figure 13. Similarly, numeric values are replaced with a representative integer value. For example, as shown in Figure 11 a value of -8 for *audio equalizer highs* belongs to bin 2, thus this value is replaced by the bin representative value of -7.

4.2.1.3 Profile integer coding

As described in the previous section, all preference values are represented by an integer value, regardless of being discretized or not. For example, the substiles language preference of "EN" is represented by an integer value of 0. A similar example from a discretized numeric preference is presented in Figure 11, where an audio equalizer highs preference of -14 is replaced by the value 0.

In the case of one attribute integer representation no problem occurs, all integer values are unique and their mapping is well known. However, in the case of a profile integer coding, combining all preferences integer representations imposes a problem. Such profile representation will have 0 for two different preferences, which makes impossible to specify to which preference zero value refers to.

Representing preference attribute values as integer values, although assures a one-to-one mapping

(uniqueness) in the content of preferences, it does not assure that in profile integer coding. We need to recode the one-to-one mapping of preference values in the content of the profile. We achieve that by aggregating all preference attribute values into one new unique dimension.

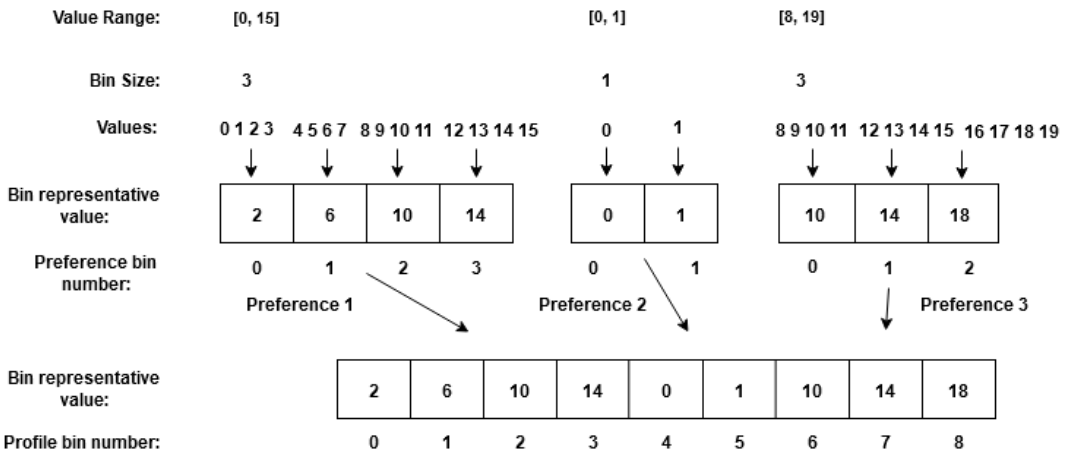

Figure 16 profile values transformation

Figure 16 illustrates an example of the aggregation process. The example shows the aggregation of three hypothetical dimensions keeping their unique one-to-one mapping between preference transformation and profile transformation. A preference2 value of 1 that is mapped to bin 1 in the content of preference value transformation is now mapped into bin 5 in the content of profile value transformation. Assuming Figure 17 user profile, after the integer coding phase, it is now represented as integer vector of values as shown in Figure 18, contextual and content information are also included in the representation.


```
{
    "user_id": 1,
    "user_profile": {"user_preferences": {"default": {"preferences": {
        "http://registry.easytv.eu/application/cs/audio/track": "en",
        "http://registry.easytv.eu/common/contrast": 100,
        "http://registry.easytv.eu/application/cs/audio/eq/bass": 10,
        "http://registry.easytv.eu/application/cs/ui/language": "en",
        "http://registry.easytv.eu/application/tts/audio/speed": 3,
        "http://registry.easytv.eu/application/tts/audio/voice": "female",
        "http://registry.easytv.eu/application/cs/ui/text/size": "20",
        "http://registry.easytv.eu/application/control/voice": true,
        "http://registry.easytv.eu/application/cs/cc/audio/subtitle": true,
        "http://registry.easytv.eu/application/cs/ui/audioAssistanceBasedOnTTS": true,
        "http://registry.easytv.eu/application/cs/accessibility/detection/character": true,
        "http://registry.easytv.eu/application/cs/ui/text/magnification/scale": true,
        "http://registry.easytv.eu/common/content/audio/language": "ca",
        "http://registry.easytv.eu/application/cs/accessibility/detection/text": true,
        "http://registry.easytv.eu/application/tts/audio/language": "en",
        "http://registry.easytv.eu/application/control/csGazeAndGestureControlCursorGuiLanguage": "ca",
        "http://registry.easytv.eu/application/cs/audio/volume": 2,
        "http://registry.easytv.eu/application/cs/accessibility/enhancement/image/type": "face-detection",
        "http://registry.easytv.eu/application/cs/cc/subtitles/font/size": 12,
        "http://registry.easytv.eu/application/cs/cc/subtitles/font/color": "#ffffff",
        "http://registry.easytv.eu/common/display/screen/enhancement/cursor/Size": 1.5,
        "http://registry.easytv.eu/application/tts/audio/volume": 21,
        "http://registry.easytv.eu/application/cs/cc/subtitles/language": "ca",
        "http://registry.easytv.eu/application/control/csGazeAndGestureControlType": "gaze_control",
        "http://registry.easytv.eu/application/cs/ui/vibration/touch": true,
        "http://registry.easytv.eu/application/control/csGazeAndGestureControlCursorGuiTextSize": 2.0,
        "http://registry.easytv.eu/application/cs/audio/eq/highs": 5,
        "http://registry.easytv.eu/application/cs/cc/subtitles/background/color": "#000000",
```

```
        "http://registry.easytv.eu/application/cs/accessibility/sign/language": "es",
        "http://registry.easytv.eu/application/cs/accessibility/detection/sound": true,
        "http://registry.easytv.eu/application/cs/accessibility/audio/description": true,
        "http://registry.easytv.eu/application/cs/audio/eq/mids": -8,
        "http://registry.easytv.eu/common/display/screen/enhancement/cursor/color": "#ffffff",
        "http://registry.easytv.eu/common/volume": 36
    }}}},
    "user_context": {
            "http://registry.easytv.eu/context/device": "pc",
            "http://registry.easytv.eu/context/light": 10,
            "http://registry.easytv.eu/context/proximity": 20,
        "http://registry.easytv.eu/context/location": "es",
        "http://registry.easytv.eu/context/time": "09:47:00"
    },
    "user_content": {
        "http://registry.easytv.eu/application/cs/accessibility/detection/text": true,
        "http://registry.easytv.eu/application/cs/accessibility/detection/face": true,
        "http://registry.easytv.eu/application/cs/accessibility/detection/sound": true,
        "http://registry.easytv.eu/application/cs/accessibility/detection/character": true,
        "http://registry.easytv.eu/application/cs/cc/subtitles/language": ["ca", "gr", "it"],
        "http://registry.easytv.eu/application/cs/audio/track": ["ca"]
    }
}
```


Figure 17 A user profile example


```
[8, 49, 52, 56, 58, 158, 163, 193, 214, 225, 230, 253, 452, 454, 456, 458, 460, 466,
468, 477, 481, 495, 499, 524, 530, 532, 536, 540, 544, 547, 549, 551, 553, 560, 564,
569, 596, 620, 564, 622, 624, 626, 628]
```


Figure 18 Profile integer representation

Each integer value represents a preference and its assigned value. Figure 19, shows a section of item to preference and their value mappings. For example, item 8 of the above profile, is a coding of the preference *http://registry.easytv.eu/common/volume=36.* More precisely, it represents the bin 8 that has range [33, 36] where the preference value (36) belongs. Some preferences group one value in each bin, thus their integer representation represents exactly that (e.g. item 158 below).

item	url	range	value
8	http://registry.easytv.eu/common/volume	33, 36	36
49	http://registry.easytv.eu/common/contrast	97, 100	100
52	http://registry.easytv.eu/common/content/audio/language	CA	ca
56	http://registry.easytv.eu/common/display/screen/enhancement/cursor/Size	1.5, 1.5	1.5
58	http://registry.easytv.eu/common/display/screen/enhancement/cursor/color	0, 167772	#ffffff
158	http://registry.easytv.eu/application/tts/audio/language	EN	en
163	http://registry.easytv.eu/application/tts/audio/speed	0, 4	3
193	http://registry.easytv.eu/application/tts/audio/volume	21, 24	21
214	http://registry.easytv.eu/application/tts/audio/voice	female	female

Figure 19 item to preference mapping

4.2.2. Frequent preferences set mining

Frequent preference sets are mined by counting their occurrences in the data set (i.e the set of user profiles included in the repository). Preference sets having $support \geq minimum\ support$ are

considered frequent. The mining process starts by generating candidate preference sets and counting their support. Assuming that available preferences are $\{p_1, p_2, p_3, \ldots . p_n\}$, each profile would contain a combination of these preferences. Then, the number of candidate preferences can be calculated as follows:

$$\binom{n}{1} + \binom{n}{2} + \binom{n}{3} + \cdots . \binom{n}{n} = \quad 2^n - 1$$

For n = 100 the number of candidate is 1.27×10^{30}. Some algorithms' implementation follows the **Apriori property** [4] principle to purge the number of generated candidates. Based on the Apriori property, all nonempty subset of a frequent itemset are also frequent. This property allows the purging of non-frequent sets reducing, thus, the number of produced candidate. Although this property is helpful in reducing the number of passes, their number still remains high. We choose to use FP-growth algorithm [5] to find frequent preferences sets due to its efficiency. FP-growth follows a different strategy: rather than counting the support of each candidate it codes the relation between data in data tree structure, called FP-tree. This tree is then used to mine frequent preference sets. Figure 20 describes how an FP-tree is generated for a set of itemsets (preferences).

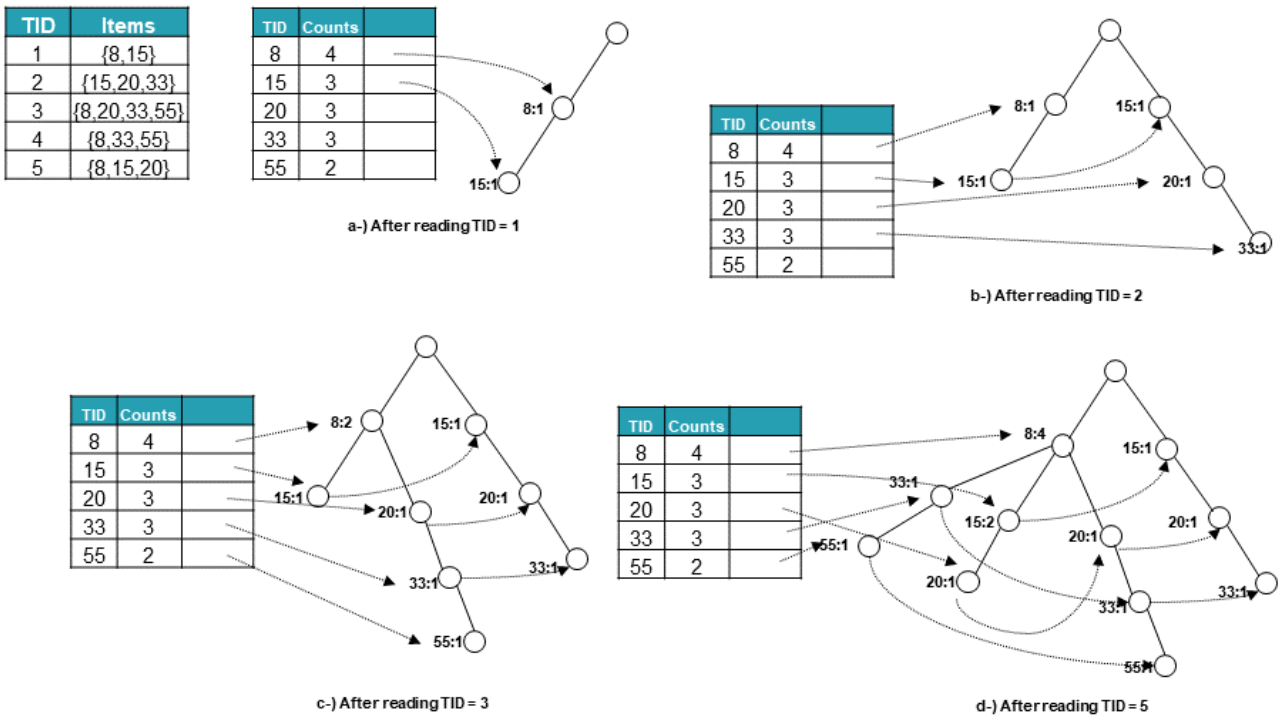

Figure 20 FP-tree

After the preprocessing step, a vector of integer values represents each user profile. As a result, mining frequent preferences sets is now reduced to mining frequent integers sets. The output of the mining process is a set of frequent preferences sets that we use to generate association rules.

4.2.3. **Association rules generation**

After generating the frequent preference sets the next step is to produce association rules. For each frequent preference set l, we generate all non-empty subsets. For each non-empty sub-set s of l

we calculate the confidence for the rule $\rightarrow (l - s)$. Valuable rules that are rules having $confidence \geq minimume\ \ confidence$. Thus, by assuming $l = \{p_1, p_2, p_5\}$ we generate the following rules $\{p_1, p_2\} \rightarrow p_5,\ \{p_1, p_5\} \rightarrow p_2,\ \{p_2, p_5\} \rightarrow p_1,\ p_1 \rightarrow \{p_2, p_5\},\ \ p_2 \rightarrow \{p_1, p_5\},\ \ p_5 \rightarrow \{p_1, p_2\}$

A rule of form $R_1\colon\ H \rightarrow B$ is a rule that says when h is true then B can be applied. H, B are the rule header and body and in our case consists of a set of preferences to be checked (in the header case) and a set of preferences to be applied in the case of the body. For example, assuming the following rule $R_1\colon\ h_1 \wedge h_2 \wedge h_3 \rightarrow b_1 \wedge b_2$, for a user profile that contains $h_1 \wedge h_2 \wedge h_3$ (the preferences and their associated values) we can then suggest him $b_1 \wedge b_2$ (these two preferences with their associated values).

After generating all association rules, we filter them out keeping only the most specific ones. For two rules, $R_1\colon\ H \rightarrow B,\ R_2\colon\ H' \rightarrow B'$ we say that R_2 is more specific than R_1 when $H \subseteq H'\ \ and\ \ B \subseteq B'$ and thus R_1 can be removed.

4.3. Rules refinement

The output of the association analysis is a set of association rules. These rules are then used for the refinement process of the set of semantic rules supported by the RBMM. The rules refinement step requires the ability to compare these two types of rules. Generated association rules are represented by integer values that correspond to preferences and their values, while RBMM rules are written in Apache Jena format (Figure 21). Thus, before starting the refinement process, we convert Jena rules into their corresponding integer representation following the same approach as described in section 4.2.1

```
[Face_detection_suggestion_rule:
 (?user rdf:type easyTV:User)
 (?user easyTV:hasPreference ?pref)
 (?user easyTV:hasSuggestedPreferences ?sugPref)
 (?pref easyTV:hasAccessibilityEnhancementImageType ?faceDetection)
 (?pref easyTV:hasCSSubtitlesFontSize ?sub_text_size)
 (?pref easyTV:hasCursorSize ?cursor_size)
 (?pref easyTV:hasCSUITestSize ?ui_text_size)
greaterThanEquals(?sub_text_size, '75'^^http://www.w3.org/2001/XMLSchema#integer)
equals(?cursor_size, '2.0'^^http://www.w3.org/2001/XMLSchema#double)
equals(?ui_text_size,   '23'^^http://www.w3.org/2001/XMLSchema#string)
equals(?faceDetection,   'none'^^http://www.w3.org/2001/XMLSchema#string)
->
(?sugPref easyTV:hasAccessibilityEnhancementImageType 'face-detection'^^http://www.w3.org/2001/XMLSchema#string)
]
```

Figure 21 Apache Jena rule example

At the following subsections, we describe the three possible output of the refinements process, namely updating, adding and removing rules. The order of the execution affects the process, where removing must be the first operation to be executed followed by updating and then adding. Let L, K be the sets of association rules and converted RBMM rules correspondently then as described in the following sections.

4.3.1. Removing rules

We considered removing rules, when these are no longer valid. A false suggestion rule is a rule that suggests preferences that contradicts the preferences suggestions of an association rule. $\forall\ R_1 \in L, R_2 \in K : R_1 : H \rightarrow B, R_2 : H' \rightarrow B' \wedge H = H' \wedge \neg B = B'$ .In other words, the association rule that contradicts RBMM's rule, is the rule that has the same header's condition with that rule, however, its body section suggests exactly the opposite set of preferences.

4.3.2. Updating rules

After the removing operation, the two sets contains only rules for updating or adding. Now to update RBMM rule $\forall\ R_1 \in L, R_2 \in K : R_1 : H \rightarrow B, R_2 : H' \rightarrow B' \wedge H = H'$ then we can update R_2 by replacing $B'\ with\ B$. In other words, the association rule used to update RBMM's rule, is the rule that has the same header's condition, however, differ in the body sections, which means that for the same set of condition it suggests different preferences. We must note there that we consider the case when $B = B'$ to be a special updating case.

4.3.3. Adding rules

All remaining association rules after the removing and updating operations have no corresponding RBMM rules, and thus are added to the rules supported by the RBMM.

5. INDICATIVE USE CASES

5.1. Profile refinement

We present a use case where we start with an active user profile and then based on his/her history of interaction a suggestion for profile modification is made. Assuming a user has activated the profile shown in Figure 3, and his/her history of interaction is shown below.

Table 5 Interaction events example


```
{{
    "user_action": {
        " http://registry.easytv.eu/application/cs/accessibility/detection/sound": true
    },
    "user_context": {
         "http://registry.easytv.eu/context/device": "tablet",
         "http://registry.easytv.eu/context/light": 10,
         "http://registry.easytv.eu/context/proximity": 20,
         "http://registry.easytv.eu/context/location": "es",
         "http://registry.easytv.eu/context/time": "09:47:00"
    }
},
{
    "user_action": {
        "http://registry.easytv.eu/application/cs/accessibility/detection/sound": true
    },
    "user_context": {
         "http://registry.easytv.eu/context/device": "tablet",
         "http://registry.easytv.eu/context/light": 10,
         "http://registry.easytv.eu/context/proximity": 20,
         "http://registry.easytv.eu/context/location": "es",
         "http://registry.easytv.eu/context/time": "11:00:00"
    }
},{    "user_action": {
        "http://registry.easytv.eu/application/cs/ui/vibration/touch": true
```

```
    },
    "user_context": {
        "http://registry.easytv.eu/context/device": "tablet",
        "http://registry.easytv.eu/context/light": 10,
        "http://registry.easytv.eu/context/proximity": 20,
        "http://registry.easytv.eu/context/location": "es",
        "http://registry.easytv.eu/context/time": "15:00:00"
    }
},{   "user_action": {
        "http://registry.easytv.eu/application/cs/accessibility/detection/sound": true
    },
    "user_context": {
        "http://registry.easytv.eu/context/device": "tablet",
        "http://registry.easytv.eu/context/light": 10,
        "http://registry.easytv.eu/context/proximity": 20,
        "http://registry.easytv.eu/context/location": "es",
        "http://registry.easytv.eu/context/time": "20:00:00"
    }
},{   "user_action": {
        "http://registry.easytv.eu/application/cs/accessibility/detection/sound": true
    },
    "user_context": {
        "http://registry.easytv.eu/context/device": "tablet",
        "http://registry.easytv.eu/context/light": 10,
        "http://registry.easytv.eu/context/proximity": 20,
        "http://registry.easytv.eu/context/location": "es",
        "http://registry.easytv.eu/context/time": "21:00:00"
    }
}}
```


Table 6 Merged events


```
{
    "user_action": {
        "http://registry.easytv.eu/application/cs/ui/vibration/touch": [null, true, null, null]
        "http://registry.easytv.eu/application/cs/accessibility/detection/sound": [true, true, null, true, true]
    },
    "user_context": {
        "http://registry.easytv.eu/context/device": ["tablet","tablet","tablet","tablet","tablet"]
        "http://registry.easytv.eu/context/light": [10, 10, 10, 10, 10],
        "http://registry.easytv.eu/context/proximity":  [20, 20, 20, 20, 20],
        "http://registry.easytv.eu/context/location": ["es","es","es","es","es"],
        "http://registry.easytv.eu/context/time": ["09:47:00” ,"11:00:00","15:00:00","20:00:00", ”21:00:00"]
    }
}
```


After sorting and merging steps (section 3.3), the result is the JSON file shown in Table 6. The process finds that the preference for sound detection occurred four times during the identified period. In relation to its value in the user profile, this preference has been regularly modified. The result then is to suggest to the user to refine the value of this preference to true.

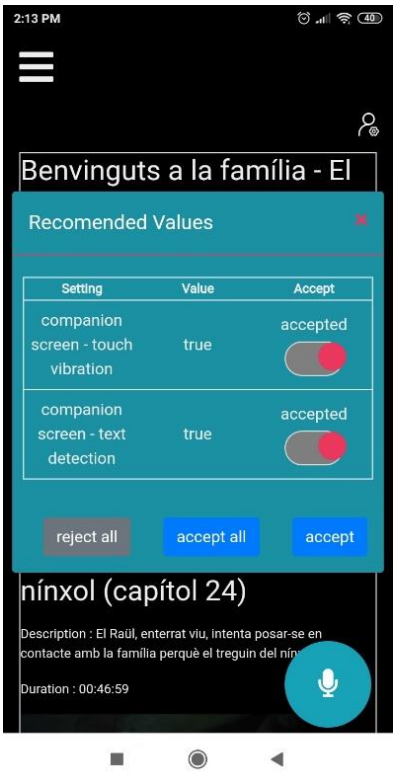

Figure 22 User profile refinements suggestions

5.2. Rules refinement

In this section, we present an indicative use case of the rules refinement process. The aim of this example is to illustrate the execution steps and produced data. In the use case description, we use a subset of actual preferences in order to reduce the amount of data produced and to make the example clearer. Preferences of Table 7 are the one used in the scenario and contain one of each preference type listed in 4.2.1, namely boolean, numeric and nominal. The set of user profiles contained in the repository is shown in Table 8.

Table 7 Indicative set of preferences

Item	Data Type	Value range	Short description
P_1	Boolean	[True, false]	Sound detection
P_2	Boolean	[True, false]	Face detection
P_3	Numeric	0 - 100	Audio volume
P_4	Nominal	[“15”, “20”, “23”]	Text size

P_5	Numeric	[1.5, 3.5] with 0.5 step	Image magnification

Table 8 User profiles repository example

ID	User profile
1	$\{P_1 = true,\ P_2 = true,\ \ P_3 = 0,\ \ P_4 = \text{"23"},\ P_5 = 3.5\}$
2	$\{P_1 = false,\ P_2 = true,\ \ P_3 = 24,\ P_4 = \text{"23"},\ \ P_5 = 3.5\}$
3	$\{P_1 = true,\ P_2 = true,\ P_3 = 1,\ \ P_4 = \text{"23"},\ P_5 = 3.5\}$
4	$\{P_1 = true,\ P_2 = false,\ P_3 = 0,\ \ P_4 = \text{"15"},\ P_5 = 2.5\}$
5	$\{P_1 = false,\ P_2 = true,\ P_3 = 20,\ \ P_4 = \text{"20"},\ P_5 = 3.5\}$
6	$\{P_1 = true,\ P_2 = false,\ P_3 = 0,\ \ P_4 = \text{"15"},\ \ P_5 = 1.5\}$
7	$\{P_1 = true,\ P_2 = false,\ P_3 = 0,\ \ P_4 = \text{"15"},\ P_5 = 1.5\}$
8	$\{P_1 = false,\ P_2 = true,\ P_3 = 20,\ \ P_4 = \text{"23"},\ P_5 = 3.5\}$
9	$\{P_1 = false,\ P_2 = true,\ P_3 = 10,\ \ P_4 = \text{"23"},\ P_5 = 3.5\}$

The preprocessing step (section 4.2.1) is the first one in the rules refinement process. This starts by reading all repository profiles and ends with converting them into an integer representation. The following tables show the program output for each preference, where statistical information and discretization are shown (section 4.2.1.2).

http://registry.easytv.eu/application/cs/accessibility/detection/sound

Attribute properties	
Range	Missing Value
[0.0 , 1.0]	-1.0

Bins histogram		
Id	0	1
Range	false	true
Center	false	true
Counts	4	5

Figure 23 Sound detection information

http://registry.easytv.eu/application/cs/accessibility/detection/face

Attribute properties	
Range	Missing Value
[0.0 , 1.0]	-1.0

Bins histogram		
Id	0	1
Range	false	true
Center	false	true
Counts	3	6

Figure 24 face detection information

http://registry.easytv.eu/common/volume

Attribute properties	
Range	Missing Value
[0.0 , 100.0]	0.0

Bins histogram																								
Id	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Range	0, 4	5, 8	9, 12	13, 16	17, 20	21, 24	25, 28	29, 32	33, 36	37, 40	41, 44	45, 48	49, 52	53, 56	57, 60	61, 64	65, 68	69, 72	73, 76	77, 80	81, 84	85, 88	89, 92	93, 96	97, 100
Center	2	7	11	15	19	23	27	31	35	39	43	47	51	55	59	63	67	71	75	79	83	87	91	95	99
Counts	5	0	1	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Statistical data					
Total	sum	Stand dev	Mean	Min	Max
9	75.0	9.7	8.3	0.0	24.0

Discretization properties			
Bins number	Bin Size	Remaining	Step
25	4	1	1.0

Values histogram	
Value	Frequency
0.0	4
1.0	1
20.0	2
10.0	1
24.0	1

Figure 25 Audio volume information with 10 bins discretization

http://registry.easytv.eu/application/cs/accessibility/magnification/scale

Attribute properties	
Range	Missing Value
[1.5 , 3.5]	0.0

Bins histogram					
Id	0	1	2	3	4
Range	1.5	2.0	2.5	3.0	3.5
Center	1.5	2.0	2.5	3.0	3.5
Counts	2	0	1	0	6

Statistical data					
Total	sum	Stand dev	Mean	Min	Max
9	26.5	0.8	2.9	1.5	3.5

Discretization properties			
Bins number	Bin Size	Remaining	Step
5	1	0	0.5

Values histogram	
Value	Frequency
2.5	1
1.5	2
3.5	6

Figure 26 Equalizer bass information with 31 bins discretization

http://registry.easytv.eu/application/cs/ui/text/size

Attribute properties	
Range	Missing Value
[0.0 , 2.0]	-1.0

Bins histogram			
Id	0	1	2
Range	15	20	23
Center	15	20	23
Counts	3	1	5

Figure 27 Text size information

From these outputs, we can identify, Figure 23, that sound detection preference is of Boolean type and the counted occurrences of false values are four and true values are five. Similarly, Figure 24 lists that face detection is of Boolean type and the counted occurrences of, false values are three and true values are six. Audio volume is a numeric preference that has been discretized into ten bins with the counted occurrences of values shown in Figure 25. A numeric preference with no discretization (or more specifically with bin size equals to one) are shown in Figure 26. Similarly, text size preference characteristics are shown in Figure 27.

After preferences discretization described above, each profile is converted into an integer representation as described in section 3.4. The output of profile integer representation is shown in Table 9, where each profile is represented as a set of integer values.

Table 9 Profile integer representation

ID	User profile
1	{1, 3, 4, 31, 36}
2	{0, 3, 9, 31, 36}
3	{1, 3, 4, 31, 36}
4	{1, 2, 4, 29, 34}
5	{0, 3, 8, 30, 36}
6	{1, 2, 4, 29, 32}
7	{1, 2, 4, 29, 32}
8	{0, 3, 8, 31, 36}
9	{0, 3, 6, 31, 36}

Table 10 Generated association rules

ID	User profile	Confidence
1	{1} => {4}	1.00
2	{4} => {1}	1.00
3	{3 } => {31,36}	0.83
4	{31} => {3,36 }	1.00
5	{36} => {3,31}	0.83
6	{3,31 } => {36 }	1.00
7	{3,36 } => {31}	0.83
8	{31,36} => {3}	1.00

Table 11 Generated rules

ID	User profile
1	`hasSoundDetection(true) => hasVolume(2)`
2	`hasVolume(2) => hasSoundDetection(true)`
3	`hasFaceDetection() => hasTextSize(23), hasMagnification(3.5)`
4	`hasTextSize(23) => hasFaceDetection(true), hasMagnification(3.5)`
5	`hasMagnification(3.5) => hasFaceDetection(true), hasTextSize(23)`
6	`hasFaceDetection(true), hasTextSize(23) => hasMagnification(3.5)`
7	`hasFaceDetection(true), hasMagnification(3.5) => hasTextSize(23)`
8	`hasTextSize(23), hasMagnification(3.5) => hasFaceDetection(true)`

With support value of 50% association analysis will found the following frequent preference set {8}, {8,36}, {3,8,36}, {0,3,8,36}, {0,8,36}, {3,8}, {0,3,8}, {0,8}, {32}, {4,32}, {1,4,32}, {1,4,29,32}, {1,2,4,29,32}, {1,2,4,32}, {4,29,32}, {2,4,29,32}, {2,4,32}, {1,32}, {1,29,32}, {1,2,29,32}, {1,2,32}, {29,32}, {2,29,32}, {2,32},

{2}, {2,4}, {1,2,4}, {1,2,4,29}, {2,4,29}, {1,2}, {1,2,29}, {2,29}, {29}, {4,29}, {1,4,29}, {1,29}, {0}, {0,31}, {0,31,36}, {0,3,31,36}, {0,3,31}, {0,36}, {0,3,36}, {0,3}, {1}, {1,36}, {1,3,36}, {1,3,31,36}, {1,3,4,31,36}, {1,3,4,36}, {1,31,36}, {1,4,31,36}, {1,4,36}, {1,3}, {1,3,31}, {1,3,4,31}, {1,3,4}, {1,31}, {1,4,31}, {1,4}, {4}, {4,36}, {3,4,36}, {3,4,31,36}, {4,31,36}, {3,4}, {3,4,31}, {4,31}, {31}, {31,36}, {3,31,36}, {3,31}, {3}, {3,36}, {36} and with confidence of 50% the association rules generated and filtered out are shown in Table 10. Association rules are then converted into Jean rules as shown in Table 11. Let us also assume that the RBMM set of rules are those shown in the following table.

Table 12 RBMM set of rules

Id	Jena rule
1	[Face_detection_suggestion_rule: (?user rdf:type easyTV:User) (?user easyTV:hasPreference ?pref) (?user easyTV:hasSuggestedPreferences ?sugPref) (?pref easyTV:hasTextSize '2.3'^^http://www.w3.org/2001/XMLSchema#integer) (?pref easyTV:hasMagnification '3.5'^^http://www.w3.org/2001/XMLSchema#string) -> (?sugPref easyTV:hasFaceDetection 'false'^^http://www.w3.org/2001/XMLSchema#boolean)]
2	[Cursor_size_suggestion_rule: (?user rdf:type easyTV:User) (?user easyTV:hasPreference ?pref) (?user easyTV:hasSuggestedPreferences ?sugPref) (?pref easyTV:hasFaceDetection 'true ^^http://www.w3.org/2001/XMLSchema#boolean) (?pref easyTV:hasMagnification '3.5'^^http://www.w3.org/2001/XMLSchema#string) -> (?sugPref easyTV:hasCursorSize 1.5 ^^http://www.w3.org/2001/XMLSchema#double)]

The refinement process will match association rules of Table 11 with those of Table 12. The removal of rules is the first step executed and based on the condition of rule removal the first rule is suggested for removal and that is because its set of headers is the same as those of the seventh rule (`hasTextSize(23), hasMagnification(3.5) => hasFaceDetection(true)`), however, the generated rule suggests enabling face detection. The second RBMM rule will be suggested for updating after being matched with the seventh generated rule (`hasFaceDetection(true), hasMagnification(3.5) => hasTextSize(23)`) and that is because both rules have the same condition, however, they differ in their body. The updated rule will then have the same conditions and the union bodies of both rules. The remaining generated rules have no matching RBMM rule and thus are added. The refined RBMM's rules are shown in the table below.

Table 13 RBMM's rules after refinements

Id	Jena rule

1	[Generated_rule_1: (?user rdf:type easyTV:User) (?user easyTV:hasPreference ?pref) (?user easyTV:hasSuggestedPreferences ?sugPref) (?pref easyTV:hasFaceDetection 'true ^^http://www.w3.org/2001/XMLSchema#boolean) (?pref easyTV:hasMagnification '3.5'^^http://www.w3.org/2001/XMLSchema#string) -> (?sugPref easyTV:hasCursorSize '1.5'^^http://www.w3.org/2001/XMLSchema#double) (?sugPref easyTV:hasTextSize '23'^^http://www.w3.org/2001/XMLSchema#string)]
2	[Generated_rule_2: (?user rdf:type easyTV:User) (?user easyTV:hasPreference ?pref) (?user easyTV:hasSuggestedPreferences ?sugPref) (?pref easyTV:hasSoundDetection 'true'^^http://www.w3.org/2001/XMLSchema#boolean) -> (?sugPref easyTV:hasVolume '2'^^http://www.w3.org/2001/XMLSchema#integer)]
3	[Generated_rule_3: (?user rdf:type easyTV:User) (?user easyTV:hasPreference ?pref) (?user easyTV:hasSuggestedPreferences ?sugPref) (?pref easyTV:hasVolume '2'^^http://www.w3.org/2001/XMLSchema#integer) -> (?sugPref easyTV:hasSoundDetection 'true'^^http://www.w3.org/2001/XMLSchema#boolean)]
4	[Generated_rule_4: (?user rdf:type easyTV:User) (?user easyTV:hasPreference ?pref) (?user easyTV:hasSuggestedPreferences ?sugPref) (?pref easyTV:hasFaceDetection 'true'^^http://www.w3.org/2001/XMLSchema#boolean) -> (?sugPref easyTV:hasTextSize '23'^^http://www.w3.org/2001/XMLSchema#integer) (?sugPref easyTV:hasMagnification '3.5'^^http://www.w3.org/2001/XMLSchema#double)]
5	[Generated_rule_5: (?user rdf:type easyTV:User) (?user easyTV:hasPreference ?pref) (?user easyTV:hasSuggestedPreferences ?sugPref) (?pref easyTV:hasTextSize '23'^^http://www.w3.org/2001/XMLSchema#integer) -> (?sugPref easyTV:hasFaceDetection 'true'^^http://www.w3.org/2001/XMLSchema#boolean)

	(?sugPref easyTV:hasMagnification '3.5'^^http://www.w3.org/2001/XMLSchema#double)]
6	[Generated_rule_6: (?user rdf:type easyTV:User) (?user easyTV:hasPreference ?pref) (?user easyTV:hasSuggestedPreferences ?sugPref) (?pref easyTV:hasMagnification '3.5'^^http://www.w3.org/2001/XMLSchema#double) -> (?sugPref easyTV:hasFaceDetection 'true'^^http://www.w3.org/2001/XMLSchema#boolean) (?sugPref easyTV:hasTextSize '23'^^http://www.w3.org/2001/XMLSchema#integer)]
7	[Generated_rule_7: (?user rdf:type easyTV:User) (?user easyTV:hasPreference ?pref) (?user easyTV:hasSuggestedPreferences ?sugPref) (?pref easyTV:hasFaceDetection 'true'^^http://www.w3.org/2001/XMLSchema#boolean) (?pref easyTV:hasTextSize '23'^^http://www.w3.org/2001/XMLSchema#integer) -> (?sugPref easyTV:hasMagnification '3.5'^^http://www.w3.org/2001/XMLSchema#double)]

6. HBBTV ACCESSIBILITY SERVICES

To further enhance the personalization process and also the accessibility of HbbTV for low vision and color blind people, CCMA developed two new accessibility services, HbbTV Screen reader and color subtitles. The screen reader is an accessibility service, based on existing CCMA HbbTV video portal that enhances the access to low vision users. Color subtitles, is an accessibility service that offers an alternative subtitles color set to ease the reading of subtitles for color-blind people

As accessibility services, both HbbTV screen reader and colored subtitles are included in the personalization process. Both services parameters are part of the user preferences as shown in Table 1, where these are tagged with HbbTV.

6.1. HbbTV Screen Reader

One of the major difficulties for blind and low vision people is the use of graphical user interfaces. A voice guide can help blind people interaction with electronic devices, as it reads the TV guide content to the user. On this basis, we have added a voice interface named HbbTV screen reader to our HbbTV service to guide the user's interaction with the device. The screen reader implementation is based on the EasyTV voice platform solution, which offers a text to speech service; the Screen Reader (SR) describes the graphical user interface considering the menu and sections' items helping users to find and consume content through voice indication. HbbTV apps in general still lack accessibility services, so this is the first time an HbbTV application covers the need of voice guiding.

The SR has been pre-tested with people of different types and levels of visual impairments gathering feedback information. In a second phase, that information helped in improving the SR throughout the process. In the next chapter, we give a detailed explanation on how it works.

6.1.1. **Overview**

The goal of the portal is to guide the user through available content. The interaction is conducted by the remote control. In order to grasp the content's organization Figure 28 shows a basic schema of the structure of the portal and Figure 29 shows how is presented to users.

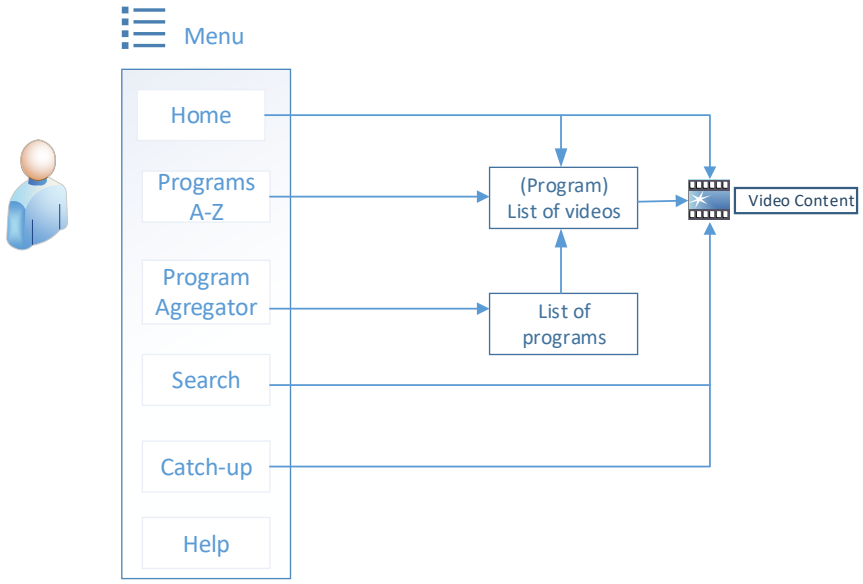

Figure 28 Content categorization schema.

The interface is composed of a menu section, which contains eight subsections that the user can select to navigate through CCMA portal. These subsections are Home, Search, Catch-up, Archive, Programs, Series, News and Sports and they are distributed in two rows and four columns.

Figure 29 HbbTV VOD portal – home

The user can navigate within the application using up-down and right-left arrows of the remote TV commander. The selected subsection is highlighted with a red box. This navigation style can be very comfortable for the general population, but it can impose an issue for people with special needs,

since the last ones do not know either where a row starts or ends, neither can appreciate where the red box is highlighting at any given moment. From user perspective, the settings must be simple. As is shown in Table 14, the available options are minimum, just enable or disable the accessibility service and speed choice (slow, normal, high). With this in mind, these adjustments are hidden from users and screen reader can easily be activated by pressing the green button ten times. Figure 30 shows HbbTV screen reader configuration page.

Table 14 HbbTV screen reader settings

Personalization of UI	Data type	Range	default
Enable/disable HBBTV screen reader	boolean	True, False	False
HBBTV screen reader speed	String	slow, "normal", "fast"	normal|

Figure 30 Screen reader configuration page.

6.1.1.1 Adaptation to existing accessibility standards

HbbTV applications are HyperText Markup language (HTML) based. HTML is the standard markup language for documents designed to be displayed in a web browser. In the implementation, we have considered the elements within the HTML document as a hierarchical structure and we have used Accessible Rich Internet Applications (ARIA) technical specification published by the World Wide Web Consortium (W3C) [6]. This specification allows to mark items within the HTML document that will be reproduced later as voice guide to help user's navigation. As we followed ARIA technical specification, we can guarantee that screen reader service will be fully compatible with future HbbTV versions. We have implemented a logical layer that detects the user-selected element and reproduces the proper audio each time. In order to define which text should be played and which not, we have structured the HTML documents in 3 views.

1. **overlay-texttospeech:** view, detects when an item has the focus or when an element is rendering to show the audios in which we are interested.
2. **arrows-navigation:** view, contains the navigable items of the HTML document.
3. **texttospeech-audio:** *view*, responsible of playing the audios.

Moreover, we have used three different types of aria tags depending on the message that we will show.

1. **aria-label**: contains the candidate text to be played.
2. **aria-labelledby:** for tagging the relevant information.
3. **aria-hidden:** is used to exclude information

6.1.1.2 Screen reader audio

To convert text to audio, we have used the EasyTV voice platform [7], which returns an audio file of the text content. HbbTV supports several standard audio formats, HE-AAC (mp4), mp3 and wav. Before connecting the application with the EasyTV voice platform we have conducted some tests with market representative TV models to determine which audio file format is more suitable. Although it would be preferable to use HE-ACC (mp4), due to the compression and the associated time response. In terms of interoperability mp3 format worked better, since some TV-sets were not able to reproduce HE-AAC properly. The choice between wav and mp3 depends on latency and accuracy, being the latency more important than accuracy for screen-reader requirements. Moreover, this results helped us parametrizing EasyTV voice service accordingly, as explained in D2.4, using mp3 for the audio file format. In addition, to make the system more effective, we applied caching techniques to reuse existing mp3 audio files, and avoid regenerating it from scratch repeatedly.

6.1.2. **Intermediate tests**

Intermediate test helped us to check the solution with real users during the project. A full description of these intermediate tests can be found here [8].Our first approach of just adding voice to an existing application, had implicit a lot of corner-cases that was not sustainable to construct a solution easy to maintain. As result, we decided to implement a specific navigation mode for SR mode. It solved a lot of issues and allowed to define some basic rules to keep this solution available beyond the project. In the current CCMA portal, the remote control arrows let users navigate and select content by moving the focus between adjacent items. Precisely the proximity method allows the users to navigate through the menus and access the desired content saving as much steps as possible. But this method is unpredictable and frustrating from blind users' perspective, as they cannot discern the menu graphical distribution. A representative example of the difficulties for visually impaired users is when they search information in the portal. Figure 31 shows the search subsection.

Figure 31 HbbTV VOD portal - search.

In this menu, users can search for multimedia content using a On screen keyboard. These types of keyboards are commonly used by the general population, but the interaction is quite difficult for vision

impaired users. Since they don't know where the keyboard starts and ends, they easily exit from it and jump to other sections. Consequently, to predict where the next element, is a big challenge.

To solve this drawback, we designed a new navigation solution. Keeping the proximity method on the default mode and the new one only available for the SR. Sounds obvious, but the solution was to redefine the navigation in accordance with the way that pc and mobile screen readers works. It means that up-down keys are reserved to jump from a section to another one, and left-right to enter on the section and traverse all the items. Moreover, we added some features to ease the navigation:

- When the user enters on a new page, a brief description will be played.
- The sections name are only played the first time that the user enters. Inside a section, the items are always navigated from first element left to right. That helps users keeping a proper order, although visually can be displayed differently.
- Dictionaries: Some titles, such as the days of the week are abbreviated on the UI. We implemented some quick translations to better reproduce and understand the days of the week.
- Fast Navigation: If the user repeatedly clicks on an item, only the last one is voice played. This is thanks an audio-queue implementation. If users decide to find new elements while a voice is playing, then current reproduction stops.
- Search virtual keyboard:
    - The new navigation solution allow navigation in alphabetical order, and hearing each focused character.
    - Hearing a 'beep' sound after selecting an element on keyboard. This notification sound helps notifying users that an element has been selected. As remote controls are not 100% effective and sometimes the TV do not process the order properly.
    - When new search results are shown, the user will be informed by a voice message that has higher priority over other navigation messages.
- Young users are used to higher screen reader voice speeds, otherwise elder and cognitive impairments population require slower speed. Thanks to the intermediate test, we were able to pre-configure the speed according their feedback.
- EasyTV CSapp will be able to launch, enable and personalize HbbTV SR.

6.1.2.1 How pages must be prepared

The results obtained from intermediate tests showed that blind people were pleased about the Screen Reader use. To maintain the compatibility with the system developed in the EasyTV project web pages must be designed following certain patterns. For instance, when working on the functional mockup of the candidate pages, it must be guaranteed that all navigable elements are visible and can be focusable as shown in figure below, and also define:

- Regions, wrapping siblings' elements.
- The element that when entering on the region that will receive the focus.
- And also organize the information on the HTML to construct coherent messages, for example: section-title-length.

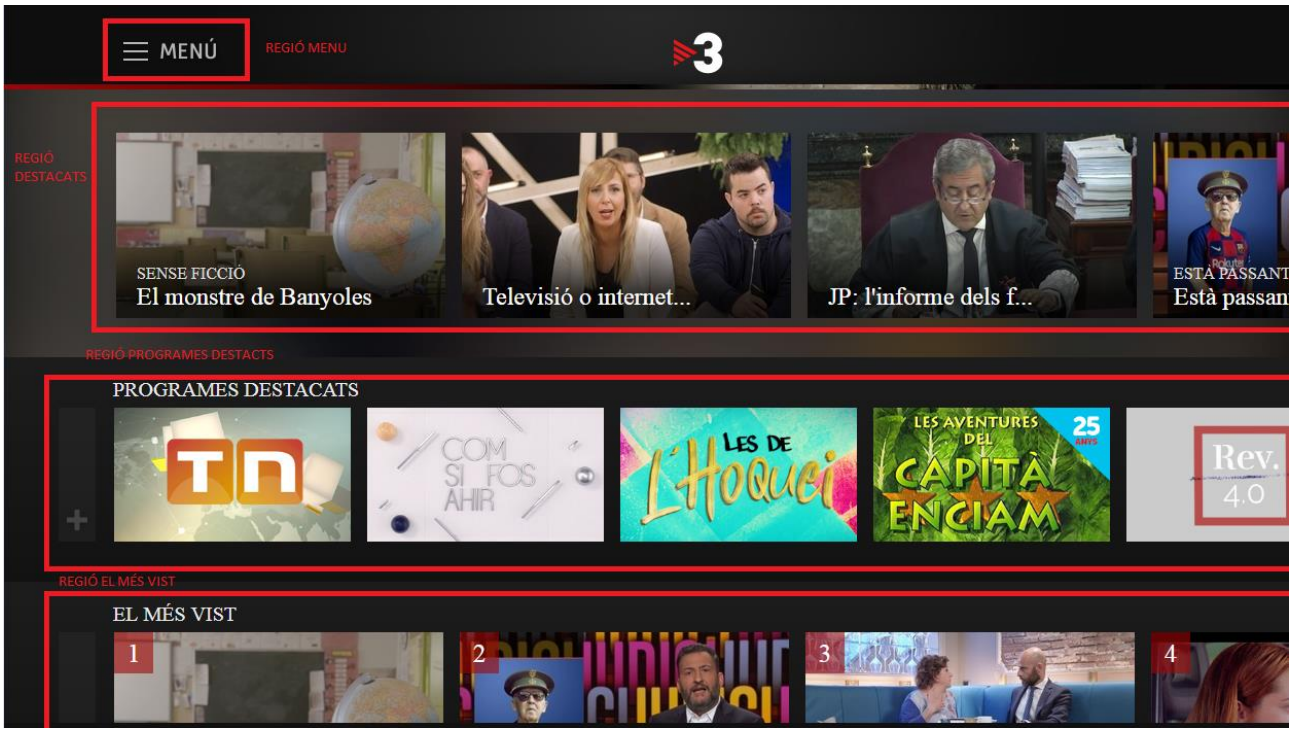

Figure 32 Region definition - navigation by sections.

6.2. HbbTV colored subtitles

HbbTV subtitles is an accessibility service for displaying subtitles on HbbTV screen. The service settings are shown in Table 15 and they allow the user to choose the language, the font size, the on screen position of the subtitles and whether he/she wants to enable background color or not. These settings are included in the user profile as shown in Table 1 and thus are considered in the personalization process. A part of HbbTV subtitles development was to propose and test a new set of color palette for color-bind people.

Table 15 HbbTV subtitles settings

Personalization of UI	Data type	Range	default
Enable/disable HBBTV subtitles	boolean	True, False	FALSE
HBBTV subtitles position	string	“up”, "down"	down
HBBTV subtitles language	string	“ca”,"en","ar", "es"	ca
HBBTV subtitles font size	string	[1, 50]	25
Enable/disable HBBTV subtitles background	boolean	True, False	FALSE

6.2.1. Subtitles alternative palette for color-blind

The idea behind this use case, was to research if we could improve the readability of subtitles to

color-blind users. Spanish law determines a color palette to differentiate actors. However, it has been confirmed during the research that color-blind users confuse that colors, which implies that they can't differentiate well the color of the speaking actor. Undoubtedly, the outcome of that research offers alternative palettes to the 3 different achromatopsia's (color blindness), which would enriched the EasyTV user model defined, to customize the subtitles presented to end users.

The default palette used for subtitles is given by 6 basic colors presenting high differences in color hue, more specifically 6 different hues with very different color, namely white, yellow, cyan or aqua, green, pink or fuchsia and red. They are ordered by their level of intensity, since intensity clearly helps readability in a variable background.

Figure 33 Default color palette.

Our *working hypothesis* is to define a specific color palette for each type of achromatopsia trying to maximize the perceived hues and intensities for color-blind subjects. Considering the reduced range of each specific achromatopsia the different number of hues is sensibly reduced, e.g. essentially 2. In order to maximize the number of hues we initially try to select the best hues for each specific achromatopsia. To select colors, we used ***Accessibility Color Wheel*** [9], which is a visual simulator that provides the perceived color by specific achromatopsia's. This simulator displays colored text on black or on white background. This simulator has been used on different displays, the most adequate selection would be with a calibrated display, but considering the uncalibrated nature of this project goal, we have preferred to work on different displays to try to overcome this intrinsic difficulty. This simulator is based on the algorithm given by Brettel et-al in [10], that builds the perceived colors for the 3 different achromatopsia's: Deuteranopia, Protanopia and Tritanopia. It is considered as one of the best models in doing so. The implementation of the simulation has been performed by Giacomo Mazzocato, who used the algorithms of Natterer et-al [11]. Using the simulator, we plot in figure 2 the perceived versions of the Default Palette. Where we can see that red is non-discriminable for deuteranopia and protanopia and green present a very low intensity for deuteranopia, which can make it very difficult to perceive. Pink also has low intensity and can present difficulties. Tritanopia perceived palette present good level of intensity but important confusions between cyan and orange, and pink and red.

DEFAULT Palette	Perceived Colour for the DEFAULT Palette		
Displayed	Deuteranopia	Protanopia	Tritanopia
#FFFFFF	WHITE (#FFFFFF)	WHITE (#FFFFFF)	WHITE (#FFFFFF)
#FFFF00	YELLOW (#FFD300)	YELLOW (#FFE900)	YELLOW (#FFB2C2)
#00FFFF	CYAN (#4C6CFF)	CYAN (#9BAEF9)	CYAN (#00D5FF)
#00FF00	ORANGE (#7D5000)	ORANGE (#FA9C00)	ORANGE (#039AFF)
#FF00FF	PINK (#112BEF)	PINK (#0009FF)	PINK (#D3030C)
#FF0000	RED (#2B1A00)	RED (#[illegible])	RED (#FF0003)

Figure 34. Perceived Default Palette (Simulation)

Using the simulator, and with the aim of the working hypothesis we proposed three different color palettes. Trying to represent the maximum number of hue differences and with two versions of the same hue, one more saturated and one more light, to ensure differences in intensity:

DEUTERANOPIA	
Displayed	Perceived Colors
#FFFFFF	WHITE (#FEFFFF)
#E18C00	ORANGE (#E08C00)
#F0FF28	SATURATED YELLOW (#FFE028)
#FFFF8E	LIGHT YELLOW (#FFF48E)
#FF77FF	LIGHT BLUE (#97AAFA)
#1FD5FF	SATURATED BLUE (#6C87FF)
#969696	GREY (#959696)

PROTANOPIA	
Displayed	Perceived Colors
#FFFFFF	WHITE (#FEFFFF)
#E18C00	ORANGE (#DF8C00)
#F7E914*	SATURATED YELLOW (#FFDD13)
#FFFF8E	LIGHT YELLOW (#FFF98D)
#FF77FF+	SATURATED BLUE (#758FFF)
#1FD5FF+	LIGHT BLUE (#98ACFC)
#969696	GREY (#959696)
#18F6FF*	LIGHT BLUE (#B1C0FB)

TRITANOPIA	
Displayed	Perceived Colors
#FFFFFF	WHITE (#FEFFFF)
#E18C00	PINK (#FF6072)
#64FFFF*	CYAN (#6AEFFF)
#FF0000*	RED (#FF0003)
#FF77FF	LIGHT PINK (#F18693)
#1FD5FF	BLUE (#22C4FF)
#969696	GREY (#959696)

Figure 35. Proposed Specific Color Palettes for different achromatopsia's

Using these 3 palettes we also derived a Universal palette, trying to keep the differences of the three different palettes into a single one.

Displayed	Perceived Colors according to different		
	Deuteranopia Simulation	Protanopia Simulation	Tritanopia Simulation
#FFFFFF	WHITE	WHITE	WHITE
#E18C00	ORANGE	ORANGE	ORANGE
#F0FF28	SATURATED YELLOW	SATURATED YELLOW	SATURATED YELLOW
#FFFF8E	LIGHT YELLOW	LIGHT YELLOW	LIGHT YELLOW
#FF77FF	LIGHT BLUE	LIGHT BLUE	LIGHT BLUE
#1FD5FF	SATURATED BLUE	SATURATED BLUE	SATURATED BLUE
#969696	GREY	GREY	GREY

Figure 36. Proposed Universal color Palette

6.2.2. **Intermediate tests**

These 4 proposed palettes (Figure 35 and Figure 36) jointly with the default palette (Figure 33) were tested in a psychophysical experiment performed with 6 subjects presenting different achromatopsia's. intermediate tests outcome can be found in [12]. In the next lines, we outline the main details of the experiment:

Subjects: 6 subjects participated in the experiment. 4 presenting deuteranopia, 1 with deuteranopia and protanopia and 1 with deuteranopia and tritanopia.

Stimuli: A video presenting a conversation between two people in an indoor scenario. Scene is presented with different subtitled versions. Each one presenting all the subtitles palettes proposed in the previous section. In this scenario we are assuming a random background for the subtitles, however, we want to note that this assumption should be revisited in further versions of the experiments, since this piece of video is presenting some bias to specific colors (in the gamut of purple, black, grey, white). The subtitles have been organized in order to have all color pairs of a specific palette appearing simultaneously. For each palette we have 15 or 21 pairs depending on the size of the palette.

Questions: Subjects were asked about three aspects on their experience: (a) Level of confusion between color pairs, (b) Readability of subtitles, and (c) Color Identification. The questions were not performed exhaustively, and therefore the answer were compiled by the comments of the subjects, providing a qualitative answer more than quantitative and accurate.

Method: The procedure to collect all the information from each subject was based on 4 steps:

Step 1. Observation of the video in the DEFAULT palette

Step 2. Observation of the previous video collecting the subject comments

Step 3. Observation of the video with the proposed palette accordingly with the subject achromatopsia.

Step 4. Observation of the previous video collecting the subject comments.

Results: The main results we can derive from the experiments are summarized in the following points:

- Overall, both palettes, the default and the adapted, show a similar performance.
- Color intensity seems an important property to improve readability
- The importance of background emerges in the developed approach.

Compiled comments has been quantitatively evaluated by the construction of the confusion matrix that contains the amount of comments about the confusion of a given pair of colors. From this matrix we have computed a confusion index (number of commented confusion/ number of possible pairs). The measured values are: (a) Index for the default palette is 5.3, (b) Index for the proposed palette is 5.7. Therefore, in general, there is no improvement and the level of confusion remains high. The main conclusion is that more research needs to be done. The level of difficulties reported by the subjects is huge, even more than we expected, and the experimental paradigm needs to be improved to collect the comments in a more accurate way in order to extract consistent information for all the subjects.

7. CONCLUSIONS AND FUTURE WORK

Improving the accuracy of the personalization process aims at improving the quality of the inferred suggestions. Better suggestions result to better user experience. The accuracy of the statistical matchmaking approach is improved by improving the quality of user profiles, while the accuracy of the knowledge-based approach is improved by refining the set of semantic rules supported by the RBMM. Data sources used in both approaches are users' available data (user profiles and corresponding interactions with the system), which may be data with high level of noise.

As with any data mining method, data plays major role in the quality of the produced results and thus in the refinement suggestions. To avoid errors and improve as much as possible the produced results, the final decision for accepting user profile refinements are taken by the user. On the other hand, RBMM's rules refinements are applied automatically. However, we are planning to improve the refinement process by taking into consideration cases that occur during the platform pilot running. As more and more users are using the platform, more data will be generated and collected and cases that have not been anticipated may occur. The feedback given from such data will be used to fine-tune the functioning of the system. Moreover, the newly added HbbTV accessibility services helped in gathering more information about user needs and preferences. Thanks to these two services, more information about the user HbbTV preferences is generated and used in the personalization process. HbbTV services have been subjected to intermediate tests, where the outcome has helped in indicating which features of each service to include and what modifications are required. The HbbTV screen reader tests results showed that the service may help at improving visually impaired users' navigation in the HbbTV portal. On the other hand, the test results show that the newly proposed color palette for HbbTV subtitles service does not improve readability. Therefore, it has been decided to exclude this feature from the HbbTV subtitles service. This experiment has allowed to establish the basis for further research in this field and has shown the huge difficulties of current subtitles for people with color blindness. The proposed palettes and the performed experiment have established an important basis for further research that should be founded on the following aims: a) The diversity of judgements between different subjects show the need for a TV that allows to the user to set individually the palette (User-specific palette). b) This work has allowed to establish the basis for a procedure to measure the adequacy of a palette for a user by measuring three different parameters: (a) confusion, (b) readability, and (c) identification. c) The subtitles background as an important factor in measuring the palette adequacy, where three approaches are proposed: (a) a background selected during the setting of a specific-palette, or (b) an adaptive subtitles that change depending on the background. D) Developing a full mathematical model of the perceived color simulation integrated in the procedure mentioned in first point, which can provide the best color perception conditions for the user.

8. REFERENCES

[1] P. S. Y. Charu C. Aggarwal, "Mining Large Itemsets for Association Rules," in *IBM T. J. Watson Research Center*, Yorktown Heights, NY 10598.

[2] M. S. A. K. V. K. Pang-Ning Tan, Introduction to Data Mining (Second Edition), Pearson, 2018.

[3] J. Han and M. Kamber, Data mining concepts and techniques, 3th edition, 2000.

[4] R. S. Rakesh Agrawal, "Fast algorithms for mining association rules," in *Very Large Data Bases (VLDB'94)*, Santiago, Chile, 1994.

[5] J. P. a. Y. Y. J. Han, "ining frequent patterns without candidate generation," in *Management of Data (SIGMOD'00)*, Dallas, May 2000.

[6] "Web Content Accessibility Guidelines (WCAG) Overview," [Online]. Available: https://www.w3.org/WAI/standards-guidelines/wcag/.

[7] EasyTV, "D2.4 – Integration of new access services in the".

[8] "D6.2 - Report on Intermediate tests: Ethical Requests, Methodology," p. 57.

[9] "Accessibility Color Wheel," [Online]. Available: https://gmazzocato.altervista.org/colorwheel/wheel.php.

[10] F. V. J. M. H. Brettel, "Computerized simulation of color appearance for dichromats," J. Opt. Soc. Am. A 14 (10) (1997) 2647–2655.

[11] "Code developed by four authors who are reachable by e-mail contact," Michael Natterer (mitch[at]gimp.org), Sven Neumann (sven[at]gimp.org), Robert Dougherty (bob[at]vischeck.com) and Alex Wade (alex[at]vischeck.com). [Online].

[12] "Report on Intermediate tests: Ethical Requests, Methodology," p. 62.