Conceptual Structure of Application for Personalized News Feeds

Viktoriia Puzankova, Oleg Faitas, Yuri Tyhun

CAD Department, Lviv Polytechnic National University, UKRAINE, Lviv, S. Bandery street 12,
E-mail: puzankova.viktoriia@gmail.com, faitas.oleg@gmail.com, yura_tyhun@mail.ru

Abstract – **This paper describes a conceptual structure of application, which based on user preferences personalize news feeds. Personalize news feeds application allows for user to save time and gives him only interesting articles, news.**

Keywords – **news, personalization, news aggregator, database, keyword, conceptual structure, MEMS.**

I. INTRODUCTION

With the rapid development of the Internet and its permeability, people get to know the latest news through online media, blogs and news feeds. For easy organization are news aggregators that allow a single place to view news from various sources (e.g. feedly, the old reader, newsblur).

There are several problems with the use of such means of aggregation:

• one and the same news illuminate by several media outlets that serve only the facts - spends time on re-read the information already obtained;

• one and the same problem covering several media, but in different languages;

• many news sites have not categorize their news feeds – reads news which are not interested in.

This paper presents the conceptual framework and algorithm application that will resolve the given problem.

II. CONCEPTUAL STRUCTURE

At the core algorithm personalize news feeds have an idea identify key words, phrases and track user responses to them. This will have base areas that interested user. With this, halt will filter out news information feeds. In case, there is no perfect method for extracting keywords from the text [1]. Especially problematic is available in languages where each token characterized by a large number of word forms, including Ukrainian. Some methods require a corpus of texts, but very difficult to create an array of texts that would reflect a broad subject area.

Keywords are commonly used for search engines and document databases to locate information and determine if two pieces of test are related to each other. Reading and summarizing the contents of large entries of text into a small set of topics is difficult and time consuming for a human, so much so that it becomes nearly impossible to accomplish with limited manpower as the size of the information grows. As a result, automated systems are being more commonly used to do this task.

This problem is challenging due to the intricate complexities of natural language, as well as the inherent difficulty in determining if a word or set of words accurately represent topics present within the text.

With the advent of the internet, there is now both a massive amount of information available, as well as a demand to be able to search through all of this information. Keyword extraction from text data is a common tool used by search engines and indexes alike to quickly categorize and locate specific data based on explicitly or implicitly supplied keywords.

One solution to this problem is the idea of using an array of edits Wikipedia [2]. User entered request divided into a number of tokens, rejecting the words of the "stop-list" (words that have no substantial load: articles, prepositions, conjunctions, pronouns, etc.). According to the list of tokens searched by title Wikipedia, including by means of elected context most relevant articles.

Based on the array of terms based on semantic graph loaded ribs, which demonstrates the closeness of terms with each other. With Hirvana-Newman, algorithm graph divided into subgraphs communities [3]. To assess the quality-partitioning column made counting the number of edges in the community and beyond. Bold keywords is in several communities (usually 1 to 3) with the highest ranks for counting scales ribs.

Each article will include shows extracted from it keywords. There are some rules that will affect the presentation of such news:

• If a user read only the headline and not moved to a detailed description news - lowers the priority of such news presentation and keywords extracted from it; a conflict may arise if headline news highlights all the required information to the user. In this case, you can use the ratio of keywords in title to its length, but this method will provide only part of the right decisions on priorities keywords;

• User moved to detailed news. In this review, it is able to:
- notice key words, which it is interested. In this case, the rating of these words will increase;
- notice key words that he was not interested to continue to show less news with keywords - words such ratings decreases;

- just read the news - rating all keywords increases, but at a smaller value than at marked them as interesting.

Keywords given rating scales. The easiest way to implement it, isolating each recording a separate keyword field where the default weight is zero, and changes in a positive or negative way, which will help sort news by user preferences.

A similar technique using Google and Facebook for the presentation of relevant contextual advertising.

Another feature is the search and news coverage on similar terms. For example, when user interest in the company STMitsroeletstronitss application must also get updates about Analog Devices, Kionix and other companies that produce MEMS, sensors, etc., that is, the companies that work in one area. In the same way, the system works, and vice versa - if people do not care MEMS particular company, the rating of all keywords related to microprocessors has reduced. In this case there is a problem, because the user may be interested in finding MEMS, but it does not interest products separate company, in the event of the user from the news advisable to display a list of reasons why this news is not interesting for him, where he can enter - it not interested in specific news or general news in this subject area.

The problem of coverage of multiple sources of the same information solved with existing methods of checking for text uniqueness. It should establish a threshold percentage, which will determine whether the news shown to the user. If these sources in different languages, the application must cover the news in the language of a given request, if the user does not put priorities, otherwise the application can inform the user about the existence article in another language. Combining these features will personalize news feeds - algorithm and conceptual diagram below.
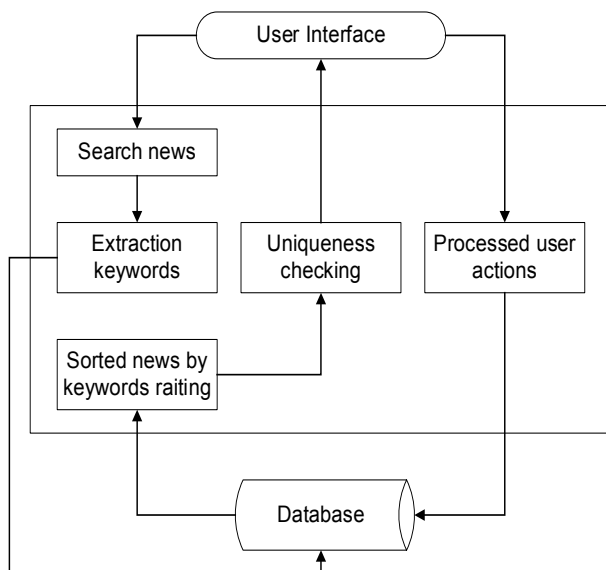

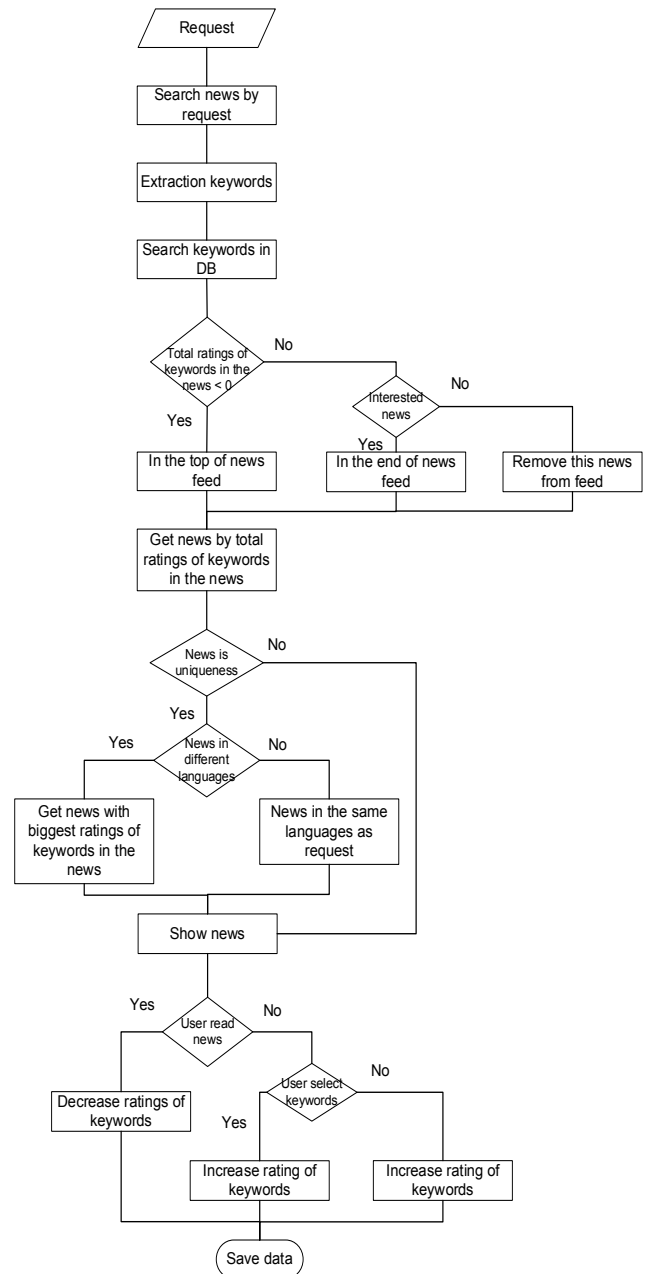Fig.1 Conceptual structure of application

Fig.2 Algorithm of application

III. SOFTWARE

This application based on ASP.NET. This technology gives us possibility to build RESTful service. Microsoft introducing new ASP.NET Core 1.0[4]. This new version of ASP.NET based on .NET Core 1 which allowed develop applications for Mac OS X and Linux.

As database, we used Microsoft SQL Server with MongoDB. This combination allowed to us to increase performance for searching data in DB. SQL Server will contained information about user, their news feed and keywords.

ASP.NET with database will hosted in separate servers, for users will developed different user interfaces: web interface, interface for Android OS, IOS and Windows

Mobile. This approach gives us more flexibility and shift some functionality on the client side.
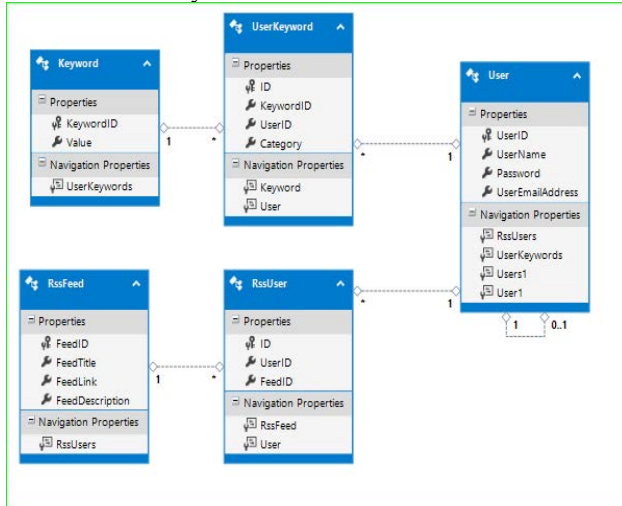

Fig.3 Database structure

Client side will developed with HTML5 with JavaScript. There are many framework that gives developer possibility develop application for IOS or Android on HTML5 (phonegap, framework7).
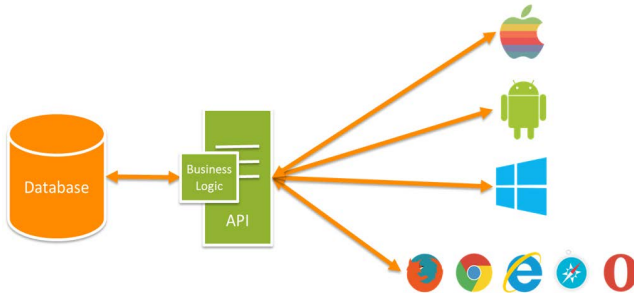

Fig.4 Architecture with a central API, which hold all the business logic [5]

Each client side communicated with database throw RESTful service. REST stands for 'Representational State Transfer' and it is an architectural pattern for creating an API that uses HTTP as its underlying communication method. .Net's Web API is an easy way to implement a RESTful web service using all of the goodness that the .net framework provides.

REFERENCES

[1] Hulth A. Combining machine learning and natural language processing for automatic keyword extraction. Stockholm University, Faculty of Social Sciences, Department of Computer and Systems Sciences (together with KTH), 2004.

[2] Anton Mykhailiuk, A creation of the linguistic ontology based on astructured electronic encyclopedic resource, Anton Mykhailiuk, Olena Mykhailiuk, Oleksiy Pylypchuk, Volodymyr Tarasenko, International Journal of Computing, 13(1) 2014, 1-2

[3] Role of Natural Language Processing in Community Structure Detection - http://www.academia.edu/768236/Role_of_Natural_Language_Processing_in_Community_Structure_Detection

[4] An Update on ASP.NET Core and .NET Core - https://blogs.msdn.microsoft.com/webdev/2016/02/01/an-update-on-asp-net-core-and-net-core/

[5] Introduction to REST and .net Web API - https://blogs.msdn.microsoft.com/martinkearn/2015/01/05/introduction-to-rest-and-net-web-api/

III. CONCLUSION

The developed algorithm personalize news feeds saves the user time and give him only interesting articles, news etc.