XLike
FP7-ICT-2011.4.2
Contract no.: 288342
www.xlike.org

XLike

Deliverable D5.2.1

Early Information Visualization Prototype

Editor:	Zhixing Li, THU.
Author(s):	Zhixing Li, THU; Peng Zhang THU, Juanzi Li, THU; Blaž Fortuna, JSI; Janez Brank, JSI; Andrej Muhic, JSI
Deliverable Nature:	P
Dissemination Level: (Confidentiality)[1]	PU
Contractual Delivery Date:	M12
Actual Delivery Date:	
Suggested Readers:	XLike project partners
Version:	1.0
Keywords:	

[1] Please indicate the dissemination level using one of the following codes:
• **PU =** Public • **PP =** Restricted to other programme participants (including the Commission Services) • **RE =** Restricted to a group specified by the consortium (including the Commission Services) • **CO =** Confidential, only for members of the consortium (including the Commission Services) • **Restreint UE =** Classified with the classification level "Restreint UE" according to Commission Decision 2001/844 and amendments • **Confidentiel UE =** Classified with the mention of the classification level "Confidentiel UE" according to Commission Decision 2001/844 and amendments • **Secret UE =** Classified with the mention of the classification level "Secret UE" according to Commission Decision 2001/844 and amendments

Disclaimer

Full Project Title:	XLike – Cross-lingual Knowledge Extraction
Short Project Title:	XLike
Number and Title of Work package:	WP5
Document Title:	D5.2.1 - Early Information Visualization Prototype
Editor (Name, Affiliation)	Zhixing Li, THU
Work package Leader (Name, affiliation)	Juanzi Li, THU
Estimation of PM spent on the deliverable:	15

Executive Summary

This document presents a visualization prototype which displays multilingual real-time news from different publishers in different languages. The main outcome of this deliverable is a set of real time data services and a visualization component whose input is well-processed structured news data. This report contains two parts: the definition of data format and services and the design of front-end component.

The prototype described in this document depends on former parts of the project such as WP2, WP3 and WP4. It is assumed here that the related language processing pipeline is well prepared. It should also be noticed that this prototype is use cases oriented to meet the requirements proposed by STA and BLP.

The developed prototype is available at http://sandbox-xlike.isoco.com/portal/ .

Table of Contents

List of Tables

List of Figures

Abbreviations

API	Application Programming Interface

Definitions

Article	An instance of a news report.
Entity	*Person*, *Organization* or *Location* contained in the *Title* or *Content* of an *Article*
Story	A collection of *Articles* which talk about the same topic or event.

1 Introduction

This deliverable (D.5.2.1) is a prototype for real time news visualization. The input of this component is the structured data extracted from news reports in former work packages. The visualization prototype will display news in different dimensions, which can be classified into two categories. The first category is metadata dimensions containing *language*, *time*, *location*, *publisher* and so on. The second category is semantic dimensions containing *entity*, *keyword*, *story* and so on.

The prototype relies on the multilingual and cross-lingual web services provided by other workpackages. Namely, it is built on top of Newsfeed feed of news articles [D1.3.1]. Each news article is sent through multilingual linguistic pipeline [D2.1.1] for basic linguistic processing (language identification, tokenization, part-of-speech, named entities), cross-lingual annotation [D3.1.1] and cross-lingual document linking [D4.1.1]. Details on this can be seen in the corresponding deliverables, and details on integration can be seen in [D6.2.1].

The visualization prototype consists of two components: data services and visualization component. The data services process news articles into structured data with techniques developed in former work packages and the visualization component provides an interface to display these structured data in a simply and clear way. Meanwhile, the visualization component will also provide some interactive functions. In section 2, the use cases is introduced. Section 3 discusses the data formats and services. The detail of visualization component is introduced in Section 4. In Section 5, a conclusion will be made. Some external components and tools can be found in annex.

2 Requirements of Use Cases

This section introduces the use cases proposed by Bloomberg and STA, which are being tackled within XLike project.

2.1 Bloomberg Use Case

I. Entity tracking

The goal of entity tracking in the Bloomberg use case is to maintain an up-to-date list of relevant company news, preferably from their home markets. The task can be roughly defined in two steps: (1) detect mentions of the entity in the multi-lingual news stream, and (2) determine which four are most suitable to be displayed on the company news profile page. The first step requires entity extraction from multi-lingual stream, while the second step requires integration and summarization across all languages with relevant articles.

II. Related or Relevant Articles tracking

The related or relevant article in Bloomberg use case is assembling an article list custom fitted for the specific user, based on his local markets and his/her history. Such list of articles can include external mainstream source from his/her local language.

2.2 STA Use Case

I. Article tracking

The goal of article tracking is to find articles same or similar to a given article. The similar articles include those which are published in different languages.

II. Topic and Entity tracking

Entity tracking can be seen as a filter applied on news reports. The reports which match, contain or are related to the given entity will be retained and visualized in the graphic interface. The visualization component should display the geospatial distribution of the matched reports.

2.3 Summary of Use Cases

Generally speaking, the use cases of STA and BLP are similar. First, both use cases require functionality to track articles by entities. Entity tracking technique is mature in monolingual environment while the situation is different in cross-lingual environment. In XLike, entities in different languages should be linked or aligned before tracking by **T3.1**. Second, both user cases require functions of tracking articles by related articles. Considering the publishers are in different languages, this part requires the cross-lingual similarity calculation which is the outcome of **T4.1**.

Besides tracking by entities (or topics) and related articles, the visualization prototype also provides a function of tracking articles by stories. A story is a set of articles which talk about a same event. In total, the visualization component will provide functions of tracking articles in the three semantic dimensions: entities, stories and publishers.

3 Data Formats and Data Services

The visualization prototype contains two components: data services serving as a backend, and the visualization component frontend. The data services are responsible for converting the real-time raw news stream into structured data. In this section, the data formats and data services used in visualization prototype are introduced. The visualization component will be introduced in Section 4.

Data services revolve around three core objects: article, story and entity. In Section 3.1, we define the formal formats of these three objects and show their relationships. In Section 3.2, six web services are introduced together with their parameters and the output format. This is followed by Section 3.3 detailing the algorithms used for story identification, since this functionality was not covered by any other workpackage. We finish with Section 3.4 describing how cross-lingual document linking [D4.1.1] was integrated into the prototype.

3.1 Data Formats

This section defines the format of the three core objects: article, story and entity. The actual formats used in the interface between data services and visualization component form a subset in order to save network traffic, when possible.

An ***article*** corresponds to one news report, and is the basic unit returned by the Newsfeed [D1.3.1]. Typically an article would correspond to a webpage, from which it was collected. However, for special feeds (e.g. STA agency feed) this is not the case. All the fields, used to describe an article, are presented in Table 1.

Table 1. The formal format of article

Attribute	Data type	Description
id	Int	The ID of the article
url	String	The url of the article
title	String	The title of the article
body	String	The content of the article, only text.
publisher	String	The source of the article
country	String	Where the article takes place, currently it is the country of publisher.
city	String	Where the article takes place, currently it is the city of publisher.
location	<double, double>[2]	The longitude and latitude of ***city***
language	String	The language of the article, follow *ISO 639-2* standard
time	String	yyyy-MM-ddThh:mm:ss, the publishing time of article, or crawling time if publishing time is unavailable.
entities	{<entity uri, int>}	A list of entities occurs in this article, each associated with its frequency in the ***title*** and ***body***
story	<storyid, double>	Which story does this article belong to and the relevance between this article and the story.
articles	{<article id, double>}	The most similar articles of this article, each associated with a similarity.

[2] <> refers a tuple, in which the order, data type and size of elements are predefined. {} refers to an array which is an infinite collection of elements which are of the same data type. Note that the element of a tuple can be an array and the element of an array can be a tuple too.

A ***story*** corresponds to a collection of ***articles*** which talk about the same topic or event. Technically, a story is a cluster of news articles, identified using the approach presented in Section 3.3. All fields, used to describe a story, are presented in Table 2.

Table 2. The formal format of entity

Attribute	Data type	Description
uri	String	The URI of the entity as identified by annotation service; in the first prototype the URIs correspond to either Wikipedia links, or links to DMoz categories.
type	String	The type of entity in enricher(a system developed by JSI)
label	{<language code, String>}	The string to be displayed of the entity in xlike languages.
keys	{<String, String>}	The key-value pairs of the entity. For example, if the entity refers to a person, the keys may contain pairs such as <birthday, 1990-01-01>, <gender, female> and so on.

An ***entity*** is a special word or phrase representing a person, a location, an organization, or some abstract concept. All fields, used to describe an entity, are presented in Table 3.

Table 3. The formal format of story

Attribute	Data type	Description
id	String	The id of story, case sensitive.
label	String	A text to be displayed of the story. Currently it is the title of the most centroid article of the story.
abstract	String	Several summary sentences of the story. Currently it contains two or three sentences of the most centroid article of the story.

3.2 Data Services

This section introduces the data services API[3] used by the visualization component. The API is implemented in form of simple web services using JSon format. The parameters and outputs of each service will be list in detail in the following subsections.

Data services are implemented on top of data infrastructure developed within [D1.3.1]. It is based on NewsMiner system, which provides an index of over last month of articles, collected by Newsfeed. NewsMiner is updated in batches of 100 articles, which in practice means around once per minute.

There are two versions of the API: a public one based on Newsfeed crawl, and a private one, used in the STA use-case, which contains private agency feeds. Considering that the privacy problem has no impact on the functionality of data services, it won't be discussed here. Instead, in deployment, there will be two separate data service sets with different access control strategies, one for public access and the other for private access.

Data services overview

This section lists data services API defining the visualization prototype backend. Both public data services and private data services follow the same standards.

[3] http://newsfeed.ijs.si/xlike/

Table 4. Data services overview

API	Key parameter	Output	Description
stories	none	A story list	Return the most popular stories(maximum 40)
entities	none	An entities list	Return the most popular entities(maximum 60)
entity	*uri=entityUri*	An entity A list of articles A list of stories A list of entities A list of STA entities A list of keywords A list of timestamps	Articles in article list should contain the specified entity. Stories in story list should contain at least one article in article list. Entities in entity list should occur in at least one article in article list, ordered by frequency. STA entities in STA entity list should occur in at least one article in article list, ordered by frequency. Keywords in keyword list should occur in at least one article in article list, ordered by frequency. Timestamps in timestamp list are 2-tuples in which the first item is the timestamp and the second item the count of articles whose time equals to this timestamp.
article	*id=articleID*	An article A list of articles A list of stories A list of entities A list of STA entities A list of keywords A list of timestamps	Articles in article list are similar articles from other languages as identified using cross-lingual document linking. Stories in story list are similar stories from other languages as identified using cross-lingual document linking. The descriptions of the rest of lists are the same as in API ***entity***.
story	*id=storyID*	A story A list of articles A list of entities A list of STA entities A list of keywords A list of timestamps	Articles in article list are contained by the specified story. Stories in story list are similar stories from other languages as identified using cross-lingual document linking The descriptions of the rest of lists are the same as in API ***entity***.
search	*q=keyword*	A list of articles A list of stories A list of entities A list of STA entities A list of keywords A list of timestamps	Articles in article list should contain the specified keyword. The descriptions of the rest of lists are the same as in API ***entity***.

From table 4, it can be seen that the outputs contain two kinds of data. The first kind is single object such as an **article**, an **entity** or a **story**. The second kind is lists of these objects. In section 3.1, formal formats for these objects are defined. However, in the actual implementation only their subset was used. For example, a list of articles returned by *search* API does not need full content of each article to be provided. On the other

hand, when retrieving one particular article using *article* API, this field is provided. Table 5 illustrates the selected subset used by each specific web service. These formats will be used by both data services and the visualization component.

Table 5. Data formats used in data services. The size of article list is defined by parameter *pagesize* (see Table 6), the size of STA entity list, keyword list follows the size of entity list (maximum 60).

Data	Practical Format	Related APIs
article	<id, url, title, body, publisher, country, city, location, language, time>	***article***
article list	<count, {<id, url, title, publisher, country, city, location, language, time >}>	***article, story, entity, search***
entity	<uri, type, {<language, label>}, {<key, value>}>	***entity***
entity list	{<uri, {language, label}, frequency>}	***article, story, entity, search, entities***
story	<id, label, abstract, count(the number of articles in this story)>	***story***
story list	{<id, label>}	***article, entity, search, stories***
keyword list	{<keyword, count>}	***article, entity, story, search***
timestamp list	{<time, count>}	***article, entity, story, search***
STA entity list	{<uri, label, frequency>}	***article, entity, story, search***

Supporting parameters

This section provides a list of miscellaneous parameters, which can be provided when calling data services API. These supporting parameters are not mandatory, but can be used for a more customizable user experience and more accurate tracking. Table 6 lists the supporting parameters and their descriptions.

Table 6. Supporting parameters

Name	Values	Type	Description
pagesize	20(default), 50, 100, 200, 500	Single value	How many articles will be returned
page	0(default), 1, …	Single value	The page offset of returned articles
lang	eng, ger, spa, chi, slv, cat, oth	Multiple values	Specify the language of articles; default value is empty which means all languages are permit.
group	general, politics, sports, culture , economy, foreignaffairs	Multiple values	Specify the groups of STA entities, works only in private API. Default value is empty which means all groups are permit. In public API, the value of this parameter is fixed as “general”.
ts	1h, 2h, 12h, 1d(default), 3d, 7d, 4w.	Single value	The time span of the returned articles. “h” for hour, “d” for day and “w” for week.
country	Country/region strings	Multiple values (up to 5)	Specify the country or region of articles, default value is empty which means all countries are permit.

For different APIs, the available supporting parameters are different. Mainly, we divided all APIs into two categories according to their outputs. The first category is **simple API** including ***stories*** and ***entities*** whose outputs are single lists. The second category is **complex API** including ***article***, ***entity***, ***story*** and ***search***. Table 7 shows the availabilities of supporting parameters over different APIs.

Table 7. The availabilities of supporting parameters

		pagesize	*page*	*ts*	*lang*	*group*	*country*
Simple APIs	***stories***	-	-	+	-	-	-
	entities	-	-	+	-	-	-
Complex APIs	***entity***	+	+	+	+	+	+
	article	+	+	+	+	+	+
	story	+	+	+	+	+	+
	search	+	+	+	+	+	+

The main difference between simple APIs and complex APIs is that the outputs of complex APIs contain a list of articles which derives the story list, entity list, STA list, entity list and timestamp list. The parameters *pagesize, page, lang, group* and *country* are available only on articles, thus they are unavailable on simple APIs.

All these APIs will be called by visualization component via URL in the form *baseURL/APIname?[key parameters][supporting parameters]*. A typical example is:

http://newsfeed.ijs.si/xlike/search?q=obama&page=0&pagesize=100&ts=1d&lang=eng&lang=chi&group=general&country=slovenia
A request like this will call the ***search*** API which will return the first **100** articles containing keyword "***obama***" reported in ***the last day*** in ***English*** language and ***Chinese*** language by ***Slovenia*** publishers. A story list, an entity list, a keyword list and a timestamp list will be returned according to the matched articles. A list of STA entities belonging to ***general*** group will be returned too according to their occurrences in the matched articles.

3.3 NewsCluster Service for Story Identification

Visualization prototype relies on the multilingual and cross-lingual services provided by other workpackages (annotation, cross-lingual document linking). In addition, the prototype required identification and grouping of articles into stories or clusters.

NewsCluster is a web service for microclustering a stream of news documents, based largely on the approach of C. Aggarwal *et al.* [AY06, AY10, AHWY03]. The service maintains a set of documents partitioned into a number of (relatively small) clusters. Old documents are periodically discarded, thereby keeping the cluster structure focused on the current state of the data stream.

Architecture

The NewsCluster web service runs as a simple HTTP server. It accepts incoming HTTP requests containing the text of new documents that need to be added to the clustering, and it reports the resulting (document ID, cluster ID) pairs to a set of zero or more "listeners", i.e. other web services whose URLs are passed to the NewsCluster service as command line parameters. The NewsCluster service periodically saves its state to disk and performs other maintenance operations, such as discarding old documents and deleting clusters that fall below a minimum size threshold. The architecture diagram is depicted on Figure 1.

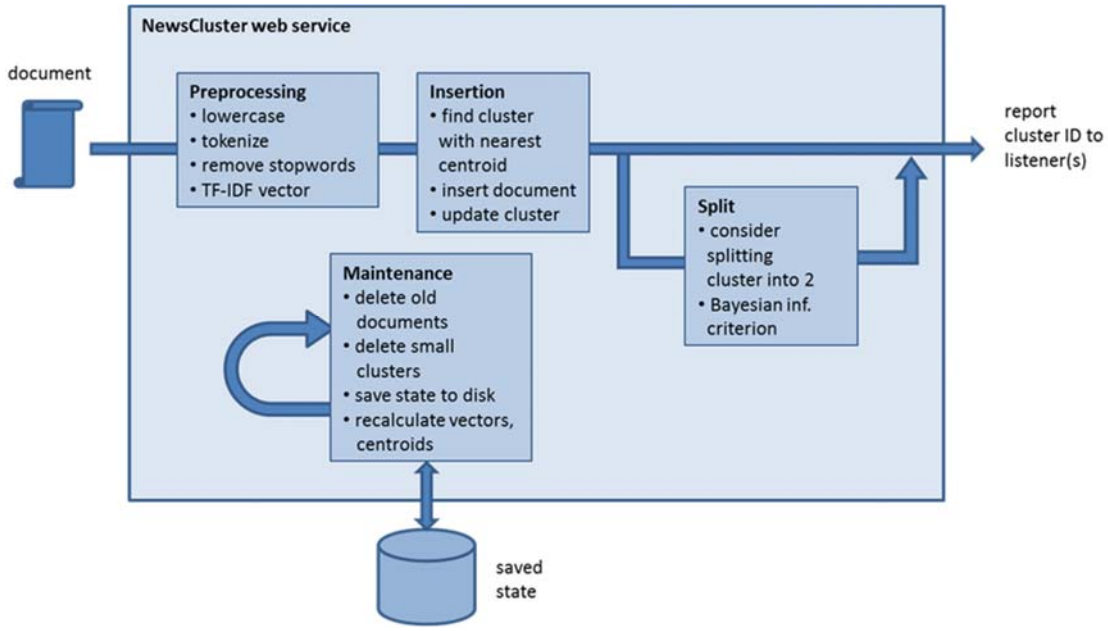

Figure 1 Architecture of NewsCluster service

Document representation

For the purposes of clustering, each document is represented by a TF-IDF feature vector. First, the text of the document is split into words using a slightly adapted version of the Unicode word breaking rules [Dav12]. Next, stopwords are removed if a stopword list for the language of the document is available; currently, the NewsCluster service has stopword lists for English, German, French, Spanish, Italian, Portuguese, and Dutch.

The remaining words are used as the basis of a bag-of-words representation: the document is represented by a sparse feature vector containing one component for each word that appears anywhere in the document set so far. The value of the component corresponding to word t in the feature vector representing the document d is defined as $TFIDF(t, d) = TF(t, d) \cdot IDF(d)$, where $TF(t, d)$ is the *term frequency* (the number of occurrences of term t in the document d) and $IDF(t) = \log(N / DF(t))$ is the *inverse document frequency*, obtained from N (the total number of documents currently maintained in the collection) and $DF(d)$ (the *document frequency* of t, i.e. the number of documents in which t occurs at least once). Finally, the resulting feature vector is normalized so that its Euclidean length equals 1.

Optionally, the NewsCluster service can use n-grams (sequences of up to n adjacent words) as features, in addition to individual words.

Internally, the service stores not only the normalized TF-IDF vector of each document, but also its TF vector and the original full text of the document.

Document weighting

The influence of a document in any clustering-related operation is weighted by a coefficient that decreases exponentially as the age of the document increases. Following [AY06], we use the concept of *half-life*, which is the time period in which a document's weight decreases by one-half. Thus, if t is the timestamp of a given document, the weight of this document at time T will be $(1/2)^{(T - t)/H}$, where H is the half-life period. The default value of H is one day, but this can be modified by the user through a command-line parameter.

Cluster representation

Consider a cluster C containing documents $d_1, ..., d_n$, represented by their respective feature vectors $\mathbf{x}_1, ..., \mathbf{x}_n$ and weights $w_1, ..., w_n$. We maintain the following aggregate statistics for each cluster:

- the sum of weights, $W = \Sigma_{i=1..n}\, w_i$

- the weighted sum of vectors, $\mathbf{S} = \sum_{i=1..n} w_i \mathbf{x}_i$
- the squared norm of **S**, i.e. $||\mathbf{S}||^2 = \mathbf{S}^T\mathbf{S}$
- the weighted sum of squares, $S_2 = \sum_{i=1..n} w_i ||x_i||^2$ (currently, all x_i are normalized to unit length, meaning that $S_2 = W$, but in general this might change if we add more features which might not participate in the same normalization).

These aggregate statistics allow us to efficiently compute various useful quantities about the cluster. For example, the centroid of the cluster can be computed as $\mathbf{c} = (1/W)\ \mathbf{S}$. The variance of the cluster, defined as $d = (1/W) \sum_{i=1..n} w_i ||\mathbf{x}_i - \mathbf{c}||^2$, can be computed efficiently as $d = S_2/W - ||\mathbf{S}||^2/W^2$. With the additional assumption that $||\mathbf{x}_i|| = 1$ for all i, we can also efficiently compute the average cosine similarity between cluster members and the centroid, defined as $a = \sum_{i=1..n} w_i \cos(\mathbf{x}_i, \mathbf{c}) / W$, where $\cos(\mathbf{x}_i, \mathbf{c}) = \mathbf{x}_i^T\mathbf{c} / (||\mathbf{x}_i|| \ ||\mathbf{c}||)$; namely, in this case it is true that $a = ||\mathbf{S}||/W$.

The aggregate statistics can also be updated efficiently when a document is added to or removed from the cluster, or when the time *T* at which the weights are computed changes.

Clustering approach

Initially, the service starts in a "pre-clustering" state, during which it only accumulates incoming clustering without trying to assign them to any clusters. When a certain number of documents has been accumulated (controlled by a command-line parameter, defaulting to 1000), we use a hierarchical bisecting *k*-means (i.e. 2-means) algorithm to obtain an initial partition into clusters, as shown in the following pseudocode:

```
start by placing all documents into one cluster;
while the number of clusters is less than the maximum initial number of clusters do:
   choose the cluster C with the maximum variance from among the current clusters;
   use bisecting k-means to split it into two subclusters C1 and C2;
   if neither C1 nor C2 is too small, replace C with C1 and C2 in our current
partition;
```

At this point, all the listeners are also informed about the initial assignments of documents to clusters. From this point onwards, the service enters its normal mode of operation. Whenever a new document arrives, we compute the cosine similarity between its feature vector and the centroids of all the clusters; the new document is assigned into the cluster where this cosine similarity is maximized.

After the addition of a new document, the cluster is considered for splitting. The conditions for this are that it must contain a sufficient number of documents and that sufficiently many additions must have occurred since the last time it was considered for splitting. If these conditions are met, we try to split the cluster into two subclusters using bisecting *k*-means. We use a variant of the Bayesian Information Criterion [PM00, MG09, Lug12] to decide whether to accept the new split or not. If the cluster is split, all the listeners are notified of the new cluster memberships for all the documents affected by the split.

Web interface

To insert a new document, a HTTP POST request should be made to the NewsCluster web service, at the URL `http://`*server*`:`*port*`/add-article`. The body of the HTTP request should contain (argument name, argument value) pairs in URL-encoded form, e.g.

```
id=12345&lang=eng&text=Lorem+ipsum+dolor+sit+amet
```

The following arguments are required: `id` (giving a unique identifier of the document; if a document with the same identifier already exists, the new document is ignored); `lang` (an ISO-639 three-letter code identifying the language of the document); `title` (optional, giving the title of the document); and `text` (containing the actual contents of the document itself).

Return value: the body of the HTTP request contains a JSON value of the form

```
{"ClusterId":"cluster identifier"}
```

where the *cluster identifier* is a string that globally uniquely identifies the cluster into which the new document has been placed. Additionally, the NewsCluster service will send the new (article ID, cluster ID) pair to any listener(s) whose URLs have been passed in the command-line parameters.

Other commands supported by the NewsCluster web service are:

- `http://`*server*`:`*port*`/save` - forces the service to immediately save its state to disk
- `http://`*server*`:`*port*`/exit` - causes the service to save its state to disk and then terminate
- `http://`*server*`:`*port*`/report` - returns an HTTP response containing an HTML report on the current state of the clusters. The optional parameter `?centroids=1` will cause cluster centroids to be included. The optional parameter `?clusterId=`*number* will generate a report on the cluster whose internal number is given by the *number* parameter; this report includes the list of documents in the cluster. Note that the reports returned by the `report` command are not intended to be machine-readable, but to be human-readable for debugging and informational purposes.

3.4 Integration of Cross-lingual Document Linking

This section describes how cross-lingual document linking integrates with the Newsfeed for the task of tracking similar articles across languages. The approach is based on Cross-Lingual Similarity Service (CLSS) and computes the similarity between English, German, Spanish and Chinese news articles. Given the newsfeed article as an input it returns IDs of top 10 most similar articles for each language in JSON format (Figure 2). The architecture is depicted on Figure 3.

```
{"id":66701562,
  "similar_articles_spa": [{"id":66701512,"sim":0.1364}, …],
  "similar_articles_deu": […],
  "similar_articles_fra": […],
  …
}.
```

Figure 2. Output format of CLSS service.

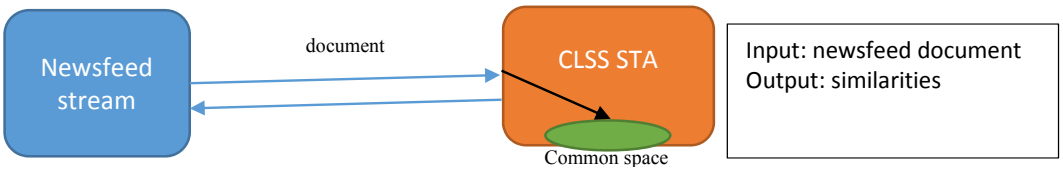

Figure 3. Diagram of intergtion between Newsfeed and CLSS.

Computation of the cross-lingual similarities is based on an aligned set of basis vectors obtained by one of two methods: latent semantic indexing (LSI) and a generalized version of canonical correlation analysis (CCA) by using an aligned multi-lingual corpus. The method enables us to map a multi-lingual collection of documents in the same (common) low dimensional space, where the similarity computation is fast and efficient. For details about the approach refer to [D4.1.1].

A circular one day buffer for each language is used to store documents projected in the common space. This enables the storage of multi-lingual documents in the low dimensional space, where the similarity computation is fast and efficient. Currently we use LSI approach in STA use-case.

4 Visualization Component

According to the functional requirements of Task 5.2 (see Table 8), the visualization prototype should provide functions to show information dynamics across sources, languages and time. However, as a component directly connected with the end users, unilateral displaying or showing is not user-friendly. To enhance user experiences, the visualization component should provide tracking functions in addition to just showing. Therefore, there should be some interfaces for end users to interact with the data provider. The layout of visualization component will be introduced in Section 4.1 to illustrate how the news data are visualized. Section 4.2 will propose the interfaces by which user can interact with data provider.

Table 8. Functional Requirement of Early information visualization.

	Early information visualization
Description	To employ techniques for text and network visualizations for real-time cross-lingual streams to show visual summary of information dynamics across sources, languages and time
Input	Documents or corpus representations from the previous stage
Output	Visualization of the input documents or corpus representations
Languages	XLike languages
Motivation	Entity tracking (BLP), Related or relevant articles (BLP), Article tracking (STA), Topic and entity tracking (STA)
Related task	**T5.2 – Information visualization.**
Evaluation	Effectiveness of visualization (user questionnaire)

4.1 Layout

The visualization component aims to visualize multi-dimensional information of news, including spatial distribution, trends, language, publisher, related entities, keywords and related stories and so on. It's a challenge to visualize all these information on one screen.

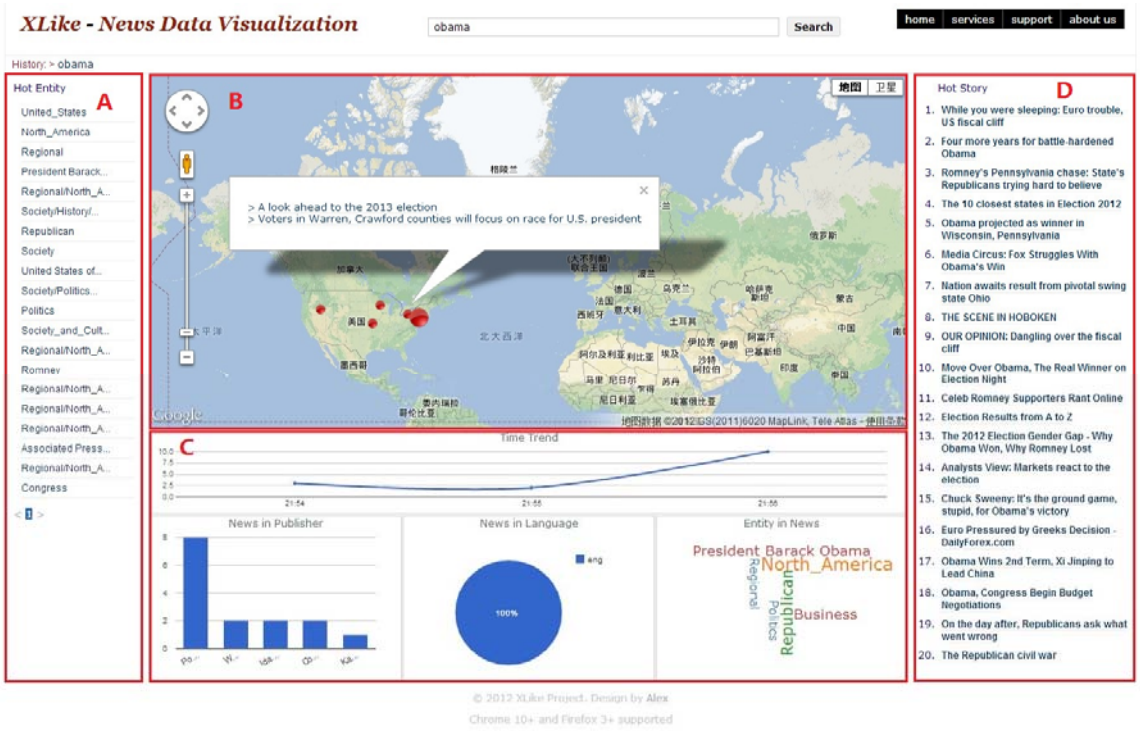

Figure 4. The layout of visualization component

Figure 4 shows the layout of visualization component. To give user an intuitive impression about the spatial

distribution of the news, we place the map in the middle of the page (Area B). Each red icon on the map represents one or more articles published or happened on the corresponding location. An article list will be popped out when clicking these icons. At bottom of the map (Area C), some important statistical data are illustrated. The first chart on the top of area C is the time trend of current request. At the bottom of area C, there are two charts illustrate the distribution of current request over publisher, language respectively. The tag cloud in the right-bottom of area C shows the keywords (and their weights) of current request. Other semantic dimensions such as related entities and stories are listed in the left (Area A) and right (Area D) respectively.

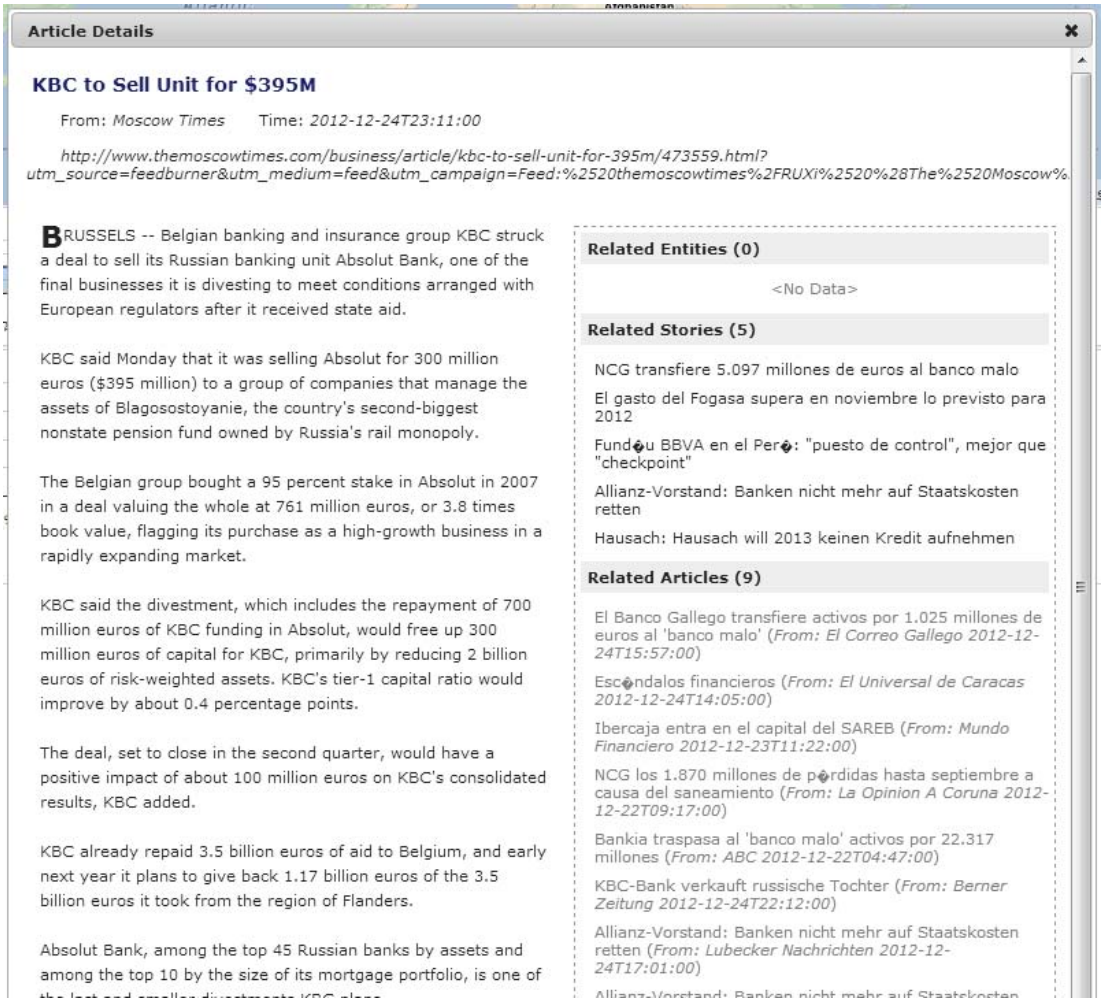

Figure 5. Display of an article with cross-lingual links to other related stories and articles

Figure 5 shows how one article is displayed. The article was identified using keyword search “Slovenia”. Besides standard information, such as title, source, URL, and body, there are also links to other related stories and articles, as identified using cross-lingual document linking. Among related articles there is also a German version of the same news, together with other articles on similar topics.

4.2 User Interactions

The visualization component provides various Computer-Human-Interfaces for end users to interacting with the data provider. This section introduces the user-interaction interfaces in detail.

- Searching

  Action: clicking the search button.

  Response: call ***search*** API and update the result on area A, B, C and D.

- Tracking by entity

  Action: clicking one item in the hot entity list or STA entity list.

  Response: call ***entity*** API and update the result on Area A, B, C and D.

- Tracking by story

  Action: clicking on one item in the story list.

  Response: call ***story*** API and update the result on Area A, B and C; unfold the article list of the clicked story.

- Tracking by keyword

  Action: clicking one item in the keyword cloud (Area C).

  Response: call ***search*** API and update the result on area A, B, C and D.

- Tracking by article

  Action: clicking on one article in the popped article list

  Response: call ***article*** API and show the related articles in the related article list.

- Preference Setting

  Action: user operations on the preference panel.

  Response: update user preferences (supporting parameters). The change will take effect at the next tracking action or searching action.

- Comparing by entities

  Action: a click on the "+" icon on the right side of an entity in the entity list.

  Response: call ***entity*** API and show the difference of the clicked entity and current entity on time trend.

Besides these positive interactions, the visualization component also provides a passive interaction to record user's tracking history (except tracking by article) and shows them in the history navigator. At most 8 tracking operations will be recorded and displayed as links. A click on these links will re-do the corresponding tracking operation.

5 Conclusion

This document presents the deliverable 5.2.1 “early information visualization prototype”. It is a show case of the research outcomes of XLIKE project in the first year. According to the requirements of use cases, we designed a set of cross-lingual data services combining the outcomes of WP1-4. These data services process raw news stream into structured data which can be used by the visualization component.

In this deliverable, we define three objects to describe news information (article, story, entity). Well-designed formal data formats are proposed in this deliverable which can enrich the information of news in various dimensions. We also defined practical formats for these objects in data services.

Six web services are developed to produce structured news data. These web services accept various parameters which are also defined in this deliverable. The output data of these web services are packed in JSon which can be easily processed by visualization component.

At the end, a visualization component which can visualize high dimensional data are developed. It is a web browser based component such that can be deployed in different platform easily. Besides displaying, the visualization component also provides convenient interfaces for users to track articles by entities, stories and similar articles.

The developed prototype is available at http://sandbox-xlike.isoco.com/portal/ .

References

[AHWY03] C. C. Aggarwal, J. Han, J. Wang, P. S. Yu. A framework for clustering evolving data streams. *Proc. of the 29th VLDB Conference*, 2003.

[AY06] C. C. Aggarwal, P. S. Yu. A framework for clustering massive text and categorical data streams. *Proc. 6th SIAM Int. Conf. on Data Mining* (SDM), 2006.

[AY10] C. C. Aggarwal, P. S. Yu. On clustering massive text and categorical data streams. *Knowledge and Information Systems*, 24(2):171-196 (2010).

[D1.3.1] Early prototype of data infrastructure, XLike deliverable

[D2.1.1] Shallow linguistic processing prototype, XLike deliverable

[D3.1.1] Early text annotation prototype, XLike deliverable

[D4.1.1] Cross-lingual document linking prototype, XLike deliverable

[D6.2.1] Early prototype, XLike deliverable

[Dav12] M. Davis (ed.) *Unicode Text Segmentation*. Unicode Standard Annex #29, Unicode 6.2.0, 12 September 2012.

[Lug12] E. Lughofer. A dynamic split-and-merge approach for evolving cluster models. *Evolving Systems*, 3:135-151 (2012).

[MG09] M. Muhr, M. Granitzer. Automatic cluster number selection using a split and merge k-means approach. *DEXA Workshops* 2009, pp. 363-7.

[PM00] D. Pelleg, A. Moore. X-means: Extending K-means with efficient estimation of the number of clusters. *Proc. ICML* 2000.

Annex A External components and tools

A.1 Google Maps

Google Maps is a set of services provided by Google. In the visualization component, the Google Maps Javascript API and Google Map API Geocoding web service are employed.

The Google Maps Javascript API lets developers embed Google Maps in your own web pages. Version 3 of this API is especially designed to be faster and more applicable to mobile devices, as well as traditional desktop browser applications.

The Google Map API Geocoding API provides interface to convert addresses (like "1600 Amphitheatre Parkway, Mountain View, CA") into geographic coordinates (like latitude 37.423021 and longitude -122.083739), which you can use to place markers or position the map. The Google Geocoding API provides a direct way to access a geocoder via an HTTP request. Additionally, the service allows you to perform the converse operation (turning coordinates into addresses); this process is known as "reverse geocoding."

Google provides a free license to these APIs but with a condition that our service must be **freely and publicly accessible** to end users.

A.2 Google Chart Tools

Google Chart Tools provide a perfect way to visualize data on your website. From simple line charts to complex hierarchical tree maps, the chart galley provides a large number of well-designed chart types. Populating your data is easy using the provided client- and server-side tools.

In this project, Charts are rendered using HTML5/SVG technology to provide cross-browser compatibility (including VML for older IE versions) and cross platform portability to iPhones, iPads and Android. **No plugins are needed**.

Annex B STA Entities

1) general section	
Slovenija	http://en.wikipedia.org/wiki/Slovenia
Ljubljana	http://en.wikipedia.org/wiki/Ljubljana
Maribor	http://en.wikipedia.org/wiki/Maribor
Koper	http://en.wikipedia.org/wiki/Koper
Celje	http://en.wikipedia.org/wiki/Celje
Kranj	http://en.wikipedia.org/wiki/Kranj
Ptuj	http://en.wikipedia.org/wiki/Ptuj
Novo mesto	http://en.wikipedia.org/wiki/Novo_mesto
Nova Gorica	http://en.wikipedia.org/wiki/Nova_Gorica
Slovenj Gradec	http://en.wikipedia.org/wiki/Slovenj_Gradec
Velenje	http://en.wikipedia.org/wiki/Velenje
Murska Sobota	http://en.wikipedia.org/wiki/Murska_Sobota
Bled	http://en.wikipedia.org/wiki/Bled
Bohinj	http://en.wikipedia.org/wiki/Bohinj
Kranjska Gora	http://en.wikipedia.org/wiki/Kranjska_Gora
reka Soča	http://en.wikipedia.org/wiki/So%C4%8Da
Portorož	http://en.wikipedia.org/wiki/Portoro%C5%BE
Izola	http://en.wikipedia.org/wiki/Izola
Piran	http://en.wikipedia.org/wiki/Piran
Pohorje	http://en.wikipedia.org/wiki/Pohorje
reka Mura	http://en.wikipedia.org/wiki/Mura_River
Kobarid	http://sl.wikipedia.org/wiki/Kobarid
predor Karavanke	http://en.wikipedia.org/wiki/Karavanke_Tunnel
Ljubelj	http://en.wikipedia.org/wiki/Ljubelj
Bovec	http://en.wikipedia.org/wiki/Bovec
Trenta	http://en.wikipedia.org/wiki/Trenta_%28valley%29
Triglav	http://en.wikipedia.org/wiki/Triglav
Otočec	http://en.wikipedia.org/wiki/Oto%C4%8Dec
Kolpa	http://en.wikipedia.org/wiki/Kupa
Lipica	http://en.wikipedia.org/wiki/Lipica
kranjska klobasa	http://en.wikipedia.org/wiki/Kranjska_klobasa http://sl.wikipedia.org/wiki/Kranjska_klobasa
Postojnska jama	http://en.wikipedia.org/wiki/Postojna_Cave
Predjamski grad	http://en.wikipedia.org/wiki/Predjama_Castle
Škocjanske jame	http://en.wikipedia.org/wiki/%C5%A0kocjan_Caves
potica	http://en.wikipedia.org/wiki/Nut_roll http://sl.wikipedia.org/wiki/Potica
štruklji	http://sl.wikipedia.org/wiki/%C5%A0truklji
prekmurska gibanica	http://en.wikipedia.org/wiki/Prekmurska_gibanica
poplava (flood)	http://en.wikipedia.org/wiki/Flood
neurje (storm)	http://en.wikipedia.org/wiki/Storm
požar (fire)	http://en.wikipedia.org/wiki/Fire

2) sport section	
Anže Kopitar	http://en.wikipedia.org/wiki/Anze_Kopitar
Planica	http://en.wikipedia.org/wiki/Planica
Tina Maze	http://en.wikipedia.org/wiki/Tina_Maze
Zlata Lisica	http://sl.wikipedia.org/wiki/Zlata_lisica
Pokal Vitranc	http://sl.wikipedia.org/wiki/Pokal_Vitranc
EP v košarki 2013	http://en.wikipedia.org/wiki/FIBA_EuroBasket_2013 http://sl.wikipedia.org/wiki/Evropsko_prvenstvo_v_ko%C5%A1arki_2013
Goran Dragić	http://en.wikipedia.org/wiki/Goran_Dragi%C4%87
Erazem Lorbek	http://en.wikipedia.org/wiki/Erazem_Lorbek
Samir Handanović	http://en.wikipedia.org/wiki/Samir_Handanovi%C4%87
Saša Vujačić	http://en.wikipedia.org/wiki/Sa%C5%A1a_Vuja%C4%8Di%C4%87
Pokljuka, biatlon	http://sl.wikipedia.org/wiki/Pokljuka
NK Maribor	http://en.wikipedia.org/wiki/FC_Maribor
Celje Pivovarna Laško	http://en.wikipedia.org/wiki/Celje_Pivovarna_La%C5%A1ko
Gorenje Velenje	http://en.wikipedia.org/wiki/RK_Gorenje
RK Krim	http://en.wikipedia.org/wiki/RK_Krim
ACH Volley	http://en.wikipedia.org/wiki/ACH_Volley_Ljubljana
Janez Brajkovič	http://en.wikipedia.org/wiki/Janez_Brajkovi%C4%8D
Lipica, tekmovanje v preskakovanju ovir, tekmovanje v dresurnem jahanju	http://en.wikipedia.org/wiki/Lipica http://en.wikipedia.org/wiki/Show_jumping
lipicanci	http://en.wikipedia.org/wiki/Lipizzaner
Polona Hercog	http://en.wikipedia.org/wiki/Polona_Hercog
Katarina Srebotnik	http://en.wikipedia.org/wiki/Katarina_Srebotnik
Peter Kauzer	http://en.wikipedia.org/wiki/Peter_Kauzer
Fiba Europe	http://en.wikipedia.org/wiki/FIBA_Europe
Eurobasket 2013	http://en.wikipedia.org/wiki/Eurobasket_2013
3) economy section	
Janez Šušteršič	http://sl.wikipedia.org/wiki/Janez_%C5%A0u%C5%A1ter%C5%A1i%C4%8D http://www.mf.gov.si/en/about_the_ministry/who_is_who/minister_of_finance/
Marko Kranjec	null
Krka	http://en.wikipedia.org/wiki/Krka_%28company%29
Petrol	http://en.wikipedia.org/wiki/Petrol_Group
Gorenje	http://en.wikipedia.org/wiki/Gorenje
Mercator	http://en.wikipedia.org/wiki/Mercator_%28retail%29
Telekom Slovenije	http://en.telekom.si/company/organization
Ipko	http://en.wikipedia.org/wiki/IPKO
One	http://en.wikipedia.org/wiki/One_%28Telekom_Slovenija_Group%29
Gibtelecom	http://en.wikipedia.org/wiki/Gibtelecom
Perutnina Ptuj	http://en.wikipedia.org/wiki/Perutnina_Ptuj
Aerodrom Ljubljana	http://en.wikipedia.org/wiki/Aerodrom_Ljubljana
Aerodrom Maribor	http://en.wikipedia.org/wiki/Maribor_Edvard_Rusjan_Airport

Adria Airways	http://en.wikipedia.org/wiki/Adria_Airways
Luka Koper	http://en.wikipedia.org/wiki/Luka_koper
Nova Ljubljanska banka (NLB)	http://en.wikipedia.org/wiki/Nova_Ljubljanska_banka http://www.nlb.si/the-bank-today
Nova Kreditna banka Maribor (Nova KBM)	http://www.nkbm.si/mission http://www.nkbm.si/novakbm-group-companies
Pipistrel	http://en.wikipedia.org/wiki/Pipistrel
Akrapovič	http://en.wikipedia.org/wiki/Akrapovi%C4%8D
Slovenske železnice	http://en.wikipedia.org/wiki/Slovenske_%C5%BEeleznice
Nuklearna elektrarna Krško (NEK)	http://en.wikipedia.org/wiki/Nuklearna_Elektrarna_Krsko
Termoelektrarna Šoštanj	http://www.te-sostanj.si/en/
Revoz	http://www.revoz.si/en/inside.cp2?cid=6CC9E0E9-BBC7-F39E-72A3-FD971F3A2537&linkid=inside
Terme Maribor	http://sl.wikipedia.org/wiki/Terme_Maribor,_d.d.
Terma Čatež	http://sl.wikipedia.org/wiki/Terme_%C4%8Cate%C5%BE
Pivovarna Laško	http://en.wikipedia.org/wiki/Pivovarna_La%C5%A1ko
Primorje	http://www.primorje.si/index.php?lng=eng
Banka Slovenije	http://en.wikipedia.org/wiki/Bank_of_Slovenia
Sloga	http://sl.wikipedia.org/wiki/Sloga
slovenske obveznice (Slovenija, obveznice)	http://en.wikipedia.org/wiki/Slovenia http://en.wikipedia.org/wiki/Bond_%28finance%29
slovensko gospodarstvo (Slovenija, gospodarstvo)	http://en.wikipedia.org/wiki/Slovenian_economy
Zavarovalnica Triglav	http://sl.wikipedia.org/wiki/Zavarovalnica_Triglav
Hidria	http://www.hidria.com/en/about-us/
Kolektor	http://www.kolektor.com/en/about-the-group
Holding Slovenske elektrarne	http://en.wikipedia.org/wiki/Holding_Slovenske_elektrarne
Lek	http://en.wikipedia.org/wiki/Lek_%28pharmaceutical_company%29
Sandoz	http://en.wikipedia.org/wiki/Sandoz
Merkur	http://sl.wikipedia.org/wiki/Merkur_Group
Tuš	http://www.tus.si/en/about-tus
Istrabenz	http://en.wikipedia.org/wiki/Istrabenz
Cinkarna Celje	http://sl.wikipedia.org/wiki/Cinkarna_Celje http://www.cinkarna.si/en/company-profile
Iskra Avtoelektrika - Letrika	http://www.letrika.com/en/group-letrika/about-group/
TAB	http://www.tab.si/teksti.php?id=1
Simobil	http://en.wikipedia.org/wiki/Simobil
Geoplin	http://en.wikipedia.org/wiki/Geoplin http://www.geoplin.si/eng/about-us
Hella Saturnus Slovenija	http://www.hella-saturnus.si
Slovenska industrija jekla	http://www.sij.si/en/who-we-are/
Metal Ravne	http://www.metalravne.com
Talum	http://en.wikipedia.org/wiki/Talum
Acroni	http://www.acroni.si/
Sava	http://www.sava.si/eng/about-sava.html

Seaway	http://sl.wikipedia.org/wiki/SeaWay
Lafarge Cement Trbovlje	http://sl.wikipedia.org/wiki/Lafarge_Cement
Elan	http://en.wikipedia.org/wiki/Elan_%28company%29
Splošna plovba	http://sl.wikipedia.org/wiki/Splo%C5%A1na_plovba
Intereuropa	http://www.intereuropa.si/index.php?page=about_us&item=1
Adria Mobil	http://en.wikipedia.org/wiki/Adria_Mobil
Alpina	http://en.wikipedia.org/wiki/Alpina_%C5%BDiri
Lisca	http://en.wikipedia.org/wiki/Lisca_%28company%29
Amis	http://www.amis.net
T-2	http://english.t-2.net/
Trimo	http://sl.wikipedia.org/wiki/Trimo
4) politics section	
Janez Janša	http://en.wikipedia.org/wiki/Janez_Jan%C5%A1a
slovenska vlada, Vlada Republike Slovenije	http://en.wikipedia.org/wiki/Government_of_the_Republic_of_Slovenia
predsednik Republike Slovenije, slovenski predsednik	http://en.wikipedia.org/wiki/President_of_Slovenia
Državni zbor RS	http://en.wikipedia.org/wiki/Parliament_of_Slovenia
Gregor Virant	http://en.wikipedia.org/wiki/Gregor_Virant
Danilo Türk	http://en.wikipedia.org/wiki/Danilo_T%C3%BCrk
Slovenska demokratska stranka (SDS)	http://en.wikipedia.org/wiki/Slovenian_Democratic_Party
Nova Slovenija (NSi)	http://en.wikipedia.org/wiki/New_Slovenia
Državljanska lista	http://en.wikipedia.org/wiki/Civic_List_%28Slovenia%29
Pozitivna Slovenija (PS)	http://en.wikipedia.org/wiki/Positive_Slovenia
Slovenska ljudska stranka (SLS)	http://en.wikipedia.org/wiki/Slovenian_People%27s_Party
Demokratična stranka upokojencev Slovenije (DsSUS)	http://en.wikipedia.org/wiki/Democratic_Party_of_Pensioners_of_Slovenia
SD	http://en.wikipedia.org/wiki/Social_Democrats_%28Slovenia%29
Zoran Janković	http://en.wikipedia.org/wiki/Zoran_Jankovi%C4%87
Karl Erjavec	http://en.wikipedia.org/wiki/Karl_Erjavec
Radovan Žerjav	http://en.wikipedia.org/wiki/Radovan_%C5%BDerjav
Borut Pahor	http://en.wikipedia.org/wiki/Borut_Pahor
slovenska policija / Slovenija, policija	http://en.wikipedia.org/wiki/Police
slovenska vojska	http://en.wikipedia.org/wiki/Military_of_Slovenia
Univerza v Ljubljani	http://en.wikipedia.org/wiki/University_of_ljubljana
Univerza v Mariboru	http://en.wikipedia.org/wiki/University_of_Maribor
Univerza na Primorskem	http://en.wikipedia.org/wiki/University_of_Primorska
slovenski profesor	http://en.wikipedia.org/wiki/Professor http://en.wikipedia.org/wiki/Slovenia
slovenski znanstvenik	http://en.wikipedia.org/wiki/Scientist http://en.wikipedia.org/wiki/Slovenia
Romi v Sloveniji	http://en.wikipedia.org/wiki/Romani_people

	http://en.wikipedia.org/wiki/Slovenia
5) foreign affairs section	
Ministrstvo RS za zunanje zadeve	http://en.wikipedia.org/wiki/Ministry_of_Foreign_Affairs_%28Slovenia%29
zunanji minister Karl Erjavec	http://en.wikipedia.org/wiki/Karl_Erjavec
Ministrstvo za obrambo RS	http://en.wikipedia.org/wiki/Ministry_of_Defence_%28Slovenia%29 http://www.mo.gov.si/en/
obrambni minister Aleš Hojs	http://www.mo.gov.si/en/about_the_ministry/leadership/
slovenska veleposlaništva in konzulati	http://en.wikipedia.org/wiki/Embassy http://en.wikipedia.org/wiki/Slovenia
Alojz Peterle	http://en.wikipedia.org/wiki/Alojz_Peterle
Ivo Vajgl	http://en.wikipedia.org/wiki/Ivo_Vajgl
Milan Zver	http://en.wikipedia.org/wiki/Milan_Zver
Tanja Fajon	http://en.wikipedia.org/wiki/Tanja_Fajon
Zofija Mazej Kukovič	http://www.zofijamazejkukovic.net/eng/about/
Mojca Kleva	http://mojcakleva.eu/about-me1/
Romana Jordan	http://en.wikipedia.org/wiki/Romana_Jordan_Cizelj
Jelko Kacin	http://en.wikipedia.org/wiki/Jelko_Kacin
Urad vlade za Slovence v zamejstvu in po svetu	http://www.uszs.gov.si/en/areas_of_activity/
Ljudmila Novak	http://en.wikipedia.org/wiki/Ljudmila_Novak http://www.uszs.gov.si/en/about_the_office/leadership/
Svetovni slovenski kongres	http://www.slokongres.com/index.php?option=com_content&view=article&id=1&Itemid=4&lang=sl
Slovenska izseljenska matica	http://sl.wikipedia.org/wiki/Slovenska_izseljenska_matica http://www.zdruzenje-sim.si/kdo_smo/sim_danes/
društvo Slovenija v svetu	http://www.drustvo-svs.si/jupgrade/index.php?option=com_content&view=article&id=458:ob-20-letnici-drutva-svs&catid=43&Itemid=79
Slovensko panevropsko gibanje	http://www.panevropa.si/o_gibanju.htm
Laris Gaiser	null
Rudolf Gabrovec	null
Jernej Sekolec	null
Ernest Petrič	http://sl.wikipedia.org/wiki/Ernest_Petri%C4%8D
Slovenci, kazenska ovadba	http://en.wikipedia.org/wiki/Slovenians http://en.wikipedia.org/wiki/Slovenia
Slovenci, smrt	http://en.wikipedia.org/wiki/Death
Ruska kapelica	http://en.wikipedia.org/wiki/Slovenia http://en.wikipedia.org/wiki/Russian_Chapel,_Vr%C5%A1i%C4%8D
6) culture section	
EPK, Maribor	http://en.wikipedia.org/wiki/European_Capital_of_Culture http://en.wikipedia.org/wiki/Maribor
Festival Ljubljana	http://en.wikipedia.org/wiki/Ljubljana_Summer_Festival http://sl.wikipedia.org/wiki/Festival_Ljubljana
Festival Brežice	http://seviqc-brezice.si/index.php/about-festival/name-of-the-festival http://seviqc-brezice.si/index.php/about-festival/programme-objectives

Liffe	http://en.wikipedia.org/wiki/Ljubljana_International_Film_Festival
Bienale industrijskega oblikovanja	http://www.bio.si/About-Biennial.aspx
Grafični bienale	http://www.mglc-lj.si/slo/index-onas.htm
festival Vilenica	http://en.wikipedia.org/wiki/Vilenica_Prize http://www.vilenica.si/about_vilenica/p/170/1/2/
Laibach	http://en.wikipedia.org/wiki/Laibach_%28band%29
Irwin	http://en.wikipedia.org/wiki/IRWIN http://sl.wikipedia.org/wiki/Irwin
Drago Jančar	http://en.wikipedia.org/wiki/Drago_Jan%C4%8Dar
Boris Pahor	http://en.wikipedia.org/wiki/Boris_Pahor
Sabina Cvilak	http://sl.wikipedia.org/wiki/Sabina_Cvilak_Damjanovi%C4%8D
Mesto žensk	http://www.cityofwomen.org/en/content/info
festival Fabula	http://en.wikipedia.org/wiki/Fabula_Award http://www.festival-fabula.org/2012/eng/
slovenski PEN center	http://sl.wikipedia.org/wiki/Slovenski_center_PEN
duo Silence	http://en.wikipedia.org/wiki/Silence_%28band%29
Marjana Lipovšek	http://en.wikipedia.org/wiki/Marjana_Lipov%C5%A1ek
Dragan Živadinov – KSEVT	http://en.wikipedia.org/wiki/Dragan_%C5%BDivadinov
Zavod Bunker	http://www.bunker.si/eng/about-bunker
Branimir Slokar	http://sl.wikipedia.org/wiki/Branimir_Slokar
Zavod Maska	http://www.maska.si/index.php?id=25&L=0&id=25
Marko Pogačnik	http://en.wikipedia.org/wiki/Marko_Poga%C4%8Dnik
Vinko Globokar	http://en.wikipedia.org/wiki/Vinko_Globokar
Perpetuum Jazzile	http://en.wikipedia.org/wiki/Perpetuum_Jazzile
Tomaž Pandur	http://www.pandurtheaters.com/#!/tomaz-pandur
festival Lent	http://sl.wikipedia.org/wiki/Festival_Lent
France Prešeren	http://en.wikipedia.org/wiki/France_Pre%C5%A1eren
Ivan Cankar	http://en.wikipedia.org/wiki/Ivan_Cankar
Janez Trdina	http://en.wikipedia.org/wiki/Janez_Trdina
Kobariški muzej	http://sl.wikipedia.org/wiki/Kobari%C5%A1ki_muzej
Hugo Wolf	http://en.wikipedia.org/wiki/Hugo_Wolf
Bolnica Franja	http://en.wikipedia.org/wiki/Franja_Partisan_Hospital
muzej Baza 20	http://www.burger.si/MuzejiInGalerije/DolenjskiMuzej/Baza20/Baza20_ENG.html
Sinagoga Maribor	http://sl.wikipedia.org/wiki/Sinagoga_Maribor
Kočevski rog	http://en.wikipedia.org/wiki/Ko%C4%8Devski_rog