

Figure 2: Algorithm

At the same time, the question of highlighting story chains in the web content of a certain industry is open: even if due to one important event, a number of less important derivative events occurred. To form a semantic core, it is easier to analyze the entire set of terms, rather than highlight a weighty primary source.

The accuracy of the approach largely depends on the quality of the primary processing of web content: the quality of syntactic models [3] and dictionaries.

Thus, for an error-free combination of terms in a category, it is necessary to use the developed algorithm and criterion for assessing the degree of proximity.

4. Software designing

Based on the requirements (figure 2), we start to design a software taking into account that there are two main users: user and administrator, which share the functionality. Using UML [7], the following use case diagram was designed (figure 3).

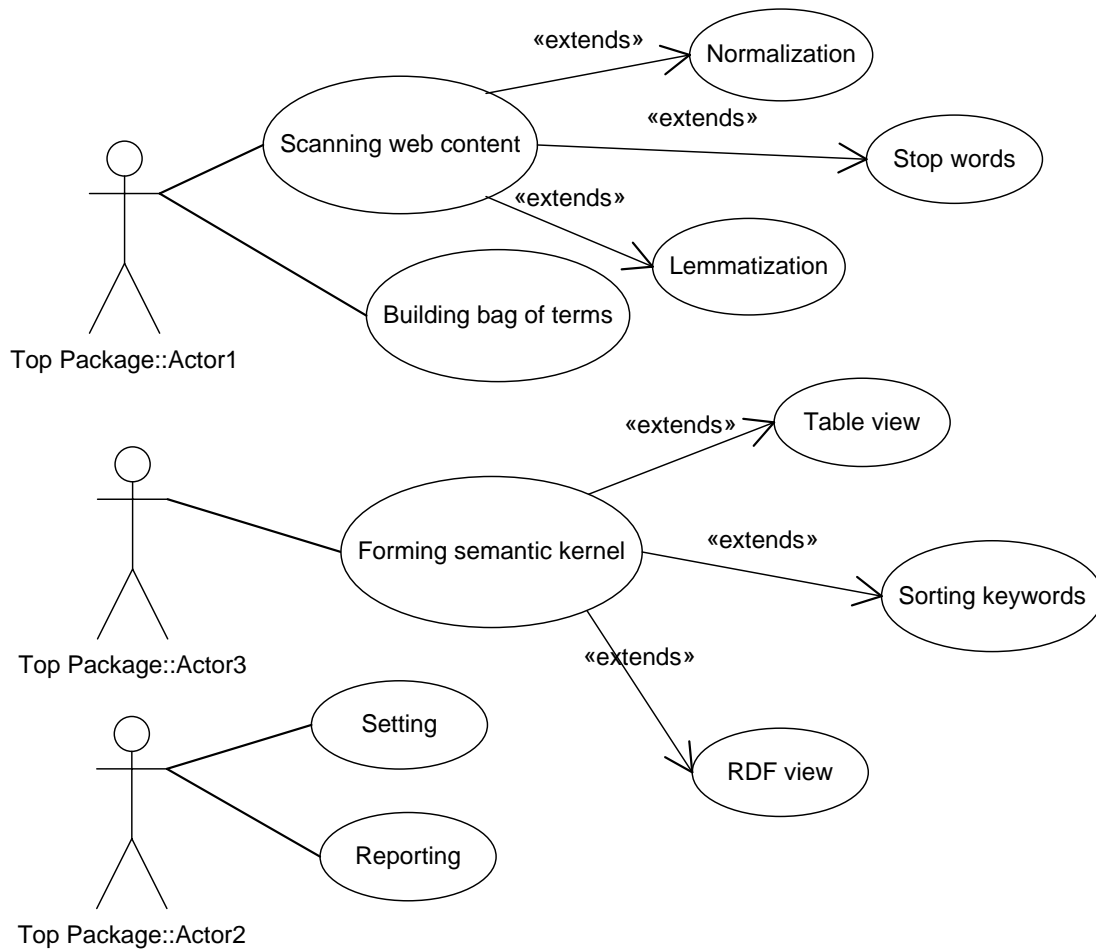


Figure 3: Use cases

The main idea is to represent the final result of the algorithm execution in the form of an RDF schema. It is a handy tool that includes the basic components of a vector \vec{n} . The operating principle of the software is shown in the diagrams - Figures 4 and 5.

As we can see, the software takes the web content of the web resource as the initial information. Next, the primary processing of web texts is performed (lemmatization and normalization). Then our clustering algorithm is applied and, as a result, an xml file is formed - an RDF schema.

5. Results

Let us look at the potential results of applying the proposed algorithm on the example of web content on the topic of astrology and psychology. Figure 6 shows an example of such content, followed by highlighting a vocabulary of terms based on the frequency of their occurrence in the text.

The web content of this site has changed three times for the last ten years. As you can see, from Google Analytics data, each such change was accompanied by a surge in user activity. In the example (figure 7), the semantic kernel was analyzed at the stage of its first change. The main task of the kernel was to fix in the minds of potential users of the site and the Internet in general that the CelestialTiming site [8] logo is associated with the direction of astrology and psychology. Subsequent changes to the core were aimed at the next two stages of the marketing of this web project.

The first stage is to promote the services of this website, namely the construction of a psychological portrait based on the user's personal data.

The second stage is to promote a new service - the school of psychology and astrology. In accordance with the stages, the semantic core of the website is also changed.

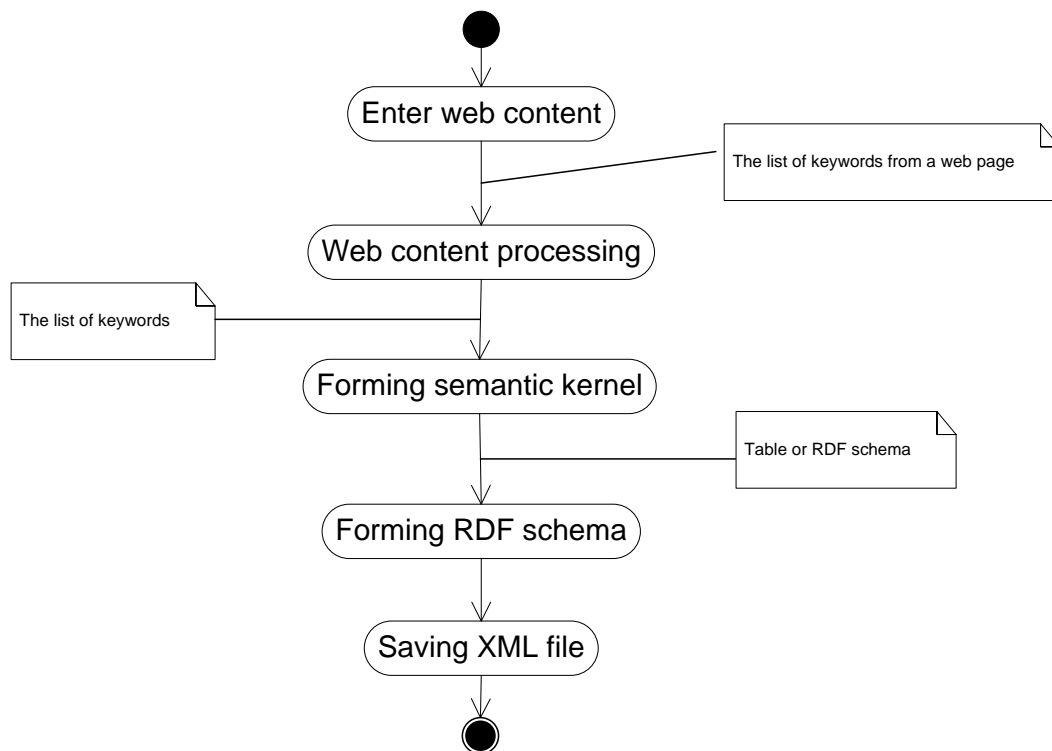


Figure 4: Activity

PROCEED	RDF
1. {"celestialtiming":["Activity", "Noun"]}	
2. {"is":["Verb", "Copula"]}	
3. {"a":["Determiner"]}	
4. {"unique":["Adjective"]}	
5. {"site":["Noun"]}	
6. {"that":["Determiner"]}	
7. {"gives":["PresentTense", "Verb"]}	
8. {"you":["Pronoun", "Noun", "Singular"]}	
9. {"a":["Determiner"]}	
10. {"glance":["Noun", "Singular"]}	
11. {"inside":["Adjective"]}	
12. {"the":["Determiner"]}	
13. {"universal":["Adjective"]}	
14. {"clock":["Noun", "Singular"]}	
15. {"and":["Conjunction"]}	
16. {"allows":["PresentTense", "Verb"]}	
17. {"you":["Pronoun", "Noun", "Singular"]}	

Figure 5: Final web form

Unfortunately, the creators of the web project made several mistakes. The first is that the kernels must be linked when changed. The second mistake is that the kernel should be changed not only on the website itself, but also on friendly links pointing to this kernel. The third mistake is that each core has its own life cycle and it is different from each other. This can be seen, for example, in figure 7. The fourth mistake is that the kernels, based on their aging effect, need changing more often than the creators of this project do. The fifth mistake is that all kernel changes must fit within the framework of a single marketing strategy.

All these comments were passed on to the developers of this web project.

Divine timing for individual success

CelestialTiming

HOME
DAILY FORECASTS
FREE HOROSCOPE
ADDITIONAL SERVICES
MY ACCOUNT
CONTACT US

THE RIGHT TIME FOR THE RIGHT ACTION
(Free daily astrology horoscope and forecasts)

Snapshot (for today) forecast
Find how different aspects of your personality change depending on the natural rhythms of nature.

Did you ever wonder why hard work doesn't always pay and why things don't always go the way you want? The answer is timing! Everything in the universe works together like parts of a clock. When all the right elements align, the right outcome shows up, when they don't, no amount of hard work will produce the desired result.

CelestialTiming is a unique site that gives you a glance inside the "universal clock" and allows you to identify the right timing for interactions and events. **CelestialTiming** provides a **dynamic psychological portrait of a person** at any point in time and offers a **path to self-discovery**, self-mastery, and **success in personal and business matters**. Enjoy the site, use the information you discover, and most of all, have success in everything you do!

FORECAST OPTIONS

The site offers several **forecast options** that provide individuals with information about how different aspects of personality change depending on the natural rhythms of the universe. The forecasts can be designed for a specific point in time or for a period of time. The site uses the individual's birth time, birth place, and current location to calculate celestial rhythms and present them on a number of scales. The scales form the foundation of the forecasts.

Click on the tabs below to read about each forecast option.

Figure 6: Testing data

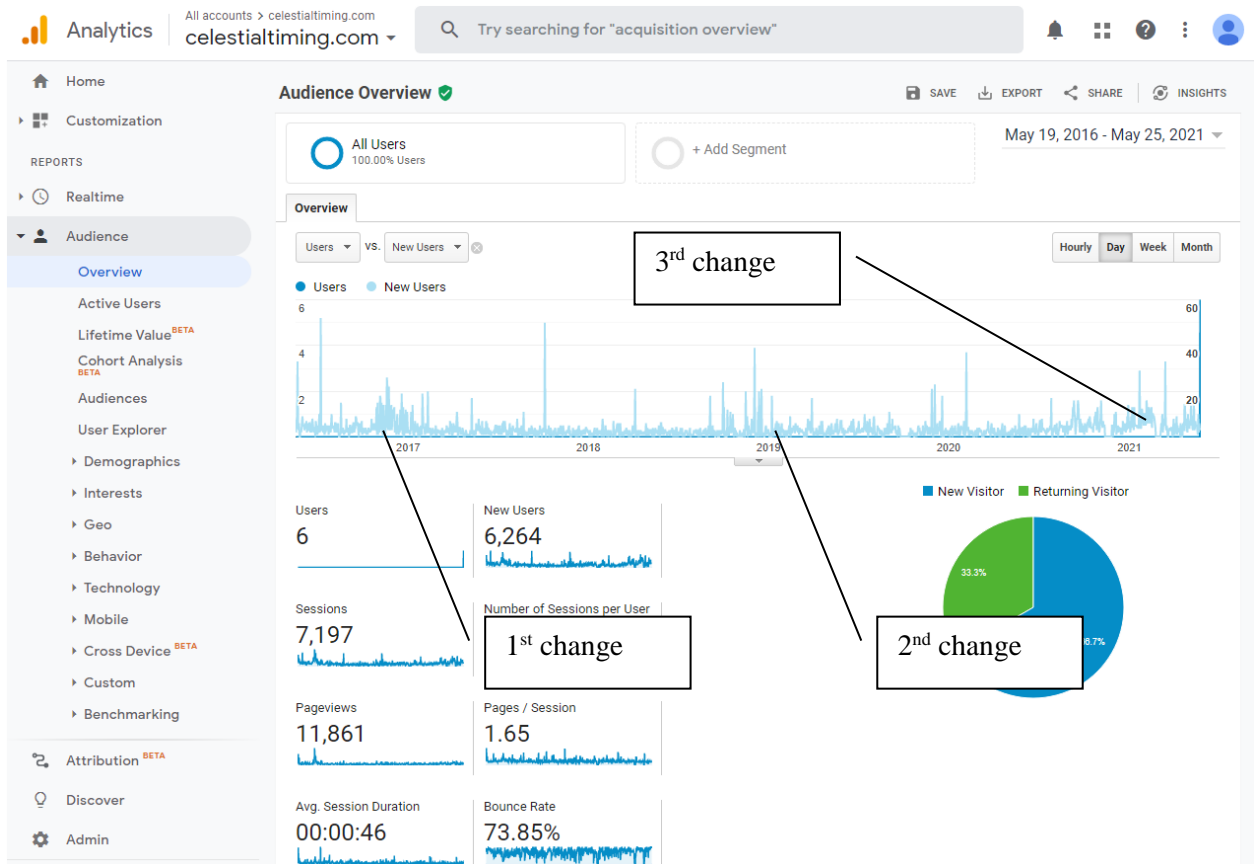


Figure 7: Google analytics data

6. Future work

Subsequent research on this topic includes:

- development of an alternative approach to the description of the semantic kernel. A promising idea is to represent the kernel in matrix form using the principles of permutations based on expert assessments of relationships between terms. In this case, the matrix contains the same elements as the search engine matrix. Then it is possible to use a vector-space model to determine the proximity of nuclei. In addition, it is promising to use a genetic algorithm to obtain an optimal semantic kernel, on a set of existing or promising ones;
- development and testing of software in the Javascript language on the WordPos platform [9]. It is required to design this software as a component with the possibility of its subsequent integration into existing content management systems such as Wordpress and Opencart;
- also a promising area of research is the analysis of profiles in social networks in order to identify semantic cores and, on their basis, search for potential buyers of a product or service.

7. References

- [1] S. Orekhov, M. Godlevsky, O. Orekhova, Theoretical fundamentals of search engine optimization based on machine learning in: Proceedings of the 13th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, ICTERI '2017, CEUR-WS, 2017, Volume 1844, pp. 23-32
- [2] Y. Goldberg. Neural Network Methods in Natural Language Processing (Synthesis Lectures on Human Language Technologies). Morgan&Claypool Publ. USA, 2017
- [3] S. Orekhov, H. Malyhon, T. Goncharenko, Using Internet News Flows as Marketing Data Component in: Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems. Volume 1: Main Conference, COLINS '2020, CEUR-WS, 2020, Volume 2604, pp. 358-373
- [4] M. Berry. Survey of Text Mining: Clustering, Classification and Retrieval. Springer, USA, 2004
- [5] E. Krupka, N. Tischby. Generalization from Observed to Unobserved Features by Clustering. Journal of Machine Learning Research, 2008, Volume 9, pp. 339-370
- [6] A. Jain, M. Murty, P. Flynn. Data clustering: A review. ACM Computing Surveys, 1999, Volume 31, No. 3, pp. 264-323
- [7] B. Rumpe. Agile modeling with UML. Springer, Germany, 2017
- [8] Web content source: Psychological self portrait is the path of self discovery based on celestial timing / Celestialtiming.com, 2021
- [9] J. Lengstorf, K. Wald. Pro PHP and jQuery. APress, USA, 2016