Noname manuscript No.
(will be inserted by the editor)

GeoHbbTV: A Framework for the Development and Evaluation of Geographic Interactive TV Contents

David Luaces · José R.R. Viqueira · Pablo Gamallo · David Mera · Julián C. Flores

Received: date / Accepted: date

Abstract Synchronizing TV contents with applications is a topic that has gained much interest during the last years. Reaching the viewers through various channels (TV, web, mobile devices, etc.) has shown to be a means for increasing the audience. Related to the above, the hybrid TV standard HbbTV (Hybrid Broadcast Broadband TV) synchronizes the broadcast of video and audio with applications that may be delivered through either the broadcast channel or a broadband network. Thus, HbbTV applications may be developed to provide contextual information for emitted TV shows and advertisements. This paper reports on the integration of the automatic generation of geographic focus of text content with interactive TV. In particular it describes a framework for the incorporation of geographic context to TV shows and its visualization through HbbTV. To achieve this, geographic named entities are first extracted form the available subtitles and next the spatial extension of

This work has been partially funded by the Galician Government (Xunta de Galicia) and FEDER funds of the EU under the Consolidation Program of Competitive Research Units (R2014/007, GPC2014/037, R2016/011).

David Luaces · José R.R. Viqueira · Pablo Gamallo · David Mera · Julián C. Flores
Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS), Universidade de Santiago de Compostela (USC), Rúa Genaro de la Fuente Domínguez s/n, 15782 Santiago de Compostela, SPAIN

David Luaces
E-mail: david.luaces@usc.es

José R.R. Viqueira
E-mail: jrr.viqueira@usc.es

Pablo Gamallo
E-mail: pablo.gamallo@usc.es

David Mera
E-mail: david.mera@usc.es

Julián C. Flores
E-mail: julian.flores@usc.es

those entities is used for the production of context maps. An evaluation strategy has been devised and used to test alternative prototype implementations for TV newscast in Spanish language. Finally, to go beyond the initial solution proposed, some challenges for future research are also discussed.

Keywords Interactive TV · HbbTV · Geographic Visualization · Geographic Annotation · Geographic Tagging

1 Introduction

The current habits of many TV watchers that simultaneously use a second device have been studied by various well known companies like Google. This current tendency of the audience to use a *second screen* is starting to be exploited by some advertising agencies. An example of the impact of the synchronization of these channels is that more than 25 million tweets were sent during each of the last three Super Bowls. The advances on the interactive TV (ITV) technologies enable the content producers to devise new strategies to improve the user experience during TV shows. Many innovative applications have been developed related to ITV, to provide enhanced view and control functionalities [4,10,13,1] , improve the accessibility [16,18,32], increase the interaction with and among the audience [14] and enrich the shows with possibly personalized [15] additional multimedia content [56,47]. Related to this, the new standard for hybrid TV HbbTV [1] (Hybrid Broadcast Broadband TV) enables the broadcast of web applications together with the standard Audio/Video (A/V) content, to be displayed together in a TV device. Thus, HbbTV applications may complement TV shows and advertisements with additional functionality. These applications may be accessible both through the broadcast channel and through broadband connections.

Many TV shows contain references to geographic entities whose locations are assumed to be known by the audience. However, in practice many viewers do not have clear knowledge of those locations. Such situation arise for example in TV newscasts, which have a plethora of references to geographic entities. To provide adequate context for some of such entities, sometimes maps are displayed on the screen. However, those maps are few and static, disappear too quickly, and are not adapted to the geographic knowledge of each user. It would be desirable therefore to give additional control to the viewer to explore the geographic context of the TV newscasts.

On the other hand, many research efforts have been devoted during the last years to the automatic enrichment of contents [37], even within the scope of TV programs [47]. More specifically, many approaches have dealt with the automatic generation of a geographic focus for text documents [21,5,43,50]. In particular, some specialized recent approaches have been proposed for news documents [52,41,20].

[1] https://www.hbbtv.org/

Although ITV approaches have been proposed to enrich the TV shows with additional content, they either do not generate and synchronize the additional content automatically [56] or they do not focus on geographic knowledge [47]. It is noticed however that, in addition to the A/V content, subtitles are also generally available. In particular, Spanish regulations require public TV services to provide subtitles for the 90% of their broadcasts. Therefore, this paper shows that automatic and synchronized generation of geographic maps from the subtitles is a feasible approach that had not been attempted yet.

In this paper, a framework that enables the generation and incorporation of geographic context within interactive TV shows is described. The framework makes use of the HbbTV technology to deliver a web application that enables the viewer to browse contextual maps adapted to the contents of the TV shows. To achieve this, each subtitle of the show is analyzed to extract geographic named entities, composed of names and coordinates, which are next combined with an available representation of the audience geography knowledge to generate the extent of the relevant map. An evaluation framework was also devised and used to test different alternatives during the implementation of a first prototype, focused on TV newscast in Spanish language. The evaluation results show that a combination of the rule based named entity extractor, implemented as part of this work, with a geocoding service achieves an F-Score of 0.82 for geographic named entity identification, and of 0.65 for map extent matching. To the best of these authors knowledge, the present work is original, since it is the first attempt reported in the literature to automate the enrichment with geographic content in the ITV context.

The main contributions of the paper are the following. i) The design of a framework and the implementation of a first system prototype to tackle with the problem of automatic generation of time evolving geographic focus from subtitles, which had not been attempted by any previous approach. ii) The modification of an already existing rule based Named Entity Recognition (NER) technology that enabled obtaining results that improve other state of the art Named Entity Recognition (NER) technologies. iii) The proposal of an evaluation strategy, which is based on the definition of well-know precision and recall measures for geographic entities and map extents. iv) The identification of various research challenges that may guide future research lines in this specific area.

The remainder of this paper is organized as follows. Section 2 provides some background on HbbTV standard and a more detailed specification of the framework requirements. Other pieces of research work related to the present one are discussed in Section 3. Section 4 provides a description of the framework components architecture. The key component of the architecture, i.e., the one in charge of the generation of the geographic focus of each subtitle is described in Section 5. A brief description of the implemented prototype for TV newscast and Spanish language is given in Section 6. Section 7 describes the evaluation strategy and discusses results with different alternative implementations. Section 8 discusses some challenges for future research and Section 9 concludes the paper.

2 Preliminaries

An overview of the HbbTV standard is provided below together with a detailed description of the requirements assumed during the design and implementation of the framework.

2.1 Hybrid Broadcast Broadband TV

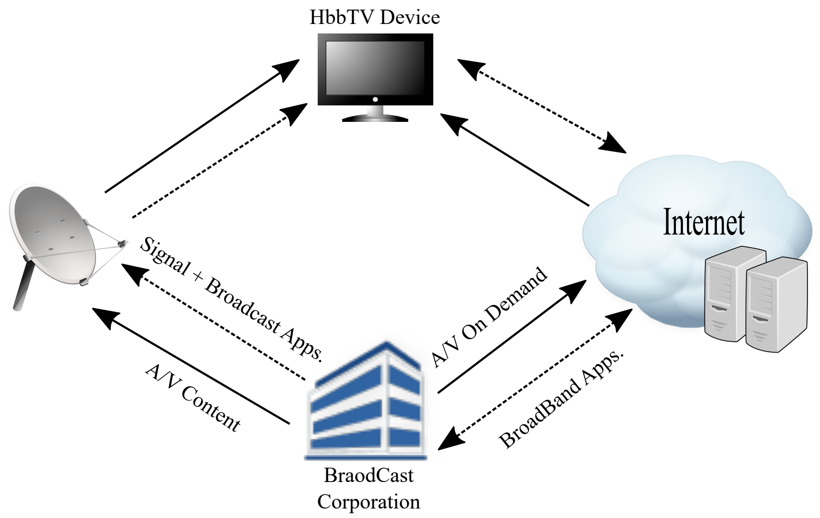


Fig. 1 HbbTV 1.2 General Architecture.

The HbbTV standard specifies the components and defines the interfaces that a hybrid broadcast broadband television systems should have. The general architecture of the system is depicted in Figure 1. As it is shown in the figure, a main difference with currently dominant smart TV platforms is that the application provider is not related to the device provider. In fact, applications and A/V contents may be provided by the same entity, therefore, they can be perfectly synchronized. This is a main advantage of this new technology with respect to conventional smart TV platforms.

Content providers use an up link to broadcast standard A/V content. Besides, applications may be delivered either through broadcast or broadband communication channels. Therefore, some applications may still reach population with limited broadband access, assuming the lack of a return broadband channel from the user device. This is still useful for some rural areas in Spain for example. Broadband applications may be either "broadcast-related" or "broadcast-independent". The former are synchronized with the A/V content by specific signaling. The latter are like conventional smart TV applications. Notice that compared to conventional web applications, HbbTV applications have limited user interaction capabilities, due to the use of standard remote control devices, but on the other hand they are suitable to both deliver video on demand and synchronize with live A/V broadcasting.

2.2 Platform Requirements

Broadly speaking, the objective of the platform is to incorporate at each time instant of a TV show, a map with geographic references that helps the viewer with relevant geographic context. Clearly, the above statement is too rough to be used as solid basis for requirements specification. In particular, the main question to be answered is: Which is the correct map to be shown at each specific time instant? Obviously, the reply to the above question depends a lot on the preferences of each viewer and on the knowledge she has about the relevant geographic area. To illustrate the above consider the following four consecutive subtitles.

- S#1: 'after stopping her on a street in north London.'
- S#2: 'The man cycled towards the girl in Enfield'
- S#3: 'saying she had been caught shoplifting and he would search her.'
- S#4: 'The attack in Sandhurst Road at 17:30 GMT on Wednesday'

It is clear that various different combinations of maps could be used to provide geographic context for the above sequence. For example, during S#1 a map with the extent of either 'London' or 'north London' could be used. The former provides better context whereas the latter is more precise. During S#2 the framework could either provide a new map with context for 'Enfield' or incorporate 'Enfield' in the map already shown for S#1. During S#3 the system must decide whether to maintain or not the map shown during S#2. Finally, regarding S#4 there is more than one 'Sandhurst Road' in London, however only one is in Enfield. Besides, the location of the street could be shown in either a map of London or Enfield or even narrower map, depending on the geographic knowledge of the viewer.

It is important to remark that the above problem has some particularities compared with the generation of geographic focus for text documents [21, 5, 43, 50]. In particular, first, the geographic focus is here used for visualization and not for spatial searching, therefore its extent must provide appropriate context adapted to the geographic knowledge of the viewer. Second, the temporal nature of the document (TV show) demands the generation of an appropriate time evolving focus.

To enable the construction of an initial solution, specific requirements where assumed in the scope of the above described general problem. In particular,

- A new map is generated for each subtitle, using only the geographic references obtained from its text.
- If a subtitle does not have any geographic reference then the map of the previous subtitle is maintained.

- The geographic extension of each map must contain the geometries of all the extracted geographic references, accompanied by the geometry of another geographic entity that provides appropriate context, as it is estimated to be well-known by the audience.
- All the subtitles are already available from the teleprompter before the emission of the TV show, therefore, real time geographic focus generation is not required.

3 Related Work

Related work on both interactive TV applications and the enrichment of content with geographic references is revised in the following two subsections. It is noticed however, that despite of the many examples of TV applications found in the literature, and of the many efforts devoted to the automatic generation of geographic references for text documents (including news documents), the automatic enrichment with geographic content for interactive TV has not been yet attempted, to the best of these authors knowledge.

3.1 Interactive TV Applications

Interactive Television (ITV) combines traditional TV streaming with applications whose functionality bring new opportunities to improve the active participation of the audience in the communication process. ITV applications may be used for many different purposes, including among others: i) provision of traditional services such as Video on Demand (VOD) and Catch-up, ii) enabling multicamera and immersive environments, iii) providing enhanced control means, iv) improving accessibility, v) enabling enhanced social interaction, vi) achieving content personalization and vii) providing additional related multimedia content. Relevant research initiatives in all the above topics are reviewed in this subsection.

Traditional applications deal with basic Video on Demand (VOD) and Catch-up services, that give the viewer some freedom to choose the content and the moment to consume it. Multicamera applications enable the audience to consume a given show from different points of view, as it is reported in the TVRing project[4]. One step beyond is achieved by the creation of more immersive TV experiences by the use of ommidirectional video and Virtual Reality (VR) technologies [4,10].

A common objective of many ITV applications is to give the user additional control over the selection of the TV channel she wants to see. Thus mobile devices may be used to provide enhanced control over the TV [4,13]. Besides, advanced search engines enable channel selection in a way that goes much beyond traditional Electronic Program Guide (EPG) functionalities [1].

A very important use of ITV technologies is related to the improvement of the accessibility to already existing applications. This is mainly due to two reasons. First, a simple and well known interface based on a conventional remote control makes it suitable for users with limited skills like elderly citizens. Second, in general, the TV signal reaches an amount of population broader than the one reached by broadband Internet communications. This is the case of some rural areas in many countries. Thus, for example, in the scope of the TV-WEB project [31] multimedia content recorded in a Content Management System is offered through a broadcast HbbTV application to reach population without Internet access. The experience of elderly users with the application is evaluated as part of the work of the project [32]. A TV Facebook application was developed in [16], whose functionality includes sharing TV show information and photos and remote control with voice and gestures. In [18], a comparative evaluation of the accessibility of conventional and TV interfaces is performed, including well-known applications like Facebook, Twitter and Accuweather. The results show that TV applications conform better with accessibility standards. Finally, a new approach to provide accessible interactive services through ITV in the context of and ambient intelligence environment is described in [24], which is specially focused on elderly users.

Another important functionality is related to the use of ITV applications to improve the interaction with and among the audience. This is the main goal that lies behind the so called Social TV [14]. The study presented in [59] shows, among other results, the general interest of the people for applications that improve social interaction within home, by sharing multimedia content among collocated viewers. Social experiences within home are also the objective of [60]. In [23], guidelines for the design of social TV prototypes are given, based on an study of the interactions of people in front of a TV. Finally, the factors that influence the likelihood of using social TV applications are analyzed in [38].

ITV capabilities can also be used for content personalization (see [15] for a review of relevant research). Broadly speaking, two major lines of work have been explored in this area, namely, content recommendation in the current context of large lists of TV channels and personal advertising [8,35].

Finally, and closely related to the work described in the present paper, ITV applications enable the enrichment of traditional TV streaming with synchronized additional multimedia content. In some second screen approaches [2] the TV shown is recognized using audio fingerprints and relevant multimedia content is next selected from some content management system to be presented in the second device. Various companion applications have been developed to support programs with many characters by means of story maps [45,46,55]. A companion second screen application for the Olympic games was evaluated in [7], whose functionality enabled both enhanced control of TV from a second device and the enrichment of the content with statistics and other relevant information. The results highlighted, among other issues, the importance of supporting parallel viewing between collocated viewers. In [56] an HbbTV application is described, which illustrates the enrichment of content in the context of informal learning scenarios in a German TV. In particular, a news program is segmented into topics of 2-5 minutes, and additional content, including geographical maps, is available for the viewer for each segment. Contrary to the approach described in the present paper, automatic generation of content is not an objective of [56].

Regarding the automatic enrichment of content, the annotation of contents with references to semantic resources has been a topic to which much attention was paid during the last decade [37]. Documents are annotated with concepts of ontologies to both enrich their content and enable semantic search. As an example, a recent approach is proposed by [62] to semantically annotate documents with linked data. In particular, document terms are linked to subgraphs of reference ontologies, which may next be navigated by the user to deepen the document topics. Closer to the present work, in [47], semantic descriptions are generated for broadcast sports videos. In particular, the closed-caption text is processed to extract information about plays and players, which is next incorporated in the video stream. Notice that the content automatically generated in [47] does not have geographic nature.

3.2 Geographic Content Enrichment

In the area of Geographic Information Retrieval (GIR) [34] the tagging of web contents with geographic references has been a challenge for researchers for many years [21, 5, 43, 50]. Two main steps are identified in the geotagging process. First, geographic entities must be identified among the terms of the input document. This phase is usually called geoparsing. Disambiguation is a major challenge during this stage. Notice for example that the same term might denote both geographic and non-geographic entities. Furthermore, two different geographic entities might share exactly the same name. Second, the geographic coordinates of all identified entities must be combined in some manner to construct a geographic focus for the whole document. Various different solutions may be adopted in this stage.

Various tools and services are already available for the geoparsing of text documents. Some examples are Yahoo PlaceSpotter (a service of the Yahoo BOSS geo services, accessible through table Geo.Placemaker of Yahoo Query Language (YQL) [2]), the Cartographic Location And Vicinity Indexer (CLAVIN) [3] and Open Calais [4].

The identification of geographic entities is generally achieved by a two step process. First, place names have to be extracted from the text [61, 29]. Usually, natural language processing (NLP) techniques are applied. In particular, specific geographic entity recognition functionality is provided by NLP tools such as the Standford Named Entity Recognizer (SNER) [25], GATE [19], Apache OpenNLP [5] and FreeLing [48]. SNER is considered as the state-of-the-art for

[2] https://developer.yahoo.com/yql/

[3] https://clavin.bericotechnologies.com/

[4] http://new.opencalais.com/

[5] https://opennlp.apache.org/

English language [33], whereas FreeLing achieved the best performance for Spanish language at the CoNLL shared task [58].

In a second step, usually called toponym resolution, geographic coordinates have to be assigned to each entity obtained from the previous phase. Digital gazetteers (see [30] for a good introduction to relevant research) and geocoding services such as Nominatim [6] and Geonames [7] are usually combined with specific heuristics. Disambiguating toponyms is a major challenge at this stage [11]. In general, a score has to be computed for each candidate toponym based on the context of the input place name. Different approaches may be adopted for the calculation of this score. A recent and novel technique for place name disambiguation has been proposed in [3].

Various specific document geotagging solutions have been proposed in the literature. In [63], the provider, content and serving locations of web documents are extracted from their content. Proposals adapted to the geotagging of short texts are also proposed. As an example, twitter messages are processed in [28]. Tweets that reference multimedia documents are analyzed in [17] to determine their geographic focus. The challenges of dealing with references to vernacular regions, which are not generally referenced by geocoding resources, are discussed in [49]. Semiautomatic geocoding of Greek persistent web pages is undertaken in [6]. Finally, the geotagging of multimedia contents has already been surveyed in [44].

More closely related to the objective of the present paper are the various recent approaches and solutions devoted to geotagging of resources with news contents. A reference system in this area is NewsStand [57,52]. RSS feeds from thousands of on-line news resources are collected, processed and visualized in interactive maps, within minutes of publication. NewsStand uses a custom-built geotagger to extract the geographic focus from the text [42]. Next, news are grouped into story clusters using a fast on-line clustering algorithm. Those story clusters are intended to have news related to the same event. Clustering is performed based on the existence of both related terms and temporally proximate dates of publication. Different stories are shown to the user depending on the position and zoom level of the map. Thus, stories are displayed according to both their significance and geographic location. A system related to the previous one is TwitterStand [53], whose main objective is the capture of tweets that correspond to late breaking news.

Another news related approach is the one described in [41]. News documents are enriched both with geographic focus and multimedia content. Again, a map based interactive interface enables the browsing of the news. Toponym extraction is based on Apache OpenNLP. Toponym resolution combines the use of Geonames with Wikipedia. Finally, a relevance measure is computed between locations and documents. Such a measure is used during the interactive geographic browsing.

[6] http://www.nominatim.org/

[7] http://www.geonames.org/

The GeoRank system [9] enables the retrieval of top-k most recent news documents related to the location of the user. Spatial proximity, temporal proximity and user preferences are used to rank the documents. A spatio-temporal pruning technique is implemented that improves the response time and efficiency of the query processor.

A final related solution is the CLIFF-CLAVIN approach described in [20]. This approach adapts the pipeline outlined in [5] for the geotagging of news. The toponym extraction and resolution stage is based on an improvement of the already existing CLAVIN tool. In particular, improvements focus on the disambiguation stage.

Notice that none of the above approaches try to enrich dynamic content, like TV streaming, with time evolving maps with visualization purposes.

4 System Architecture

An overview of the system architecture is depicted in the diagram of Figure 2. As it is shown in the figure, the system is divided into three main components, namely a *Broadcaster*, a *Geographic Focus Generator* and a *Geo Context Viewer*.

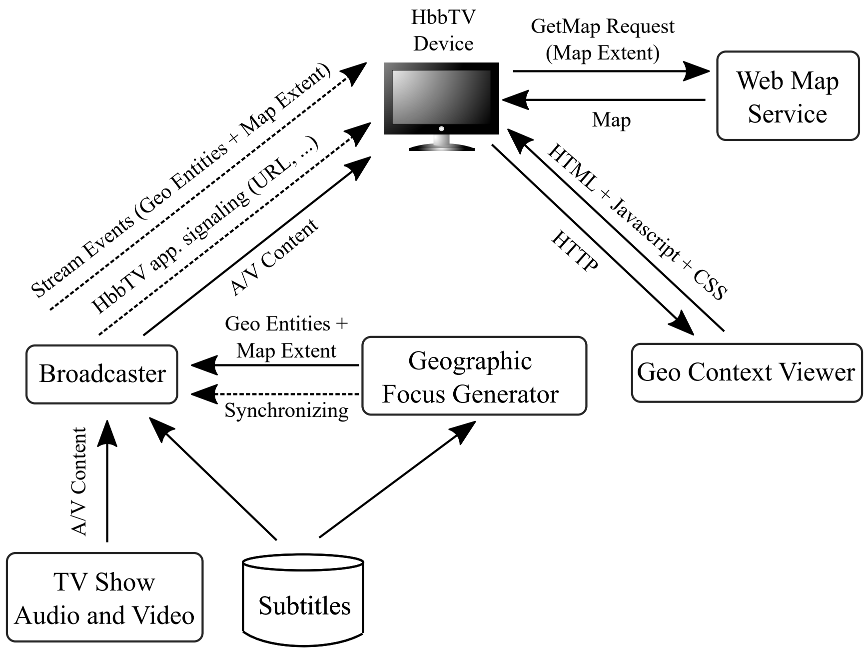


Fig. 2 System Architecture.

Broadcaster: This component obtains the A/V content and the subtitles, and broadcasts them through an up-link. Besides, synchronously, it gets from the *Geographic Focus Generator* the geographic entities and map extent that defines the context map to be displayed at each time instant. Such entities and map extent are also delivered as stream events, which are synchronized with the A/V content. Finally, the URL of the "broadcast-related" HbbTV application used to display the show together with the appropriate map is also transfered through the up-link in the form of HbbTV application signaling.

Geographic Focus Generator: This is the key component of the framework. It has to analyze the text of each subtitle in order to identify occurrences of references to geographic named entities. For each such reference it obtains both the name of the entity and the coordinates of the point that defines its location in the earth surface. Next, it generates a minimum map extent (rectangle) for the subtitle that contains both all the extracted entities and some additional context entity that is assumed to be known by the audience. Notice that although real-time requirements are not assumed, this component must be synchronized in some way with the broadcaster, in order to send each focus at its appropriate time instant. A more detailed description of the functionality of this component is given in Section 5.

Geo Context Viewer: This is a "broadcast-related" HbbTV application implemented with HTML, Javascript and CSS that enables the visualization of the newscast with maps produced using the context information retrieved by the *Geographic Focus Generator*. Once the device is tunned to the specific TV channel, it receives through the broadcast interface the A/V content of the TV show and the HbbTV application signaling, which contains the URL of this application, required to download it through the broadband interface. Once the application has started, it listens to the Stream Events of the broadcast interface to obtain the geographic entities and map extents transmitted at each instant. Upon the reception of each event, the application uses the map extent to send a GetMap request to an external Web Map Service (WMS). The response map image is combined with the geographic entities to produce the desired map in the GUI, which also integrates the A/V content.

It is remarked that, at this stage, it is not an objective for the system to provide a professional fully functional *Geo Context Viewer*, which should include both a more friendly GUI and additional pieces of functionality.

5 Automatic Geographic Focus Generation

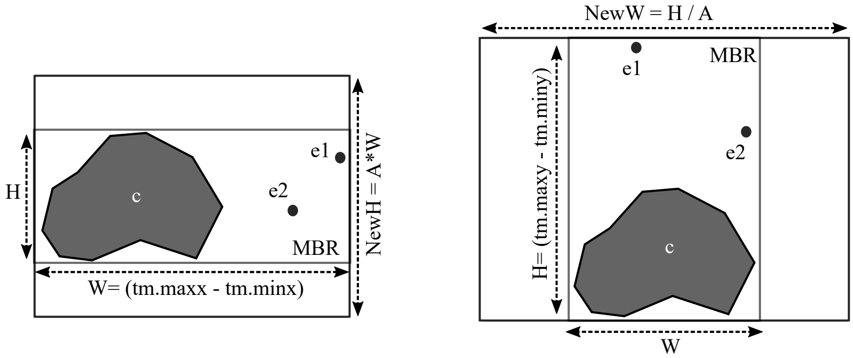
This section describes in some detail the functionality of the key component of the framework, i.e., the one that generates the temporal evolution of the geographic focus of the TV Show, according to the requirements described in Subsection 2.2.

Let us provide a more detailed description of the inputs and outputs of the component. The input of the component is a sequence of subtitles S, where the text and time instant of each subtitle $s \epsilon S$ are denoted $s.text$ and $s.time$, respectively. The output of the component is a sequence of triples $(time, E, m)$. Component $time$ is the time instant of the corresponding input subtitle. E is a set of geographic named entities $\{e_i\}$ extracted from the text of the subtitle. Each e_i contains both a name ($e_i.name$) and a point ($e_i.point$). Component m is a minimum bounding rectangle that contains both all the points $e_i.point$ and some geometry c from a list of context geometries C, which are assumed to be well-known by the audience. It is reminded that real-time requirements are not assumed, given that the above subtitles are available from the teleprompter before the emission of the show. However, the audio and video content of the TV show and the generated focus must be synchronized in the emission. This is achieved by stamping each focus with the time instant of its corresponding subtitle. If some time instant does not have a result focus, then the focus of the previous instant is maintained in the emission. The pseudocode of Algorithm 1 describes in more detail the behavior of the component.

Algorithm 1 Geographic Focus Generation.

```
 1: procedure GeoFocus(S, C, A)
 2:     for each s ϵ S do
 3:         E ← ExtractGeoEntities(s.text)
 4:         if E ≠ ∅ then
 5:             for each c ϵ C do
 6:                 tm ← mbr(E, c)
 7:                 if A > (tm.maxy − tm.miny)/(tm.maxx − tm.minx) then
 8:                     cy ← (tm.maxy − tm.miny)/2
 9:                     NewH ← A ∗ (tm.maxx − tm.minx)
10:                     tm ← rectange(tm.minx, cy − NewH/2, tm.maxx, cy + NewH/2)
11:                 end if
12:                 if A < (tm.maxy − tm.miny)/(tm.maxx − tm.minx) then
13:                     cx ← (tm.maxx − tm.minx)/2
14:                     NewW ← (tm.maxy − tm.miny)/A
15:                     tm ← rectange(cx − NewW/2, tm.miny, cx + NewW/2, tm.maxy)
16:                 end if
17:                 if m isnull ∨ area(tm) < area(m) then
18:                     m ← tm
19:                 end if
20:             end for
21:             return (s.time, E, m)
22:         end if
23:     end for
24: end procedure
```

Input parameters S, C and A contain respectively the set of input subtitles, the well-known context geographic entities and the aspect relation (height/width) of the output map area. Function *ExtractGeoEntities* in line 3 is applied to the subtitle text to obtain a set of geographic entities E. An step beyond the state of the art geoparsing technology is achieved in this key step in the current prototype implementation, as it will be shown in Section 6. This is a key part of this component whose main problem is the resolution of ambiguities. In particular two types of ambiguities have to be solved in two different phases. First, entities have to be recognized in the text and ambiguities between geo and non-geo entities must be solved (‘Santiago’ as name of either a place or a person). Second, coordinates have to be assigned to geographic entities and ambiguities between various geo entities have to be solved (‘Santiago’ as either Santiago de Chile or Santiago de Compostela).



(a) A > Height / Width. Height has to be increased to reach aspect A

(b) A < Height / Width. Width has to be increased to reach aspect A

Fig. 3 Illustration of aspect ratio adaption.

A geographic focus is generated only if E is not empty. In such case, for each geometry $c \in C$, first a Minimum Bounding Rectangle (MBR) tm that contains both all the objects in E and c is computed (see line 6). Lines 7-16 adapt the shape of tm to the input aspect ratio A. In particular, if the objective aspect ratio A is lower than ratio between height and width of tm (see line 7), then the height of tm must me increased to reach the objective aspect A (see lines 8-10 and Figure 3(a)). Similarly, if the objective aspect ratio A is greater than the ratio between height and width of tm (see line 12), then the width of tm must me increased to reach the objective aspect A (see lines 13-15 and Figure 3(b)).

Lines 17-19 are included to choose the rectangle of minimum area from all the rectangles tm generated for the different context geometries in C. Such minimum rectangle is returned in line 21 together with the relevant subtitle time stamp ($s.time$) and the set of geographic objects E. Each returned triple is pushed into a queue shared between the present component and the *Broadcaster*. Such queue prevents the *Broadcaster* from having to wait for the focus generation, which would cause undesirable delays in the streaming. Such delays might still arise if subtitles are generated on the fly, for example in live shows. Such a real time scenario is a challenging problem for future research which affects the way Algorithm 1 is currently designed, as will be discussed in Section 8.

6 Prototype Implementation

This section describes the current prototype implementation of the main components of the framework, namely the *Broadcaster*, the *Geographic Focus Generator* and the *Geo Context Viewer*.

6.1 TV Show and Application Broadcasting

An important issue to be solved was the simulation of a TV show broadcast in the laboratory. To achieve this, currently available stream generation and modulation hardware and software for PC were used (see the scheme of Figure 4). First, the audio and video content is combined with both the *Geo Context Viewer* HbbTV application signaling and the stream events containing the geographic focus. This process is performed with the aid of the open source MPEG2 transport stream data generator OpenCaster [8], which was installed in a server PC with a Linux operating system. Then a DVB-T signal is generated using a DekTec DTA-2111 [9] modulator installed in the server. An HbbTV receiver Televes ZAS Hbb [10] was used to decode the TV signal to be displayed in a conventional screen through an HDMI input.

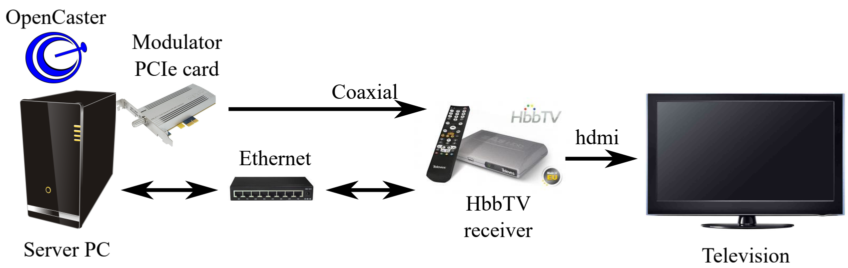


Fig. 4 *Broadcaster* simulation.

6.2 Automatic Generation of Geographic Focus

The core of the automatic generation of geographic focus is the implementation of Algorithm 1 described in Section 5. A key part of this algorithm is the extraction of geographic named entities from the subtitle texts, i.e., function *ExtractGeoEntities* in the above algorithm. To implement this function four alternative technologies were tried, whose impact in the result is shown in the evaluation results presented in Section 7. A first basic implementation

[8] http://www.avalpa.com/the-key-values/15-free-software/33-opencaster
[9] http://www.dektec.com/products/PCIe/DTA-2111/
[10] http://www.televes.es/en/catalogo/producto/hd-dtt-stb-tv-demand-access-zas-hbb

delegated the extraction of geographic entities in an already built geographic entity extraction technology. Two additional implementations combined state of the art general purpose Named Entity Recognition (NER) tools [25, 48] with a geocoding service. The last implementation, which is one of the main contributions of this work, evolved another NER technology [27] to improve its effectiveness when applied to this specific problem.

Implementation based on Yahoo PlaceSpotter (YahooPS): The Yahoo PlaceSpotter geo-entity extractor was directly used to parse each subtitle and extract geographic entities with both names and coordinates. PlaceSpotter is part of the Yahoo BOSS geo services [11], which can be queried with the Yahoo Query Language (YQL). In particular, its functionality may be freely accessed through the table Geo.Placemaker [12].

Implementation based on Stanford Named Entity Recognizer (SNER): The Stanford Named Entity Recognizer (SNER) [25], an open source tool [13], was used to obtain entity names. The SNER implements the linear chain Conditional Random Field (CRF) statistical method for sequence modelling, using the Viberbi algorithm as inference procedure for local structures. The current version is also provided with models that include distributional similarity based features, i.e. models are also supplied with semantic information. However, the use of semantic features requires more memory and makes the system less efficient. Specialized extractors for person, organization and location are supported by SNER. It also supports models for various languages, including Spanish. Our current implementation uses SNER to obtain entity names classified by the tool as locations. Since it is a supervised strategy, the effectiveness of this tool is very much dependent on the specific model available for the required language. The location names obtained by SNER are next submitted to the Nominatim geocoding service to obtain coordinates for them. The final geo-geo disambiguation, i.e. toponym resolution, is then performed by this service. SNER is considered as the state-of-the-art for English language [33].

Implementation based on FreeLing Named Entity Recognizer (FNER): This implementation is similar to the previous one, except that SNER was replaced by the FreeLing Named Entity Recognizer (FNER). FreeLing [48] is a suite of language analysers (written in C++) which includes similar modules than the SNER system and also has tools for other tasks such as phonetic encoding or syntactic analysis. Most of FreeLing modules analyse data in Catalan, Spanish, Portuguese, English, or French (among others). The named entity classifier of FNER is an AdaBoost classifier [12] which achieved the best performance for Spanish language at the CoNLL shared task [58].

[11] `https://developer.yahoo.com/boss/geo/`

[12] `https://developer.yahoo.com/boss/geo/docs/free_YQL.html`

[13] https://nlp.stanford.edu/software/CRF-NER.shtml

Implementation based on Linguakit NER (LNER): In this approach, we have designed and implemented an open-source named entity recognizer (NER), relying on open knowledge resources and linguistic rules, which is multilingual, working for English, Spanish, Portuguese, and Galician languages. The strategy underlying the NER system consists of two steps, namely *identification* and *classification* of the named entities. First, named entities are identified combining finite state automata with language dependent dictionaries and gazetteers. Second, the entities identified in the previous step are now classified as either person, organization, location or miscellaneous.

In order to carry out named entity classification, the system is provided with gazetteers extracted from Freebase[14] and DBpedia[15]), as well as with rules for semantic (geo-non geo) disambiguation. Disambiguation rules are applied on previously identified named entities in order to solve ambiguity (i.e. entities appearing in several gazetteers) and unknown entities (i.e. they are not in any gazetteer). For instance, if an ambiguous or unknown entity appears close to the noun "president" and preposition "of", it is likely to be classified as an organization, however if it co-occurs with the noun "town", it could be a location. Words like "president" and "town" are semantic *triggers* that were automatically extracted and classified by making use of semi-structured information from the Wikipedia, namely infoboxes and category trees.

The first version of our named entity classifier, reported in [26], was implemented as an external module of the linguistic suite FreeLing [48]. The module was taking as input the named entity identifier of FreeLing. Later, we have also implemented an entity identifier based on finite state automata, which has been integrated into the entity classifier to build a NER system [27]. In the present work, we have integrated the NER module into LinguaKit[16], our multilingual linguistic suite which also includes two tasks required by the NER module, namely basic analysis (sentence segmentation, tokenization, and splitting) and PoS tagging. *LNER* is the LinguaKit NER tool we have implemented to be evaluated and compared later with other NER tools in Sec 7. Once the entities are identified and classified by LNER, as in the previous alternatives, Nominatim is used for geocoding and geo-geo disambiguation.

Other NER systems, such as OpenNLP were not considered since according to previous evaluations [27], its effectiveness was poorer than that of SNER, FNER and previous versions of LNER.

6.3 Interactive Visualization of Geographic Context

The visualization of the generated geographic context is done through an HbbTV application hosted in the server that is accessed by the HbbTV receiver through the Ethernet link. An screenshot of the GUI of this application displayed in a television is given in Figure 5. Notice that the A/V content is

[14] http://www.freebase.com
[15] http://www.dbpedia.org
[16] https://github.com/citiususc/Linguakit

displayed in the bottom left side of the GUI whereas the relevant generated map is rendered in the top right side. The present version is using a WMS provided by the Open Street Map (OSM) initiative. It is remarked that, at this stage, it is not an objective of the framework to provide a professional fully functional Geo Context Viewer, which should include both a more friendly GUI and additional pieces of functionality.

Fig. 5 *Geo Context Viewer* prototype demo.

7 Experimental Evaluation

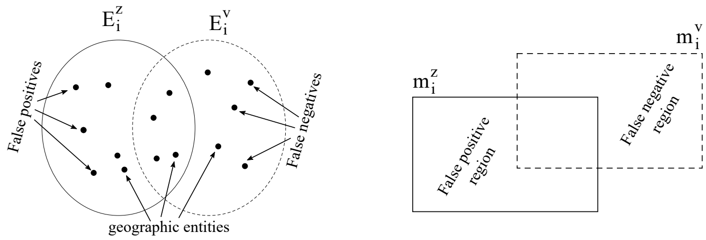
In order to test the effectiveness of the *Geographic Focus Generation* component, an evaluation strategy was devised and applied in a experiment where different implementation alternatives were compared. It is reminded that evaluating the efficiency (result time) of the component is out of the scope of the paper, since real-time requirements are not assumed. Besides, since a professional fully functional Geo Context Viewer is not an objective of this work, an evaluation of its GUI using appropriate Human Computer Interaction (HCI) guidelines is also out of the scope. Further research issues related to real-time and GUI design are discussed in Section 8.

7.1 Evaluation Strategy

Two different aspects of the functionality are evaluated, namely, the set of geographic entities retrieved and the map extent generated for each time instant of the TV show. To achieve this, first a reference dataset with the results expected by some viewer or set of viewers must be constructed. Such dataset consists of a set geographic entities and a map extent, identified by the viewer for each subtitle time instant. A web GIS application was developed to assist test viewers in the construction of reference datasets used for the evaluation. The viewer selects a TV show from a list. Next, some controls enable her to browse the available subtitles. For any subtitle, the viewer may define the extent of the required map, by performing zoom in/out and pan operations. Besides, she may search Nominatim to select the set of geographic entities $\{e_i\}$ to be included in the map, each entity will have a name $e_i.name$ and a location $e_i.point$.

Next, measures of *Precision* and *Recall*, widely known and used in the area of information retrieval, were formalized for both geographic named entities and map extents. In particular, the geographic named entity precision ($EP^{z,v}$) and recall ($ER^{z,v}$) of a system implementation z with respect to the desires of a viewer v are respectively defined by the following two formulas.

$$EP^{z,v} \equiv \frac{\sum_{i=1}^{k} \frac{|E_i^v \cap E_i^z|}{|E_i^z|}}{k} \qquad ER^{z,v} \equiv \frac{\sum_{i=1}^{k} \frac{|E_i^v \cap E_i^z|}{|E_i^v|}}{k} \tag{1}$$



(a) Entity Precision/Recall (b) Map Precision/Recall

Fig. 6 Illustration of Precision and Recall measures.

The number of subtitles in the reference dataset is denoted by k and E_i^v, E_i^z are the sets of geographic entities respectively identified by the viewer v and retrieved by the implementation z for the time instant of subtitle s_i. Notice that $EP^{z,v}$ is computed as a ratio between the entities that are identified by both viewer and system and the entities identified by the system. Therefore it increases as the number of false positives decrease, i.e. when the entities identified by the system and not by the viewer are fewer. This matches the definition of the measure of precision used in the area of Information Retrieval. Similarly, $ER^{z,v}$ is computed as a ratio between the entities that are identified by both viewer and system and the entities identified by the viewer. Therefore, it increases as the number of false negatives decrease, i.e. when the entities identified by the viewer and not by the system are fewer. In other words, it gives a measure of how complete is the response of the system compared with the one of the viewer. Again, this matches the definition of the recall measure used in the area of Information Retrieval. The above measures are graphically illustrated in Figure 6(a).

The map extent precision ($MP^{z,v}$) and recall ($MR^{z,v}$) of a system implementation z with respect to the desires of a viewer v are respectively defined by the following two formulas.

$$MP^{z,v} \equiv \frac{\sum_{i=1}^{k} \frac{area(m_i^v \cap m_i^z)}{area(m_i^z)}}{k} \qquad MR^{z,v} \equiv \frac{\sum_{i=1}^{k} \frac{area(m_i^v \cap m_i^z)}{area(m_i^v)}}{k} \tag{2}$$

Again, k denotes the number of subtitles in the reference dataset and m_i^v, m_i^z are the rectangles defining the map extent respectively identified by viewer v and generated by implementation z for the time instant of subtitle s_i. Notice that $MP^{z,v}$ is computed as the ratio between the area of the intersection of the rectangles of system and viewer and the area of the rectangle retrieved by the system. Therefore, it increases as the area of the region retrieved by the system that is out of the one selected by the viewer decreases (this could be thought as the area of the false positive region). This idea resembles the one behind the measure of precision used in Information Retrieval. Similarly, $MR^{z,v}$ is computed as the ratio between the area of the intersection of the rectangles of system and viewer and the area of the rectangle defined by the viewer. Therefore, it increases as the area of the region identified by the viewer that is out of the rectangle given by the system decreases (this could be thought as the area of the false negative region). This idea resembles the one behind the measure of recall used in Information Retrieval. The above measures are graphically illustrated in Figure 6(b).

To perform an evaluation experiment for a system implementation z, first each test viewer v must use the available dataset construction web GIS application to define sets of entities E_i^v and rectangles m_i^v for each subtitle s_i of the chosen TV show. The implementation of the *Geographic Focus Generation* component z to be evaluated is executed to obtain the sets of entities E_i^z and rectangles m_i^z for each subtitle s_i.

To calculate geographic entity precision and recall, as defined above, it is important to define a test for equality between geographic entities. Ideally, two geographic entities are exactly the same if both their name and their coordinates match. However, due to precision issues exact matching of coordinates use to fail. In the current solution, it is assumed that two horizontal coordinates x_1 and x_2 of two different entities match if their absolute difference $|x_1 - x_2|$ is lower that the 5% of the width of the map extent where they are displayed. A similar assumption is also considered for vertical coordinates y.

Clearly, the solution to the proposed problem is not unique, as different viewers may want different map extends and distinct geographic entities for the same subtitle. Given that, in general, there is not an agreement between viewers, it is not possible to construct a system whose output matches the ideal response of each viewer. Therefore, to be able to get a fair evaluation of a system implementation, a measure of the degree of average agreement between viewers is needed, to establish it as an objective for the system result. To achieve this, a measure of the compatibility of the viewers is obtained for both geographic named entities and map extents, as an arithmetic mean of the geographic named entity and map extent precisions between each pair of viewers. Formally, if V is the set of viewers and v_i, v_j denote viewers of V, then these compatibilities are defined by the formulas.

$$EC^V \equiv \frac{\sum_{i,j} EP^{v_i,v_j}}{|V|^2} \qquad MC^V \equiv \frac{\sum_{i,j} MP^{v_i,v_j}}{|V|^2} \tag{3}$$

Given that the condition $EP^{v_i,v_j} = ER^{v_j,v_i}$ holds for any $v_i, v_j \in V$, then the above compatibilities can alternatively be defined using the recall as follows.

$$EC^V \equiv \frac{\sum_{i,j} ER^{v_i,v_j}}{|V|^2} \qquad MC^V \equiv \frac{\sum_{i,j} MR^{v_i,v_j}}{|V|^2} \tag{4}$$

Broadly speaking, the response of each viewer is evaluated with respect to the response of all the other viewers, in order to get an objective for the system precision and recall. Thus, an implementation will behave perfectly (in precision and recall) if it reaches the mean value obtained by the viewers. Therefore, the final measures of geographic named entity and map extent precision and recall for a given implementation p are obtained by normalizing the measures with the above compatibilities. Notice that this way, we are assuming that the system is not supposed to behave better than one viewer does with respect to the others. Notice also that after the normalization, values of precision and recall above 1 are possible (for implementations that behave better than the average of the viewers in that specific measure). Formally, the final measures for a system implementation z are defined as follows.

$$EP^z \equiv \frac{\sum_i EP^{v_i,z}}{|V| * EC^V} \qquad ER^z \equiv \frac{\sum_i ER^{v_i,z}}{|V| * EC^V} \tag{5}$$

$$MP^z \equiv \frac{\sum_i MP^{v_i,z}}{|V| * MC^V} \qquad MR^z \equiv \frac{\sum_i MR^{v_i,z}}{|V| * MC^V} \tag{6}$$

Based on the above, F-score measures are defined in the standard way.

$$EF^z \equiv 2 * \frac{EP^z * ER^z}{EP^z + ER^z}, \quad MF^z \equiv 2 * \frac{MP^z * MR^z}{MP^z + MR^z}. \tag{7}$$

It is noticed that the above strategy provides measures with well-known semantics of precision and recall for both the geographic entities and the map extent generated by each alternative implementation. The definition of those measures for the evaluation of the map extent generation is an original contribution of the present paper.

Implementation	EP	ER	MP	MR
YahooPS	0.74	0.74	0.83	0.52
SNER	0.94	0.72	0.97	0.42
FNER	0.84	0.75	0.90	0.46
LNER	0.83	0.84	0.87	0.54

Table 1 Values of precision and recall of the three alternative implementations for geographic named entity extraction (EP, ER) and map extent generation (MP, MR).

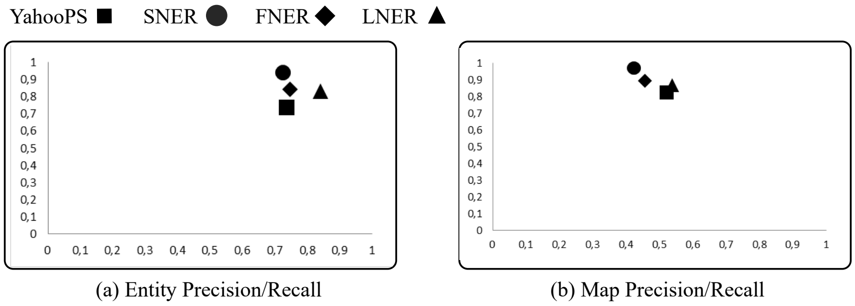


Fig. 7 Graphical representation of precision and recall of the three alternative implementations for (a) geographic named entity extraction and (b) map extent generation.

7.2 Evaluation Experiments

The strategy defined in the previous subsection was tested during the evaluation of alternative implementations of the *Geographic Focus Generation* component with datasets of TV newscasts in Spanish language.

Seven different newscasts of the Spanish public television TVE (around 1100 subtitles each, i.e., near to 8000 subtitles in total) were tagged by seven different viewers. To compute the compatibilities between viewers (EC^V and MC^V), one of the available newscasts was tagged by three different viewers. Using the expressions in Equation 3, the following compatibilities were automatically obtained:

$$EC^V = 0.82, \quad MC^V = 0.73.$$

Therefore, the viewers had a reasonably good compatibility with respect to the geographic named entities that they wanted to see in each map, but also with respect to the region covered by the map. The values of normalized precision and recall for both geographic named entity recognition and map extent generation are given in Table 1 and graphically represented in Figure 7. As it is shown, each of the four alternative implementations, i.e., those based respectively on Yahoo PlaceSpotter (YahooPS), Stanford Named Entity Recognizer (SNER), FreeLing Named Entity Recognizer(FNER) and Linguakit Named Entity Recognizer (LNER), was evaluated with respect to the dataset tagged by the viewers. With respect to geographic named entity recognition (see Figure 7(a)), the solution based on LNER outperforms the one based on YahooPS in both precision and recall. LNER has a precision similar to the one of FNER, but is has better recall. Finally, the solution based on SNER improves the precision at the cost of decreasing recall.

Regarding map extent generation (see Figure 7(b)), the tendency is more or less maintained, with an overall lose of recall in all the implementations with respect to the geographic named entity extraction. Therefore, in general, the different implementations generate maps in the zones required by the viewers, however, the size of those maps is smaller than the one demanded by them.

Implementation	EF	MF
YahooPS	0.72	0.62
SNER	0.78	0.57
FNER	0.77	0.59
LNER	0.82	0.65

Table 2 Values of F-score of the three alternative implementations for geographic named entity extraction (EF) and map extent generation (MF).

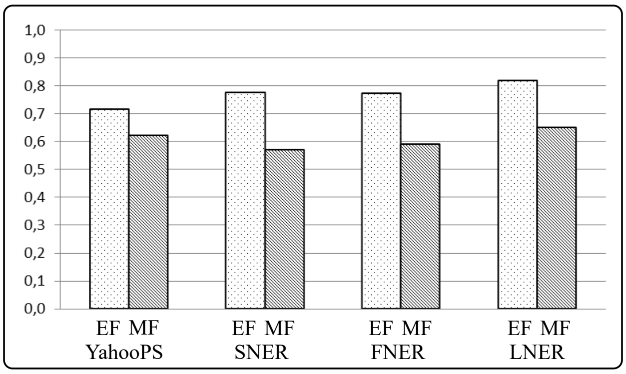


Fig. 8 Graphical representation of F-score of the three alternative implementations for geographic named entity extraction (EF) and map extent generation (MF).

Finally, to have a single value of the effectiveness of each implementation alternative, the F-score is shown in Table 2 for both geographic named entity extraction (EF) and map extent generation (MF) (Those values are also graphically represented in Figure 8). The four implementations have reasonably good results, but the solution based on LNER outperforms the other three in both facets. As it will be argued in the following section, the solution based on LNER is also the one that, in these authors' opinion, has more potential for further improvement.

8 Research Challenges

Based on the experience gained during the developing and evaluation of the initial solution to the problem described in the above sections, some challenges for future research work are now discussed.

Looking at the results of the experiment, there is still room for the improvement. A first issue would be to deal with the short textual context given by each subtitle to solve potential ambiguities. This is not a new issue that has already been faced for example with tweets [40, 51]. Beyond that, more sophisticated solutions might group various subtitles for better disambiguation and to provide a single context map for each group. This behaviour has manually been implemented in [56], where news programs are segmented into 2-5 minutes topics. To achieve an automatic segmentation of the content, already existing approaches for thematic clustering [36] and segmentation [39] of text documents would have to be reviewed. Finally, semantic relationships between terms and entities defined in existing semantic web sources could be exploited to improve system recall, deriving for example the geographic entity ‘United Kingdom’ from term ‘Britannic’. It is estimated that the approach based on rules and linguistic resources followed by LNER has big potential, enabling the integration of named entity recognition (geo / non geo disambiguation) and toponym resolution (geo / geo).

Some specific desirable behaviors of the system cannot be achieved without the use of personalization [15]. Notice that the iterative nature of the HbbTV standard enables the implementation of viewer relevance feedback features. Such functionality would be specially important for the approximation of which is the geographic knowledge of each viewer, to define accordingly appropriate list of context geographic entities. It is reminded that in the current implementation, a single list of context geographic entities is available, common for all the audience. To the best of these authors knowledge, personalization has not been considered in any content enrichment approach for ITV.

Another important challenge that may be required to overcome in some applications is related to real-time requirements for the automatic focus generation. Thus, in live shows were subtitles (of low quality in many cases) are generated from the audio, the automatic focus generation must work in real time and may not use future subtitles for disambiguation. Therefore, the response time of Algorithm 1 for each subtitle must be definitely below one second (one subtitle per second is a typical rate).

Finally, different GUI strategies should be implemented and evaluated from the perspective of HCI, including the use of one and two screens. It is noticed that the present system could easily be migrated to a second screen architecture not based on HbbTV. In that case, technologies like WebSockets would be used to deliver the geographic context to the user device (PC, smartphone, tablet, etc.) synchronously with the A/V newscast content. The effect of presenting additional information in a second screen over the understanding of the content has already been studied [22, 54]. In [22], the participants were positive about the value of companion applications for understanding the content, however they pointed out their concern with the distraction and the "knowing where to look" problem that the second screen introduces. It was also remarked that increasing the interaction with the second screen application disrupted understanding of the TV program. In [54], three different strategies were compared, namely, an only TV approach, a second screen application with manual look-up of contents and the same second screen application with synchronized automatic look-up. The results showed that the second screen had a negative impact in both understanding and program liking. The automatic look-up approach was better than the manual one, since it helped to improve visual attention to the main TV content. It was estimated that the different levels of cognitive load required by the three strategies had an important influence in the observed results. Further work in this line should evaluate different alternatives with one and two screens and with user profiles of different skills both with respect to their geographic knowledge and to their abilities to manipulate devices. Beyond that, other aspects of the interface would have to be taken into account in a more detailed evaluation, including alternative control devices (mobile devices, remote controls, etc.) and alternative levels of user control over the displayed map and its navigation.

9 Conclusions

The design, implementation and evaluation of a real framework for the incorporation of geographic context to TV shows was described. The automatic generation of geographic focus from available subtitles is identified as the key component of the system. The framework is original, since it is the first approach that attempts automatic generation of geographic context for content enrichment in the scope of ITV applications. An evaluation framework was devised to test the effectiveness of different alternative implementations using various geoparsing technologies. Such an evaluation framework, which incorporates both textual and geographic measures of precision and recall is by its own an original contribution of the paper. An F-score of 0.82 is reached by the best implementation for geographic named entity extraction, which is also a main contribution of this work that is based on previous work of the authors. The same implementation reaches also the maximum F-score 0.65 for map extent generation. Finally, research challenges are also identified to work towards a framework that supports both personalized behavior that might adapt to the preferences of each viewer and possibly real-time focus generation. An in-depth evaluation of the ITV interface, considering different interaction alternatives, is of major importance to get a solid product from the base of the present research result. Finally, the framework could also be adapted to be applied to programs whose action takes places in fictitious spaces, such as movies and series. To achieve this, resources like gazetteers that contain fictitious geographic entities with coordinates in the fictitious space would have to be constructed.

References

1. Abreu, J., Almeida, P., Silva, T., Oliveira, R.: Discovering tv contents in a second screen app: Perspectives from portuguese and brazilian markets. Procedia Computer Science **64**, 1240 – 1247 (2015). DOI https://doi.org/10.1016/j.procs.2015.08.508. URL http://www.sciencedirect.com/science/article/pii/S1877050915026435. Conference on ENTERprise Information Systems/International Conference on Project MANagement/Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN / HCist 2015 October 7-9, 2015
2. Abreu, J.F., Almeida, P., Silva, T.: Enriching second-screen experiences with automatic content recognition. In: VI Interactive Digital TV Congress – IV Iberoamerican Conference on Applications and Usability of Interactive TV, Mallorca, Spain (2015)
3. Adelfio, M.D., Samet, H.: Geowhiz: Toponym resolution using common categories. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL'13, pp. 532–535. ACM, New York, NY, USA (2013). DOI 10.1145/2525314.2525321. URL http://doi.acm.org/10.1145/2525314.2525321
4. Aguilar, M., Pamplona, P., Fernández, S.: Tv-ring and immersiatv: Present and future of television. In: M.J. Abásolo, F.J. Perales, A. Bibiloni (eds.) Applications and Usability of Interactive TV, pp. 9–14. Springer International Publishing, Cham (2016)
5. Amitay, E., Har'El, N., Sivan, R., Soffer, A.: Web-a-where: Geotagging web content. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '04, pp. 273–280. ACM, New York, NY, USA (2004). DOI 10.1145/1008992.1009040. URL http://doi.acm.org/10.1145/1008992.1009040
6. Angel, A., Lontou, C., Pfoser, D., Efentakis, A.: Qualitative geocoding of persistent web pages. In: Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '08, pp. 20:1–20:10. ACM, New York, NY, USA (2008). DOI 10.1145/1463434.1463460. URL http://doi.acm.org/10.1145/1463434.1463460
7. Anstead, E., Benford, S., Houghton, R.J.: Many-screen viewing: Evaluating an olympics companion application. In: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, TVX '14, pp. 103–110. ACM, New York, NY, USA (2014). DOI 10.1145/2602299.2602304. URL http://doi.acm.org/10.1145/2602299.2602304
8. Athanasiadis, E., Mitropoulos, S.: A distributed platform for personalized advertising in digital interactive tv environments. Journal of Systems and Software **83**(8), 1453 – 1469 (2010). DOI https://doi.org/10.1016/j.jss.2010.02.040. URL http://www.sciencedirect.com/science/article/pii/S0164121210000683. Performance Evaluation and Optimization of Ubiquitous Computing and Networked Systems
9. Bao, J., Mokbel, M.F.: Georank: An efficient location-aware news feed ranking system. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL'13, pp. 184–193. ACM, New York, NY, USA (2013). DOI 10.1145/2525314.2525336. URL http://doi.acm.org/10.1145/2525314.2525336
10. Van den Broeck, W., Pamplona, P., Fernandez Langa, S.: Developing a multi-device immersive tv-experience: User expectations and practices. In: Proceedings of the 2Nd International Workshop on Multimedia Alternate Realities, AltMM '17, pp. 25–30. ACM, New York, NY, USA (2017). DOI 10.1145/3132361.3132362. URL http://doi.acm.org/10.1145/3132361.3132362
11. Buscaldi, D.: Approaches to disambiguating toponyms. SIGSPATIAL Special **3**(2), 16–19 (2011). DOI 10.1145/2047296.2047300. URL http://doi.acm.org/10.1145/2047296.2047300
12. Carreras, X., Marquez, L., Padró, L., Padró, M.: Named entity extraction using adaboost. In: COLING-02 proceedings of the 6th Conference on Natural Language Learning (2002)
13. Cesar, P., Bulterman, D.C., Geerts, D., Jansen, J., Knoche, H., Seager, W.: Enhancing social sharing of videos: Fragment, annotate, enrich, and share. In: Proceedings of the

16th ACM International Conference on Multimedia, MM '08, pp. 11–20. ACM, New York, NY, USA (2008). DOI 10.1145/1459359.1459362. URL http://doi.acm.org/10.1145/1459359.1459362
14. Cesar, P., Geerts, D.: Past, present, and future of social tv: A categorization. In: 2011 IEEE Consumer Communications and Networking Conference (CCNC), pp. 347–351 (2011). DOI 10.1109/CCNC.2011.5766487
15. Chorianopoulos, K.: Personalized and mobile digital tv applications. Multimedia Tools and Applications **36**(1), 1–10 (2008). DOI 10.1007/s11042-006-0081-8. URL https://doi.org/10.1007/s11042-006-0081-8
16. Coelho, J., Rito, F., Duarte, C.: "you, me & tv" — fighting social isolation of older adults with facebook, tv and multimodality. International Journal of Human-Computer Studies **98**, 38 – 50 (2017). DOI https://doi.org/10.1016/j.ijhcs.2016.09.015. URL http://www.sciencedirect.com/science/article/pii/S1071581916301276
17. Compton, R., Keegan, M.S., Xu, J.: Inferring the geographic focus of online documents from social media sharing patterns. CoRR **abs/1406.2392** (2014). URL http://arxiv.org/abs/1406.2392
18. Costa, D., Carriço, L., Duarte, C.: The differences in accessibility of tv and desktop web applications from the perspective of automated evaluation. Procedia Computer Science **67**, 388 – 396 (2015). DOI https://doi.org/10.1016/j.procs.2015.09.283. URL http://www.sciencedirect.com/science/article/pii/S1877050915031294. Proceedings of the 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion
19. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V., Aswani, N., Roberts, I., Gorrell, G., Funk, A., Roberts, A., Damljanovic, D., Heitz, T., Greenwood, M.A., Saggion, H., Petrak, J., Li, Y., Petersa, W., Derczynski, L.: Developing Language Processing Components with GATE Version 8. University of Sheffield Department of Computer Science (2014)
20. D'Ignazio, C., Bhargava, R., Zuckerman, E., Beck, L.: Cliff-clavin: Determining geographic focus for news articles. In: NewsKDD workshop, 201h ACM SIGKDD conference on Knowledge Discovery and Data Mining (2014)
21. Ding, J., Gravano, L., Shivakumar, N.: Computing geographical scopes of web resources. In: Proceedings of the 26th International Conference on Very Large Data Bases, VLDB '00, pp. 545–556. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2000). URL http://dl.acm.org/citation.cfm?id=645926.672013
22. Dowell, J., Malacria, S., Kim, H., Anstead, E.: Companion apps for information-rich television programmes: representation and interaction. Personal and Ubiquitous Computing **19**(7), 1215–1228 (2015). DOI 10.1007/s00779-015-0867-7. URL https://doi.org/10.1007/s00779-015-0867-7
23. Ducheneaut, N., Moore, R.J., Oehlberg, L., Thornton, J.D., Nickell, E.: Social tv: Designing for distributed, sociable television viewing. International Journal of Human–Computer Interaction **24**(2), 136–154 (2008). DOI 10.1080/10447310701821426
24. Epelde, G., Valencia, X., Carrasco, E., Posada, J., Abascal, J., Diaz-Orueta, U., Zinnikus, I., Husodo-Schulz, C.: Providing universally accessible interactive services through tv sets: implementation and validation with elderly users. Multimedia Tools and Applications **67**(2), 497–528 (2013). DOI 10.1007/s11042-011-0949-0. URL https://doi.org/10.1007/s11042-011-0949-0
25. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL '05, pp. 363–370. Association for Computational Linguistics, Stroudsburg, PA, USA (2005). DOI 10.3115/1219840.1219885. URL http://dx.doi.org/10.3115/1219840.1219885
26. Gamallo, P., Garcia, M.: A resource-based method for named entity extraction and classification. LNCS **7026**, 610–623 (2011)
27. Garcia, M., Gamallo, P.: Yet Another Suite of Multilingual NLP Tools, pp. 65–75. Springer International Publishing, Cham (2015). DOI 10.1007/978-3-319-27653-3_7. URL https://doi.org/10.1007/978-3-319-27653-3_7
28. Gelernter, J., Balaji, S.: An algorithm for local geoparsing of microtext. Geoinformatica **17**(4), 635–667 (2013). DOI 10.1007/s10707-012-0173-8. URL http://dx.doi.org/10.1007/s10707-012-0173-8

29. Goldberg, D.W., Wilson, J.P., Knoblock, C.A.: Extracting geographic features from the internet to automatically build detailed regional gazetteers. Int. J. Geogr. Inf. Sci. **23**(1), 93–128 (2009). DOI 10.1080/13658810802577262. URL http://dx.doi.org/10.1080/13658810802577262
30. Goodchild, M.F., Hill, L.L.: Introduction to digital gazetteer research. Int. J. Geogr. Inf. Sci. **22**(10), 1039–1044 (2008). DOI 10.1080/13658810701850497. URL http://dx.doi.org/10.1080/13658810701850497
31. Guna, J., Stojmenova, E., Kos, A., Pogačnik, M.: The tv-web project - combining internet and television – lessons learnt from the user experience studies. Multimedia Tools and Applications **76**(20), 20,377–20,408 (2017). DOI 10.1007/s11042-016-3243-3. URL https://doi.org/10.1007/s11042-016-3243-3
32. Guna, J., Stojmenova-Duh, E., Pogačnik, M.: Users' viewpoint of usability and user experience testing procedure - gaining methodological insights in a case of an interactive hbbtv application. Multimedia Tools and Applications **76**(15), 16,125–16,143 (2017). DOI 10.1007/s11042-016-3898-9. URL https://doi.org/10.1007/s11042-016-3898-9
33. Jiang, R., Banchs, R.E., Li, H.: Evaluating and combining name entity recognition systems. In: Proceedings of the Sixth Named Entity Workshop, NEWS@ACL 2016, Berlin, Germany, August 12, 2016, pp. 21–27 (2016). DOI 10.18653/v1/W16-2703. URL https://doi.org/10.18653/v1/W16-2703
34. Jones, C.B., Purves, R.S.: Geographical information retrieval. Int. J. Geogr. Inf. Sci. **22**(3), 219–228 (2008). DOI 10.1080/13658810701626343. URL http://dx.doi.org/10.1080/13658810701626343
35. Kim, J., Ahn, S.J.G., Kwon, E.S., Reid, L.N.: Tv advertising engagement as a state of immersion and presence. Journal of Business Research **76**, 67 – 76 (2017). DOI https://doi.org/10.1016/j.jbusres.2017.03.001. URL http://www.sciencedirect.com/science/article/pii/S0148296317300887
36. Kim, S., Wilbur, W.J.: Thematic clustering of text documents using an em-based approach. Journal of Biomedical Semantics **3**(3), 1–13 (2012). DOI 10.1186/2041-1480-3-S3-S6
37. Kiryakov, A., Popov, B., Terziev, I., Manov, D., Ognyanoff, D.: Semantic annotation, indexing, and retrieval. Web Semant. **2**(1), 49–79 (2004). DOI 10.1016/j.websem.2004.07.005. URL http://dx.doi.org/10.1016/j.websem.2004.07.005
38. Krämer, N.C., Winter, S., Benninghoff, B., Gallus, C.: How "social" is social tv? the influence of social motives and expected outcomes on the usage of social tv applications. Computers in Human Behavior **51**, 255 – 262 (2015). DOI https://doi.org/10.1016/j.chb.2015.05.005. URL http://www.sciencedirect.com/science/article/pii/S0747563215003672
39. Lamprier, S., Amghar, T., Levrat, B., Saubion, F.: Using an evolving thematic clustering in a text segmentation process. Journal of Universal Computer Science **14**(2), 178–192 (2008)
40. Li, C., Weng, J., He, Q., Yao, Y., Datta, A., Sun, A., Lee, B.S.: Twiner: Named entity recognition in targeted twitter stream. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '12, pp. 721–730. ACM, New York, NY, USA (2012). DOI 10.1145/2348283.2348380
41. Li, Z., Wang, M., Liu, J., Xu, C., Lu, H.: News contextualization with geographic and visual information. In: Proceedings of the 19th ACM International Conference on Multimedia, MM '11, pp. 133–142. ACM, New York, NY, USA (2011). DOI 10.1145/2072298.2072317. URL http://doi.acm.org/10.1145/2072298.2072317
42. Lieberman, M.D., Samet, H.: Multifaceted toponym recognition for streaming news. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11, pp. 843–852. ACM, New York, NY, USA (2011). DOI 10.1145/2009916.2010029. URL http://doi.acm.org/10.1145/2009916.2010029
43. Lieberman, M.D., Samet, H., Sankaranarayanan, J., Sperling, J.: Steward: Architecture of a spatio-textual search engine. In: Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems, GIS '07, pp. 25:1–25:8. ACM, New York, NY, USA (2007). DOI 10.1145/1341012.1341045. URL http://doi.acm.org/10.1145/1341012.1341045

44. Luo, J., Joshi, D., Yu, J., Gallagher, A.: Geotagging in multimedia and computer vision—a survey. Multimedia Tools and Applications **51**(1), 187–211 (2011). DOI 10.1007/s11042-010-0623-y. URL https://doi.org/10.1007/s11042-010-0623-y
45. Murray, J., Goldenberg, S., Agarwal, K., Chakravorty, T., Cutrell, J., Doris-Down, A., Kothandaraman, H.: Story-map: Ipad companion for long form tv narratives. In: Proceedings of the 10th European Conference on Interactive TV and Video, EuroITV '12, pp. 223–226. ACM, New York, NY, USA (2012). DOI 10.1145/2325616.2325659. URL http://doi.acm.org/10.1145/2325616.2325659
46. Nandakumar, A., Murray, J.: Companion apps for long arc tv series: Supporting new viewers in complex storyworlds with tightly synchronized context-sensitive annotations. In: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, TVX '14, pp. 3–10. ACM, New York, NY, USA (2014). DOI 10.1145/2602299.2602317. URL http://doi.acm.org/10.1145/2602299.2602317
47. Nitta, N., Babaguchi, N., Kitahashi, T.: Generating semantic descriptions of broadcasted sports videos based on structures of sports games and tv programs. Multimedia Tools and Applications **25**(1), 59–83 (2005). DOI 10.1023/B:MTAP.0000046382.62218.e1. URL https://doi.org/10.1023/B:MTAP.0000046382.62218.e1
48. Padró, L., Stanilovsky, E.: Freeling 3.0: Towards wider multilinguality. In: Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey (2012)
49. Pasley, R.C., Clough, P.D., Sanderson, M.: Geo-tagging for imprecise regions of different sizes. In: Proceedings of the 4th ACM Workshop on Geographical Information Retrieval, GIR '07, pp. 77–82. ACM, New York, NY, USA (2007). DOI 10.1145/1316948.1316969. URL http://doi.acm.org/10.1145/1316948.1316969
50. Purves, R.S., Clough, P., Jones, C.B., Arampatzis, A., Bucher, B., Finch, D., Fu, G., Joho, H., Syed, A.K., Vaid, S., Yang, B.: The design and implementation of spirit: A spatially aware search engine for information retrieval on the internet. Int. J. Geogr. Inf. Sci. **21**(7), 717–745 (2007). DOI 10.1080/13658810601169840. URL http://dx.doi.org/10.1080/13658810601169840
51. Ritter, A., Clark, S., Mausam, Etzioni, O.: Named entity recognition in tweets: An experimental study. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11, pp. 1524–1534. Association for Computational Linguistics, Stroudsburg, PA, USA (2011)
52. Samet, H., Sankaranarayanan, J., Lieberman, M.D., Adelfio, M.D., Fruin, B.C., Lotkowski, J.M., Panozzo, D., Sperling, J., Teitler, B.E.: Reading news with maps by exploiting spatial synonyms. Commun. ACM **57**(10), 64–77 (2014). DOI 10.1145/2629572. URL http://doi.acm.org/10.1145/2629572
53. Sankaranarayanan, J., Samet, H., Teitler, B.E., Lieberman, M.D., Sperling, J.: Twitterstand: News in tweets. In: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '09, pp. 42–51. ACM, New York, NY, USA (2009). DOI 10.1145/1653771.1653781. URL http://doi.acm.org/10.1145/1653771.1653781
54. Schaap, G., Kleemans, M., Cauwenberge, A.V.: Second screening for news: Effects of presentation on information processing and program liking. Computers in Human Behavior **84**, 76 – 85 (2018). DOI https://doi.org/10.1016/j.chb.2018.02.006. URL http://www.sciencedirect.com/science/article/pii/S0747563218300645
55. Silva, P., Amer, Y., Tsikerdanos, W., Shedd, J., Restrepo, I., Murray, J.: A game of thrones companion: Orienting viewers to complex storyworlds via synchronized visualizations. In: Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, TVX '15, pp. 167–172. ACM, New York, NY, USA (2015). DOI 10.1145/2745197.2755519. URL http://doi.acm.org/10.1145/2745197.2755519
56. Strzebkowski, R., Bartoli, R., Spielvogel, S., Schmidt, D.: Hbbtv-based augmented information television with segment-linked related content on tv and 2nd screen. In: Proceedings of the 2nd International Workshop on Interactive Content Consumption at the ACM International Conference on Interactive Experiences for TV and Online Video (TVX2014) (2014). URL https://wsicc.net/2014/
57. Teitler, B.E., Lieberman, M.D., Panozzo, D., Sankaranarayanan, J., Samet, H., Sperling, J.: Newsstand: A new view on news. In: Proceedings of the 16th ACM SIGSPATIAL

International Conference on Advances in Geographic Information Systems, GIS '08, pp. 18:1–18:10. ACM, New York, NY, USA (2008). DOI 10.1145/1463434.1463458. URL http://doi.acm.org/10.1145/1463434.1463458

58. Tjong Kim Sang, E.F.: Introduction to the conll-2002 shared task: Language-independent named entity recognition. In: Proceedings of the 6th Conference on Natural Language Learning - Volume 20, COLING-02, pp. 1–4. Association for Computational Linguistics, Stroudsburg, PA, USA (2002). DOI 10.3115/1118853.1118877. URL https://doi.org/10.3115/1118853.1118877
59. Tsekleves, E., Whitham, R., Kondo, K., Hill, A.: Investigating media use and the television user experience in the home. Entertainment Computing **2**(3), 151 – 161 (2011). DOI https://doi.org/10.1016/j.entcom.2011.02.002. URL http://www.sciencedirect.com/science/article/pii/S1875952111000061. User experiences in the new TV landscape
60. Vanattenhoven, J., Geerts, D.: Social experiences within the home using second screen tv applications. Multimedia Tools and Applications **76**(4), 5661–5689 (2017). DOI 10.1007/s11042-016-3646-1. URL https://doi.org/10.1007/s11042-016-3646-1
61. Vasardani, M., Winter, S., Richter, K.F.: Locating place names from place descriptions. International Journal of Geographical Information Science **27**(12), 2509–2532 (2013). DOI 10.1080/13658816.2013.785550
62. Vidal, J.C., Lama, M., Otero-García, E., Bugarín, A.: Graph-based semantic annotation for enriching educational content with linked data. Knowledge-Based Systems **55**, 29 – 42 (2014). DOI http://dx.doi.org/10.1016/j.knosys.2013.10.007. URL http://www.sciencedirect.com/science/article/pii/S0950705113003183
63. Wang, C., Xie, X., Wang, L., Lu, Y., Ma, W.Y.: Web resource geographic location classification and detection. In: Special Interest Tracks and Posters of the 14th International Conference on World Wide Web, WWW '05, pp. 1138–1139. ACM, New York, NY, USA (2005). DOI 10.1145/1062745.1062907. URL http://doi.acm.org/10.1145/1062745.1062907