Textual Content Categorizing Technology Development Based on Ontology

Vasyl Lytvyn[0000-0002-9676-0180]1, Victoria Vysotska[0000-0001-6417-3689]2,
Bohdan Rusyn[0000-0001-8654-2270]3, Liubomyr Pohreliuk[0000-0003-1482-5532]4,
Pavlo Berezin[0000-0003-1869-5050]5, Oleh Naum[0000-0001-8700-6998]6

[1,2,5]Lviv Polytechnic National University, Lviv, Ukraine
[1]Silesian University of Technology, Gliwice, Poland
[3-4]Karpenko Physico-Mechanical Institute of the NAS Ukraine
[6]Ivan Franko Drohobych State Pedagogical University, Drohobych, Ukraine
Vasyl.V.Lytvyn@lpnu.ua[1], Victoria.A.Vysotska@lpnu.ua[2],
rusyn@ipm.lviv.ua[3], liubomyr@inoxoft.com[4], pavlo.berezin@gmail.com[5],
oleh.naum@gmail.com[6]

Abstract. The methods and means of using ontologies within systems for the categorization of textual content were created. Also, a method for optimizing the definition of which rubrics best relate to a certain text content was developed. The intellectual system that will use the methods developed earlier, as well as other research results was implemented. The results will allow users to easily filter their text content. The system developed has intuitive user interface.

Keywords: ontology, content, text categorization, text content, information technology, computer science, intellectual system, intelligent system, text content categorization, user interface, text document, text classification, information system, machine learning algorithm, user unfriendly interface, content categorization system.

1 Introduction

The task of content analysis is becoming more and more relevant in connection with the rapid growth of the popularity of the World Wide Web, as well as the exponential increase in the amount of content inside the network [1]. Priority is information that is intended for a person. It is for these reasons that the automation of categorization of the text is an important task [2]. The main problem of categorization manually is the considerable time and effort of the person who conducts it [3]. Also, the challenging is the unification of the categories to which the text content belongs. Categorization automation solves these problems by [4]:

1. Simplifying the search for the required information;
2. Unification of the categories;
3. Improving the understanding of the content;
4. No need for human intervention in text categorization.

The aim of work is designing and developing a system of text categorization. The following structure of research was created to achieve the goal of work:

1. Research methods of constructing a system of text categorization.
2. Research on ontology languages.
3. Analysis of the finished decisions in the field of categorized text.
4. Search and analysis of existing ontologies.
5. Analysis of machine learning algorithms.
6. Creating an algorithm for determining relevant categories in text content.

Object of research is the process of creation of intellectual systems of text categorization, the main purpose of which is convenient and qualitative classification of the text content. The subject of the study is means, methods and ways of developing intelligent systems of text content text categorization using an ontological approach. The main requirement of the system is to eliminate the need for manual categorization of the text. The second requirement is to create a user-friendly system for text categorization. The final result is a system that will allow users to quickly and accurately categorize text content. Expected results of the development of such a system are:

1. Development of methods and tools for using ontologies in the text categorization systems.
2. Development of methods for optimizing the definition and improvement of the relevance of the categorization to the text.
3. Development of a system that would use both existing and innovative methods of categorization.
4. Development of user-friendly and clear user interface of the system.

2 Key concepts analysis

In order to understand the design and development of the intellectual system, the definition of this term should be specified. So, under the term «intellectual system» means a technical or software system that is capable of solving tasks that are considered to be traditional. These tasks belong to a specific subject area, knowledge of which is stored in the memory of such system. An intelligent system usually includes three basis blocks: knowledge base; mechanism of decision making; intelligent interface. Intellectual system can be understood as intelligent information systems with intellectual support that solves the problem without the participation of the decision maker, in contrast to the intellectualized system in which the operator is present [5].

If we analyze the main methods of content categorization, namely text, one can conclude that one of the most successful methods of categorization is methods used by ontologies. It is also worth noting that these systems, for the most part, do not use the full spectrum of opportunities and benefits of ontologies, which provides great prospects for future developments in this direction. If to compare ontologies with other methods of constructing knowledge bases, you can easily see the benefits of the first one. Ontology is a standard of knowledge engineering, which has proven itself as one of the best methods for representing objective knowledge [6].

However, in the field of ontologies there is a set of unresolved problems, the solution of which will allow the development of fast and efficient systems for working out the text, namely its categorization. The list of such tasks includes [7]:

1. Tasks of the criteria for filling and optimizing ontologies;
2. Modeling processes for processing information resources and the emergence of new knowledge based ontologies;
3. Assessment of the novelty of ontology knowledge.

Under the formal model of ontology, *O* understand the three of the following form:

$$O = <C, R, F>, \tag{1}$$

where *C* is a finite set of concepts (concepts, terms) of the domain (PO), which is given by the ontology of *O*; $R: C \rightarrow C$ - a finite set of relations between concepts (concepts, terms) of a given software; *F* - finite set of functions interpretation (axiomatization, restrictions) defined on concepts or ways ontology *O*. It is worth noting that the set *C* is finite and non-empty, and *F* and *R* must be finite [8]. To improve the system's results, it is necessary to extend the ontology to the following form [9]:

$$O = <C, R, F, W, L>, \tag{2}$$

where *W* is the importance of the concepts *C*, and *L* is the importance of the relationship *R*. Such an expansion improves the system's results at times, since categories the importance of meeting a particular concept differ for different categories. The same is with the concepts: the connection between them may have different meanings [10].

In most cases, the graph is often used to provide ontologies (often a conceptual graph). In the graph, vertices are the concept of software, and arcs are the relation between different concepts. Depending on whether the axioms of the concepts are defined, the vertices are divided into interpreted and not interpolated. The arches (relation) can be vertical or horizontal. With the help of vertical arcs, the taxonomy of the concepts of software is given. The requirement of horizontal arcs is to determine the set of values and the area of definition of relations. In general, the structure of the ontology can be defined by four categories of elements: concept; relation; axioms; attributes. Concepts (classes) is general categories are organized according to the hierarchy. The class can be considered as the union of all representatives of a certain entity. That is, each class describes a group of entities united by common properties. Defining which class belongs to one class is one of the most common tasks in systems that use an ontological approach. Such a task is called categorization. Aattributes are an ontology element that represents a certain class. It is a specific element that belongs to one category. In elements of ontology there is a specific hierarchy. At the lowest level, there are specific representatives (attributes). There are categories above the instances. Above them are the relation between categories. In the top of hierarchy there are axioms and rules that combine all these steps. Below is a schematic representation of this hierarchy (Fig. 1). In order to build ontological model, first of all, it is necessary to define the hierarchy of concepts (set *C*). Also, during the construction of ontology, as an infological model, experts in subject areas should participate. For qualitative ontology construction it is necessary that these specialists skillfully use abstraction and combination. Also, when constructing an ontology, it is necessary to construct atomic concepts from a set of differential attributes. For convenience, when constructing ontologies, classification is often used. The classification is the method of streamlining knowledge. Using classification approach, we divide objects into groups based on their structure and behavior. In object-oriented analysis, by means of determining the general properties of objects, the simplicity of the architecture of the model system is achieved. It is because of the simplicity of the infological model of the system that key mechanisms and abstractions are easy to find. In modern studies, it is considered that there is no ideal hierarchy of classes, as well as the correct classification of objects. Since there are no strict methods and rules for classification of objects and their classes. This is due to the fact that is a compromise solution to choose classes with which the system will operate [11].

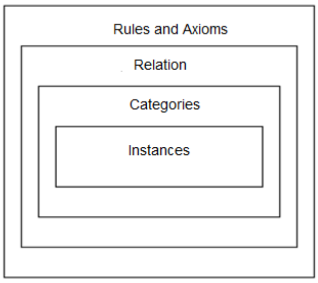


Fig. 1. Hierarchy of elements of ontology.

It can be assumed that there is no system for text content categorization that does not use parsing or keyword typing. The key is a word, or a stable expression of the natural language, which expresses some aspect of the content of the document; a word that contains an important semantic load. Such a word can act as a key when searching for information. Parsing is a process in which the input sequence of characters is analyzed. The purpose of parsing is to break up the grammatical structure according to the prescribed formal grammar. In this analysis, the content becomes a kind of data structure. Often, as a data structure, a tree is used that repeats the syntactic structure of the input data. This is due to the fact that such a structure is very well suited for further processing. Stemming is algorithms work by cutting off the word to the base. This is achieved by rejecting the suffix, ending, and other auxiliary parts of the word. Although the results of the sedation, often reminiscent of the root of the word, the emulation algorithms are based on other principles than algorithms for determining the root of the word. Therefore, the word after stamping may be different from the morphological root of the word. In the tasks of information retrieval and linguistic morphology, the process of sedation is often used.

The stemming algorithms contain the 2 most common problems [12]:

1. Over-stemming is when two words with different stems are stemmed to the same root. This is also known as a false positive.
2. Under-stemming is when two words that should be stemmed to the same root are not. This is also known as a false negative.

Stemming algorithms attempt to minimize each type of error, although reducing one type can lead to increasing the other [13].

3 Recent researches and publications analysis

After analyzing modern works [4-8, 21-27], we can sum up that developments in the field of construction and use of ontologies are actively improving. However, it is worth noting that there are very little researches on the use of ontologies in decision-making systems. In such systems, ontologies will help to make optimal decisions, since they will allow better processing of information resources of the domain area of the system. The work [14, 28-32] conducts the review and solution of the problem of searching for methods for developing and processing resources for intelligent Internet systems. Such methods will allow the development of such software tools, which greatly facilitate the development, distribution and development of Web-systems. These methods were developed as a result of the analysis of features, patterns and dependencies in the processing of information resources of e-business systems. It can also be seen that scientific research on this topic is rather local [20-33]. This creates a certain contradiction, as the development of IT and related fields is very fast, and a small amount of scientific works points to a number of problems in scientific circles. As a result, there is a delay in the development of this direction, because of the lack of theoretical information that leads to the problems of those people who are engaged in a practical part of the research. This, in turn, can lead to a situation in which, due to the lack of development of the field, specialists will no longer use ontologies in their systems. Although ontologies are a very promising direction in certain classes of tasks. The end of the twentieth century became the beginning of scientific research in the field of practical use of ontologies in the design and operation of information systems. Studies on this topic are actively under way. Formal mathematical models of ontologies and their basic theoretical foundations were developed in works by [15]:

1. T. Gruber, A. Pérez-Gómez, J. Salton, who proposed to consider ontologies in the form of a three-dimensional tuple.
2. N. Guarino, M. Shambard, P. Folts, who were looking for ways to refine ontologies and developed methods for constructing them.
3. J. Sowa, who first used and proposed the use of a conceptual graph.
4. M. Montes-Gomez, who for the first time used a conceptual graph for image ontologies.
5. K. Jones, J.V. Rogushina, E. Mena, A. Kaufmana, R. Knappa, M. Boris, A. Kalli, M. Yu. Uvarova, I.P. Norenkov, who investigated the use of ontologies in functioning systems.
6. T. Andreasen, T. Berners-Lee, O. Lassila, D. Handler, who investigated the problems of building intelligent systems that would be based on ontologies.
7. O.V. Palagina, A.V. Anisimova, A. Gladun, who offered methods and ways of processing the Ukrainian language.

After analyzing the scientific studies of foreign and our scientists, one can conclude that in the area of processing of information resources, such aspects as assessing the quality of ontology, extracting knowledge from heterogeneous sources and developing methods of integration between ontologies are the main ones in the direction of research ontologies. Modern areas of researches that are related to the teaching of ontologies, as well as their practical use in intelligent information systems, are:

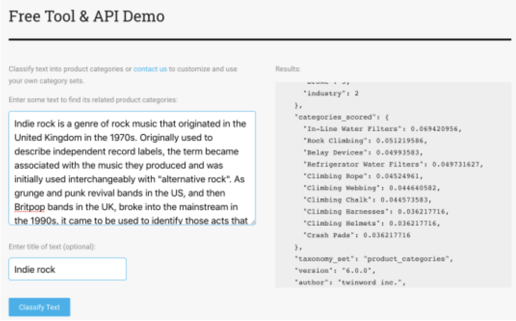
1. Learning ontologies based on the analysis of texts in the natural language [2, 11-13].
2. Methods and ways of using ontologies in the construction and use of decision-making systems [14].
3. Application development that would allow us to conveniently develop ontologies manually, or develop them automatically (Ontosaurus, OntoEdit, Protégé) [15-17].
4. The solution of practical problems, which are based on requests to knowledge bases, using ontologies [2, 18].
5. Creating and improving ontology description languages (RDF, OWL, XML, DAML + OIL) [19].

4 Finished software products analysis

A thorough search of sites with the possibility of the text categorization was carried out for the analysis of finished software. After searching for such sites, one can conclude that the text classification market has few rivals for the system under development [34-42]. In spite of the fact that there are several libraries for categorizing, as well as several open APIs that include the possibility of classifying the text, full-fledged systems that would allow the simple user to thoroughly refine your text. There are also several proprietary solutions, but they are quite expensive and for a user, the price is not entirely justified. All found products face a number of challenges: user-unfriendly interface; access to only one language; lack of possibility of saving the result; impossibility to load text in a file. The following systems will be considered: TwinWord, uClassify and Ailien. The first found software product is TwinWord, which almost immediately showed a number of problems for ordinary users. The first problem is the availability of only one language (English). Also, the accuracy of categorizing is rather low, on the test text among the 10 categories proposed by the system, none of them answered the subject of the text or its keywords. And the last, but not the slightest problem of this system is a user-unfriendly interface (Fig. 2). The next system is uClassif». This system showed itself not much better than the previous one. It is also available in English only. The quality of categorizing is better, but the results were still far from expected. As a result of the classification of this system, it can be understood that it is more based on keywords. Also, the common problem with the previous system is user-unfriendly interface (Fig. 3). The last system is Aylien. This system is much better than the previous two and can be considered as the best one. Although there are no other languages than English, it gives users a taxonomy to select the text to be categorized. There are two taxonomy to choose from:

1. IPTC News Codes - The International Press Telecommunications Council for News Categorization;
2. IAB QAG - Quality Assurance Guidelines from the Interactive Advertising Bureau (IAB).

The results of the categorizing were also better than in the previous systems and were fairly accurate. Also, this system has a much friendlier user interface (Fig. 4-5).



Free Tool & API Demo

Classify text into product categories or contact us to customize and use your own category sets.

Enter some text to find its related product categories:

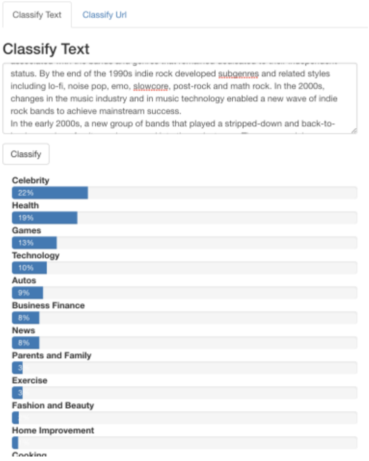
Indie rock is a genre of rock music that originated in the United Kingdom in the 1970s. Originally used to describe independent record labels, the term became associated with the music they produced and was initially used interchangeably with "alternative rock". As grunge and punk revival bands in the US, and then Britpop bands in the UK, broke into the mainstream in the 1990s, it came to be used to identify those acts that

Enter title of text (optional):

Indie rock

Results:

```
    "industry": 2
  },
  "categories_scored": {
    "In-Line Water Filters": 0.069420956,
    "Rock Climbing": 0.051219586,
    "Belay Devices": 0.04993583,
    "Refrigerator Water Filters": 0.049731627,
    "Climbing Rope": 0.04524961,
    "Climbing Webbing": 0.044640582,
    "Climbing Chalk": 0.044573583,
    "Climbing Harnesses": 0.036217716,
    "Climbing Helmets": 0.036217716,
    "Crash Pads": 0.036217716
  },
  "taxonomy_set": "product_categories",
  "version": "6.0.0",
  "author": "twinword inc.",
```

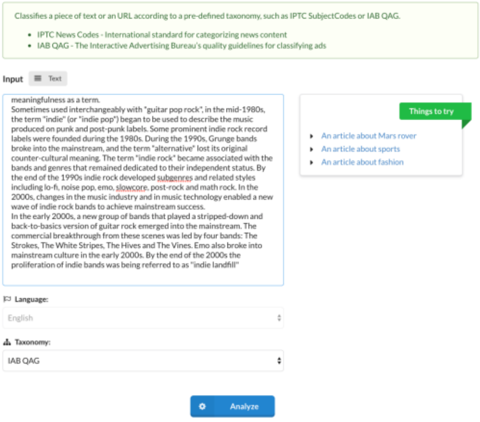


Classify Text

status. By the end of the 1990s indie rock developed subgenres and related styles including lo-fi, noise pop, emo, slowcore, post-rock and math rock. In the 2000s, changes in the music industry and in music technology enabled a new wave of indie rock bands to achieve mainstream success.
In the early 2000s, a new group of bands that played a stripped-down and back-to-

Celebrity
22%
Health
19%
Games
13%
Technology
10%
Autos
9%
Business Finance
8%
News
8%
Parents and Family
3
Exercise
3
Fashion and Beauty
Home Improvement
Cooking

Fig. 2. The user interface of TwinWord service **Fig. 3.** User interface for uClassify service



Classifies a piece of text or an URL according to a pre-defined taxonomy, such as IPTC SubjectCodes or IAB QAG.

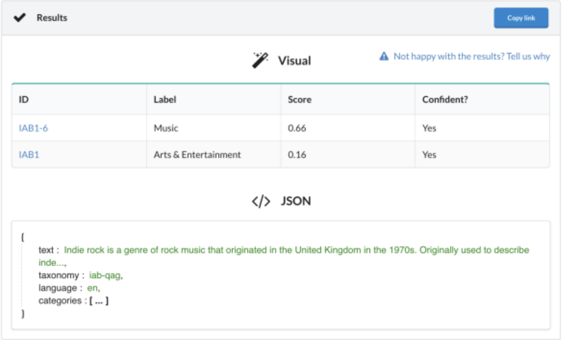
- IPTC News Codes - International standard for categorizing news content
- IAB QAG - The Interactive Advertising Bureau's quality guidelines for classifying ads

meaningfulness as a term.
Sometimes used interchangeably with "guitar pop rock", in the mid-1980s, the term "indie" (or "indie pop") began to be used to describe the music produced on punk and post-punk labels. Some prominent indie rock record labels were founded during the 1980s. During the 1990s, Grunge bands broke into the mainstream, and the term "alternative" lost its original counter-cultural meaning. The term "indie rock" became associated with the bands and genres that remained dedicated to their independent status. By the end of the 1990s indie rock developed subgenres and related styles including lo-fi, noise pop, emo, slowcore, post-rock and math rock. In the 2000s, changes in the music industry and in music technology enabled a new wave of indie rock bands to achieve mainstream success.
In the early 2000s, a new group of bands that played a stripped-down and back-to-basics version of guitar rock emerged into the mainstream. The commercial breakthrough from these scenes was led by four bands: The Strokes, The White Stripes, The Hives and The Vines. Emo also broke into mainstream culture in the early 2000s. By the end of the 2000s the proliferation of indie bands was being referred to as "indie landfill"

Language: English

Taxonomy: IAB QAG

Fig. 4. Input field in the Aylien system



ID	Label	Score	Confident?
IAB1-6	Music	0.66	Yes
IAB1	Arts & Entertainment	0.16	Yes

```
{
    text : Indie rock is a genre of rock music that originated in the United Kingdom in the 1970s. Originally used to describe
    inde...,
    taxonomy : iab-qag,
    language : en,
    categories : [ ... ]
}
```

Fig. 5. Results window in the Aylien system

5 System Analysis of the Study Object

The main purpose of the system is the text content categorization. The main objectives of this system are the user-friendly interaction with the system, and qualitative categorization/clasification of the content. In order to achieve each of these objectives, we will break them down. This partition will look like this:

1. Interaction with users.
   - Authorization
   - Sending request for categorizing.
   - Reviewing categorized/classified articles.
   - Filtering articles by sections.
2. Categorization content.
   - Processing a categorization request.
   - Searching key essentials.
   - Analyzing found key entities.
   - Defining categories for articles.
   - Optimizing weights for categories.

Fig. 6 shows a class diagram for the design of a system. On Fig. 7 the Use case diagram for the text content categorization system is depicted. Fig. 8 shows the activity diagram of the intellectual system.

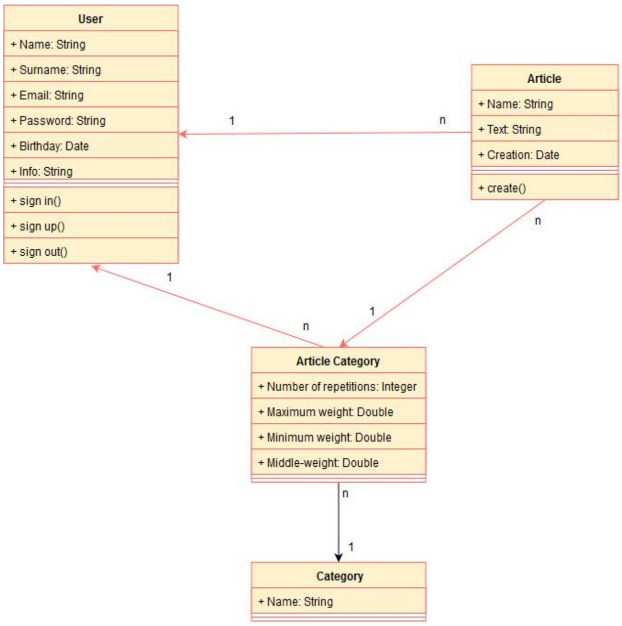


Fig. 6. Class Diagram

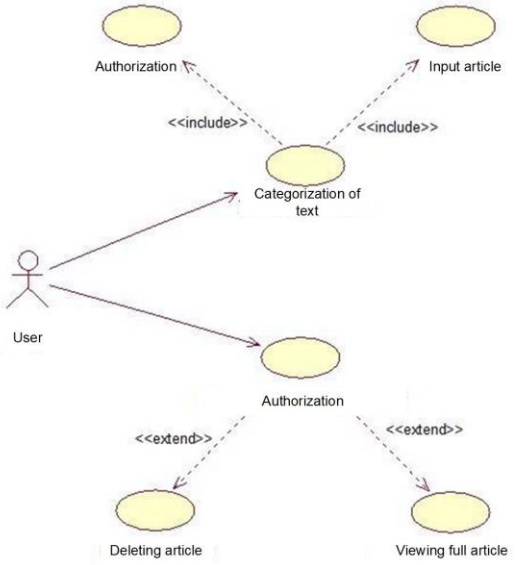


Fig. 7. Use case diagram

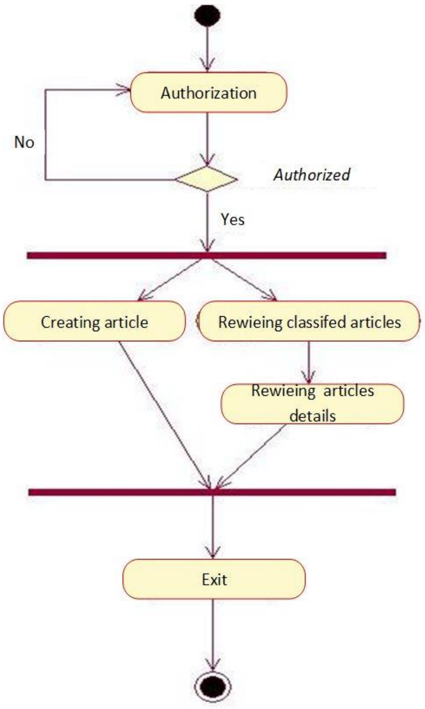


Fig. 8. Activity Diagram

In the text content categorization system an object that changes its state over time is the article itself. Fig. 9 shows the state diagram of the article.

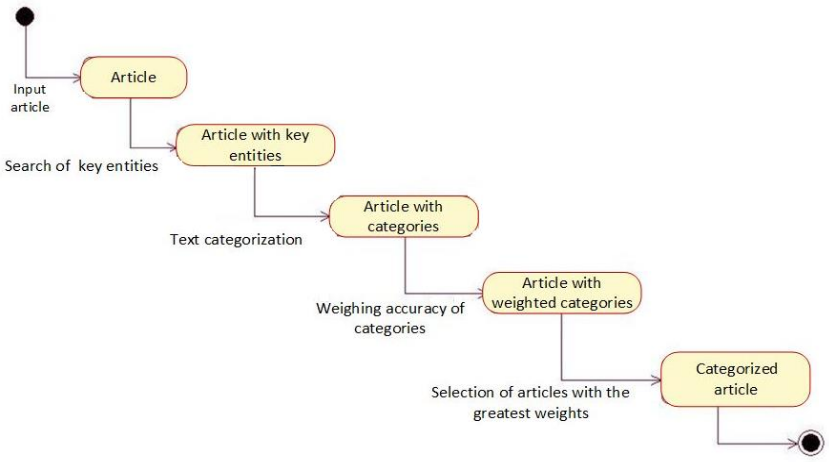

Article
Input article
Article with key entities
Search of key entities
Article with categories
Text categorization
Article with weighted categories
Weighing accuracy of categories
Categorized article
Selection of articles with the greatest weights

Fig. 9. State diagram of an article

A sequence diagram shows a set of messages arranged in time sequence. Also, this diagram depicts which messages are transmitted between objects during an action. In UML, when constructing such a diagram, processes and objects are represented in the form of vertical lines, and messages between them are represented as horizontal lines. Messages must be ordered at the time of departure. Fig. 10 shows the sequence diagram for the text content categorization system.

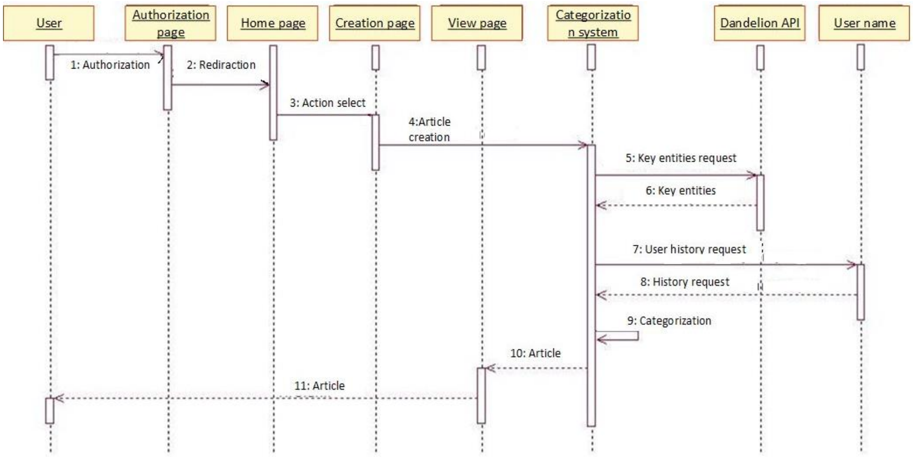


Fig. 10. A sequence diagram

Since the diagram as a result of such association is rather voluminous, in order not to confuse people who will work with this diagram, the actions are numbered. The numbering begins with 1 and continues on the message movement in the system. Collaboration diagram is one of the most comprehensive. This diagram is most commonly used by system designers since they can see the overall picture of the system they are developing. Fig. 11 shows a collaboration diagram.

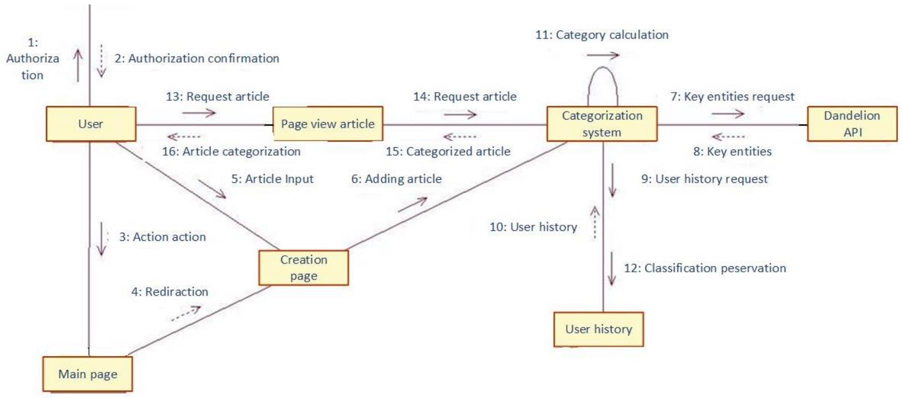


Fig. 11. Collaboration Diagram

The latest diagram was a deployment diagram. Just like the previous diagram, this diagram is more technical. As a component diagrams, it is very popular among software developers, especially among DevOps. Because it is their competence to start and configure the system's nodes. This diagram can also be used in software development, in order to outline which nodes are needed by the system, which can facilitate the development, since no extra work will be done. Fig. 12 shows the deployment diagram of the intellectual system of categorization of the text content.

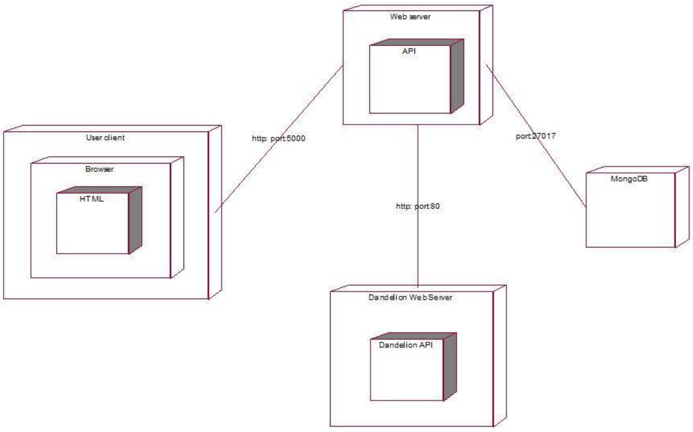


Fig. 12. Deployment Diagram

6 Statement and justification of the problem

As described in the previous section, there are already systems that would implement the possibility of categorizing text content. There are different types of systems, both public and closed systems, access to which must be purchased. However, none of the systems can save results, although they can be used to improve the results. That is why it should be used to improve the competitiveness of the system. The first and most obvious problem is the user interface. Many existing systems have a big problem in this direction. And this should be used in the development of this system. The user interface is a very important part of the system, since it makes no sense to design a system that nobody will use. The next problem is the ability for users to save their articles. This can improve the system time, since it will not be necessary to rearrange articles. And if the user loses the result of the previous categorizing, he will always be able to view the previous results. The third problem also applies to the preservation of information. We can use the history of categorizations to improve its quality. Knowing to which categories are the user's articles belongs, we can improve the accuracy of classification. That is, if a user often writes on a single topic, then the possibility that his new article belongs to this topic is much bigger than that the article belongs to a completely new topic for the user. To solve this problem, you can connect a machine learning algorithm that could analyze the user's history.

7 Software product description

For full operation of the system, a permanent connection with the server part of the system is required. Also, on the server should be a mandatory access to the Internet, because the server uses external resources. The system is based on the MERN stack using the JavaScript programming language. The user interface is a web page where the user can log in and take advantage of the system.

The purpose of the system is to automate the classification of the text content. The classification problem is quite acute in modern realities, since manual classification is a time-consuming process, and correctly selected columns improve both SEO and user-friendliness of the system. The system being developed uses the Dandelion API. Dandelion API - an open API designed for text analysis. In this system, Dandelion API is used to get the essence of the text. The entity that returns this API is part of the DBPedia ontology. Dandelion API is a multilingual API. It supports a stable analysis of 7 languages and more than 50 languages are in active development. DBPedia is a crowd-sourced community effort to extract structured content from the information created in various Wikimedia projects. Most systems of text classification have a limited number of available classifications, but the system developed thanks to DBPedia will have almost the maximum available at the moment the list of classifies.

The main part of the whole process of analyzing the text on the server is to process the essentials derived from DandelionAPI. For large text, DandelionAPI returns many entities among which the most important ones to be identified, as well as determine which sections they belong to. The solution to the task involves the following steps:

1. Definition of the language of the text.

2. Splitting the text into pieces.
3. Configure Dandelion API request.
4. Sending Dandelion API request.
5. Getting results from the Dandelion API.
6. Rejecting entities with low confidence.
7. Finding and removing entities that are alternate names of other entities.
8. Obtaining categories from entities.
9. Counting the number of times each of the categories met.
10. Finding the maximum weight for each category.
11. Finding the weight of each category.
12. Correction of weight of each category according to the user's history.
13. Sifting categories with low weights.
14. Saving categories to user history.
15. Saving categories as clasifications for text content.

The problem text classification is relevant because none of the top sites for blogs / articles does not offer automatic categorization of text, which means that users or moderators of the data need to manually categorize content on each site. Content classification is an important function of such a resource, as it will simplify the navigation, search and filtering content to users. The advantages of the system are:

- Multimodality.
- No need for support.
- Automation of categorizing process.
- No constraints on the subject of convent.
- Ability to improve the system.

Data in the system is divided into 3 categories:

1. Inputs, which the user enters.
2. Input that is the result of a request to the Dandelion API.
3. Output data, which the system generates.

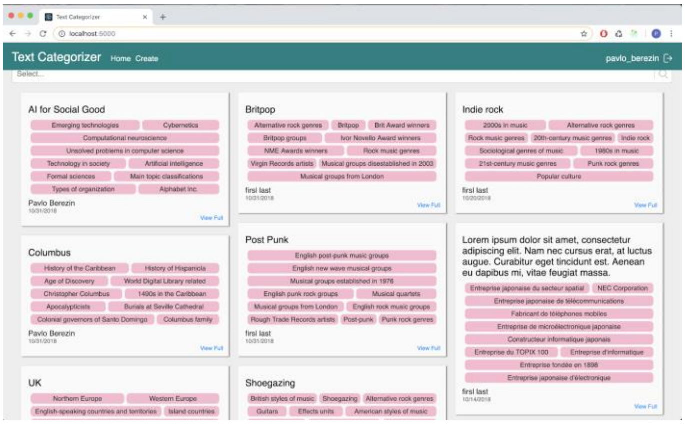
Inputs made by the user are user data, as well as text content for classification. The input, that is the result of a request to the Dandelion API - is the data received by the system after the analysis of the user's text from the Dandelion API. Output is classified articles that are received upon a user request. Integrated data is data that affect the operation of the algorithm. In this system it is the history of user’s categories.

8 User guide

The «Text Categorizer» is a website. To access to the website Internet is needed. The website is available on both personal computers and mobile devices. The target audience of this application is people who generate or consume text content. All articles and their classifications are publicly available on the site. To work with the system as the user, it is only necessary to fill in all the information about the article (name, text), after which the system will automatically select the category’s. The user does not need any actions to do that. Due to the high automation of the process, there is only one type of user in the system: Author. To work on the Author system you need to be authenticated. After passing authentication, the author will be able to: view the classified articles; filter articles by categories; categorize his own articles.

To access the application, the user must first go to the site address, then authenticate to the system. After passing the authentication process, the user will go directly on the main page of the application (Fig. 13). On the main page, the user can see the cards of the classified articles. Each card contains the main information about the article: title, author, date of creation, category. At the top of all pages of the system, the user can see the header. The left-handed menu consists of the following buttons:

- Home is click below to be taken directly to the homepage.
- Create is click below to be taken directly to the creation page.



AI for Social Good

Emerging technologies | Cybernetics | Computational neuroscience | Unsolved problems in computer science | Technology in society | Artificial intelligence | Formal sciences | Main topic classifications | Types of organization | Alphabet Inc.

Pavlo Berezin
10/31/2018

Britpop

Alternative rock genres | Britpop | Brit Award winners | Britpop groups | Ivor Novello Award winners | NME Awards winners | Rock music genres | Virgin Records artists | Musical groups disestablished in 2003 | Musical groups from London

first last
10/31/2018

Indie rock

2000s in music | Alternative rock genres | Rock music genres | 20th-century music genres | Indie rock | Sociological genres of music | 1980s in music | 21st-century music genres | Punk rock genres | Popular culture

first last
10/20/2018

Columbus

History of the Caribbean | History of Hispaniola | Age of Discovery | World Digital Library related | Christopher Columbus | 1490s in the Caribbean | Apocalypticists | Burials at Seville Cathedral | Colonial governors of Santo Domingo | Columbus family

Pavlo Berezin
10/31/2018

Post Punk

English post-punk music groups | English new wave musical groups | Musical groups established in 1976 | English punk rock groups | Musical quartets | Musical groups from London | English rock music groups | Rough Trade Records artists | Post-punk | Punk rock genres

first last
10/31/2018

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam nec cursus erat, at luctus augue. Curabitur eget tincidunt est. Aenean eu dapibus mi, vitae feugiat massa.

Entreprise japonaise du secteur spatial | NEC Corporation | Entreprise japonaise de télécommunications | Fabricant de téléphones mobiles | Entreprise de microélectronique japonaise | Constructeur informatique japonais | Entreprise du TOPIX 100 | Entreprise d'informatique | Entreprise fondée en 1898 | Entreprise japonaise d'électronique

first last
10/14/2018

UK

Northern Europe | Western Europe | English-speaking countries and territories | Island countries

Shoegazing

British styles of music | Shoegazing | Alternative rock genres | Guitars | Effects units | American styles of music

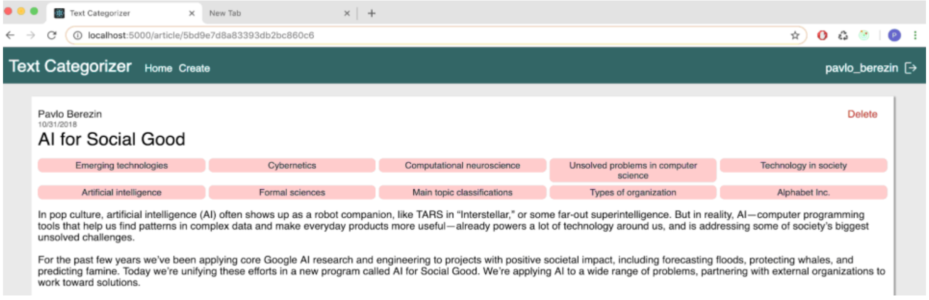
Fig. 13. Application Homepage

On the right side, the user can see under which username he or she is locked, as well as the exit button. In the bottom right corner of each card there is a button «View Full», which allows you to go to the page detailed review article (Fig. 14). On the page's detailed article view, you can see the full article. In the upper right corner of the page is a button «Delete», which function is to delete the article from the system. Also, the user has the option to click on one of the categories of the article and return to the main page where the category to which the user has clicked will be included in the article search filter.

There is an option to add a category to the filter list by using several methods:

- When moving from a detailed view page;
- Clicking on the category on the article card on the main page;
- Entering a category in the search box input.

When entering the category in the search box, the user can enter any value. However, one can select only one of the existing categories (Fig. 15). During updating of the list of categories, on the main page are filtered and are appeared only selected classifications. (Fig. 16).



Pavlo Berezin
10/31/2018

AI for Social Good

Emerging technologies	Cybernetics	Computational neuroscience	Unsolved problems in computer science	Technology in society
Artificial intelligence	Formal sciences	Main topic classifications	Types of organization	Alphabet Inc.

In pop culture, artificial intelligence (AI) often shows up as a robot companion, like TARS in “Interstellar,” or some far-out superintelligence. But in reality, AI—computer programming tools that help us find patterns in complex data and make everyday products more useful—already powers a lot of technology around us, and is addressing some of society’s biggest unsolved challenges.

For the past few years we’ve been applying core Google AI research and engineering to projects with positive societal impact, including forecasting floods, protecting whales, and predicting famine. Today we’re unifying these efforts in a new program called AI for Social Good. We’re applying AI to a wide range of problems, partnering with external organizations to work toward solutions.

Fig. 14. Detailed article view page

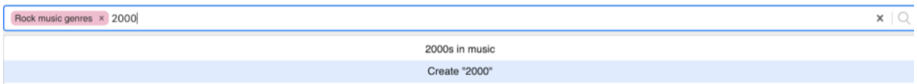
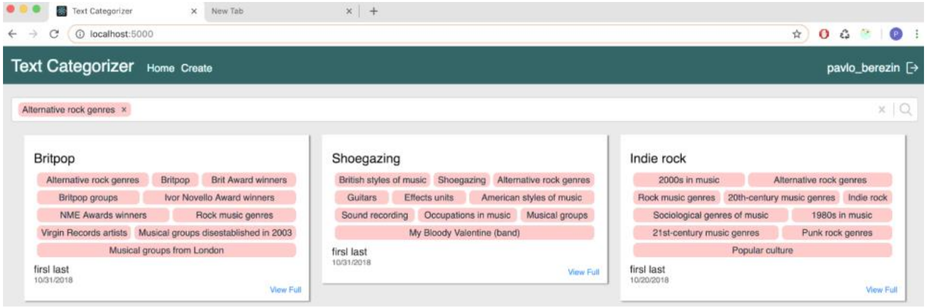


Fig. 15. Input field for filtration



Britpop

Alternative rock genres · Britpop · Brit Award winners · Britpop groups · Ivor Novello Award winners · NME Awards winners · Rock music genres · Virgin Records artists · Musical groups disestablished in 2003 · Musical groups from London

firsl last
10/31/2018
View Full

Shoegazing

British styles of music · Shoegazing · Alternative rock genres · Guitars · Effects units · American styles of music · Sound recording · Occupations in music · Musical groups · My Bloody Valentine (band)

firsl last
10/31/2018
View Full

Indie rock

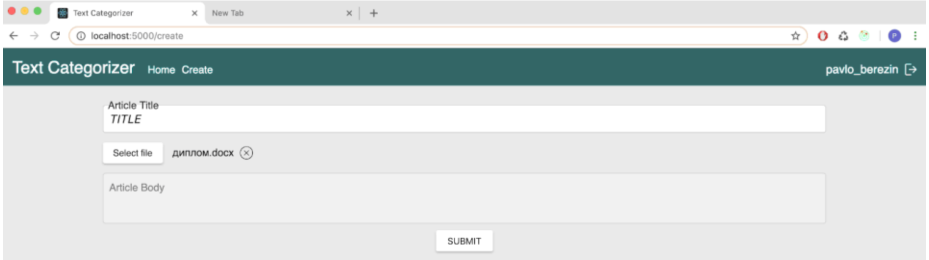
2000s in music · Alternative rock genres · Rock music genres · 20th-century music genres · Indie rock · Sociological genres of music · 1980s in music · 21st-century music genres · Punk rock genres · Popular culture

firsl last
10/20/2018
View Full

Fig. 16. Filtered Articles

To add an article to the system, go to the «Create» page (Fig. 4.5). This page has an article entry form. The first input field is the title of the article. Then the user has a choice:

- To download the text document by clicking the «Select file» button.
- To enter the article text manually in the «Article Body» entry field.



Article Title
TITLE

диплом.docx

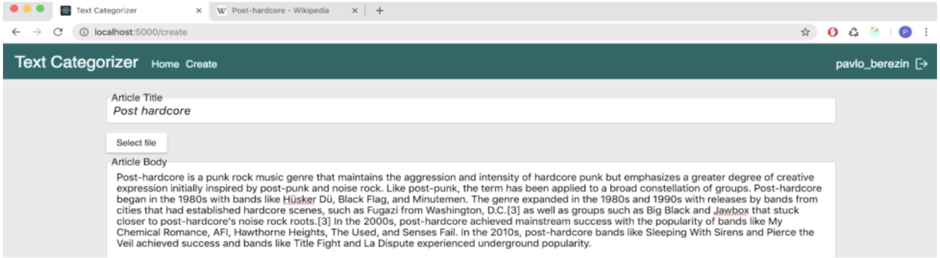
Article Body

Fig. 17. Article creation page.

Next, the user must click on the «Submit» button, after which the article will automatically be classified down and saved in the system. The user will be redirected to the main page of the site.

9 Test Case Analysis

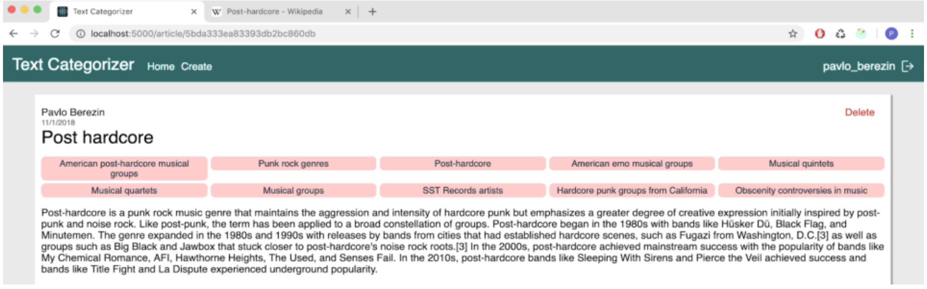
In order to check the work of the website one must consider the main types of work with the system. An important part of the developed system is the possibility of multi-language use. Therefore, it is worth checking out the work with the major world languages: English; German; Italian; French You also need to check whether it is possible to download articles as a text document. Each author will start work the same: authorization, transition to the main page, the click of a «Create» button, after which will be done the main work with the system. The system`s work was first reviewed in English (Fig. 18-19) and in German (Fig. 20-21).



Article Title
Post hardcore

Article Body
Post-hardcore is a punk rock music genre that maintains the aggression and intensity of hardcore punk but emphasizes a greater degree of creative expression initially inspired by post-punk and noise rock. Like post-punk, the term has been applied to a broad constellation of groups. Post-hardcore began in the 1980s with bands like Hüsker Dü, Black Flag, and Minutemen. The genre expanded in the 1980s and 1990s with releases by bands from cities that had established hardcore scenes, such as Fugazi from Washington, D.C.[3] as well as groups such as Big Black and Jawbox that stuck closer to post-hardcore's noise rock roots.[3] In the 2000s, post-hardcore achieved mainstream success with the popularity of bands like My Chemical Romance, AFI, Hawthorne Heights, The Used, and Senses Fail. In the 2010s, post-hardcore bands like Sleeping With Sirens and Pierce the Veil achieved success and bands like Title Fight and La Dispute experienced underground popularity.

Fig. 18. Writing an article in English

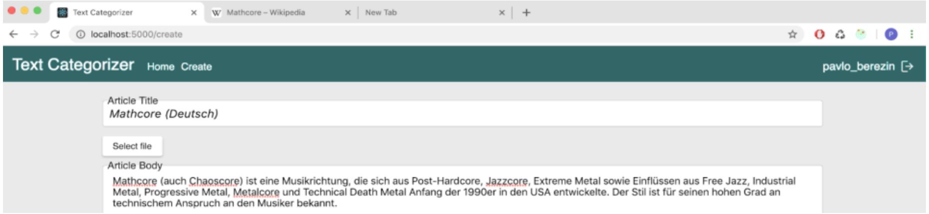

Text Categorizer
Post-hardcore - Wikipedia
localhost:5000/article/5bda333ea83393db2bc860db
Text Categorizer
Home
Create
pavlo_berezin
Delete

Pavlo Berezin
11/1/2018
Post hardcore

American post-hardcore musical groups | Punk rock genres | Post-hardcore | American emo musical groups | Musical quintets | Musical quartets | Musical groups | SST Records artists | Hardcore punk groups from California | Obscenity controversies in music

Post-hardcore is a punk rock music genre that maintains the aggression and intensity of hardcore punk but emphasizes a greater degree of creative expression initially inspired by post-punk and noise rock. Like post-punk, the term has been applied to a broad constellation of groups. Post-hardcore began in the 1980s with bands like Hüsker Dü, Black Flag, and Minutemen. The genre expanded in the 1980s and 1990s with releases by bands from cities that had established hardcore scenes, such as Fugazi from Washington, D.C.[3] as well as groups such as Big Black and Jawbox that stuck closer to post-hardcore's noise rock roots.[3] In the 2000s, post-hardcore achieved mainstream success with the popularity of bands like My Chemical Romance, AFI, Hawthorne Heights, The Used, and Senses Fail. In the 2010s, post-hardcore bands like Sleeping With Sirens and Pierce the Veil achieved success and bands like Title Fight and La Dispute experienced underground popularity.

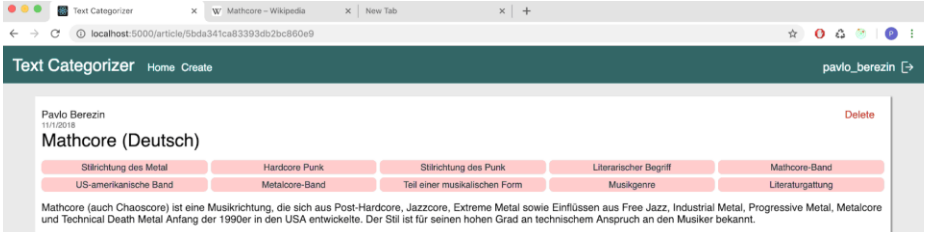
Fig. 19. View the result of categorization of English text.



Article Title
Mathcore (Deutsch)

Article Body
Mathcore (auch Chaoscore) ist eine Musikrichtung, die sich aus Post-Hardcore, Jazzcore, Extreme Metal sowie Einflüssen aus Free Jazz, Industrial Metal, Progressive Metal, Metalcore und Technical Death Metal Anfang der 1990er in den USA entwickelte. Der Stil ist für seinen hohen Grad an technischem Anspruch an den Musiker bekannt.

Fig. 20. Writing an article in German



Pavlo Berezin
11/1/2018

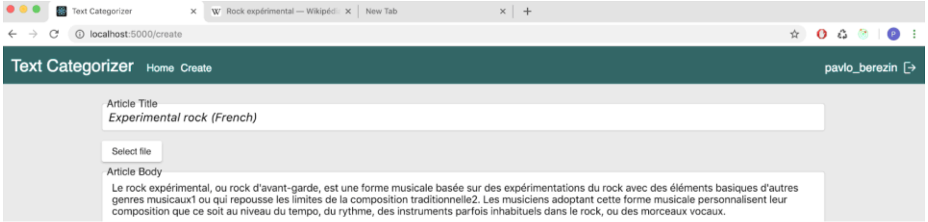
Mathcore (Deutsch)

Stilrichtung des Metal · Hardcore Punk · Stilrichtung des Punk · Literarischer Begriff · Mathcore-Band
US-amerikanische Band · Metalcore-Band · Teil einer musikalischen Form · Musikgenre · Literaturgattung

Mathcore (auch Chaoscore) ist eine Musikrichtung, die sich aus Post-Hardcore, Jazzcore, Extreme Metal sowie Einflüssen aus Free Jazz, Industrial Metal, Progressive Metal, Metalcore und Technical Death Metal Anfang der 1990er in den USA entwickelte. Der Stil ist für seinen hohen Grad an technischem Anspruch an den Musiker bekannt.

Fig. 21. View the result of the categorization of the German text

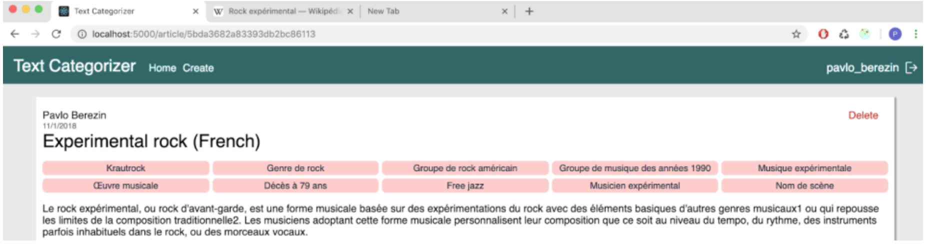
Subsequently, the article was verified in French (Fig. 22-23).



Article Title
Experimental rock (French)

Article Body
Le rock expérimental, ou rock d'avant-garde, est une forme musicale basée sur des expérimentations du rock avec des éléments basiques d'autres genres musicaux1 ou qui repousse les limites de la composition traditionnelle2. Les musiciens adoptant cette forme musicale personnalisent leur composition que ce soit au niveau du tempo, du rythme, des instruments parfois inhabituels dans le rock, ou des morceaux vocaux.

Fig. 22. Writing an article in French language


Text Categorizer
Rock expérimental — Wikipédi
New Tab
localhost:5000/article/5bda3682a83393db2bc86113
Text Categorizer
Home
Create
pavlo_berezin
Delete

Pavlo Berezin
11/1/2018

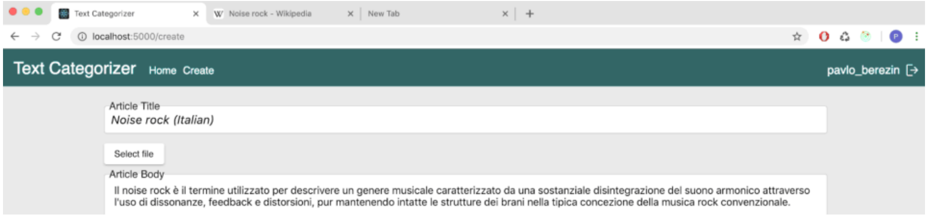
Experimental rock (French)

Krautrock · Genre de rock · Groupe de rock américain · Groupe de musique des années 1990 · Musique expérimentale
Œuvre musicale · Décès à 79 ans · Free jazz · Musicien expérimental · Nom de scène

Le rock expérimental, ou rock d'avant-garde, est une forme musicale basée sur des expérimentations du rock avec des éléments basiques d'autres genres musicaux1 ou qui repousse les limites de la composition traditionnelle2. Les musiciens adoptant cette forme musicale personnalisent leur composition que ce soit au niveau du tempo, du rythme, des instruments parfois inhabituels dans le rock, ou des morceaux vocaux.

Fig. 23. Viewing the result of the categorization of the French text

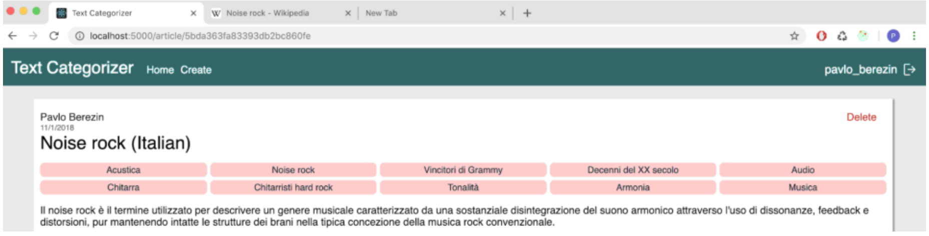
Subsequently, the article was handwritten in Italian (Fig. 24-25).



Article Title
Noise rock (Italian)

Article Body
Il noise rock è il termine utilizzato per descrivere un genere musicale caratterizzato da una sostanziale disintegrazione del suono armonico attraverso l'uso di dissonanze, feedback e distorsioni, pur mantenendo intatte le strutture dei brani nella tipica concezione della musica rock convenzionale.

Fig. 24. Writing an article in Italian



Pavlo Berezin
11/1/2018

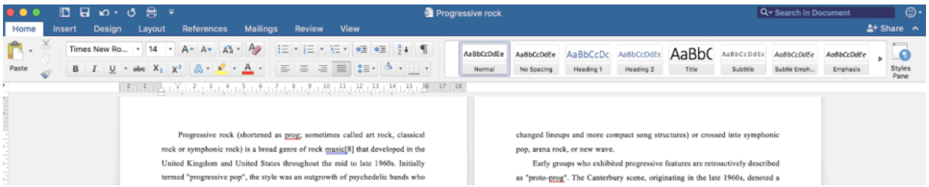
Noise rock (Italian)

Acustica	Noise rock	Vincitori di Grammy	Decenni del XX secolo	Audio
Chitarra	Chitarristi hard rock	Tonalità	Armonia	Musica

Il noise rock è il termine utilizzato per descrivere un genere musicale caratterizzato da una sostanziale disintegrazione del suono armonico attraverso l'uso di dissonanze, feedback e distorsioni, pur mantenendo intatte le strutture dei brani nella tipica concezione della musica rock convenzionale.

Fig. 25. View the result of the Italian text categorization The

The ability of downloading a document was checked (Fig. 26-28).

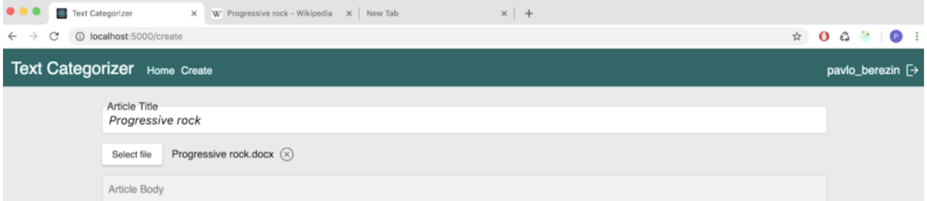


Progressive rock (shortened as prog; sometimes called art rock, classical rock or symphonic rock) is a broad genre of rock music[8] that developed in the United Kingdom and United States throughout the mid to late 1960s. Initially termed "progressive pop", the style was an outgrowth of psychedelic bands who

changed lineups and more compact song structures) or crossed into symphonic pop, arena rock, or new wave.

Early groups who exhibited progressive features are retroactively described as "proto-prog". The Canterbury scene, originating in the late 1960s, denoted a

Fig. 26. Text document

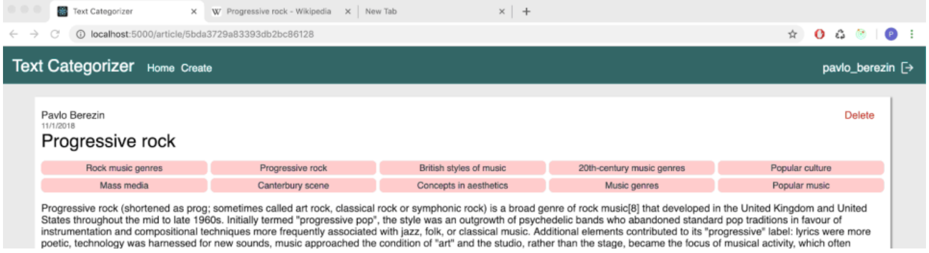


Article Title
Progressive rock

Progressive rock.docx

Article Body

Fig. 27. Loading a text document



Text Categorizer
Progressive rock - Wikipedia
New Tab
localhost:5000/article/5bda3729a83393db2bc86128
Text Categorizer
Home
Create
pavlo_berezin
Delete

Pavlo Berezin
11/1/2018

Progressive rock

Rock music genres	Progressive rock	British styles of music	20th-century music genres	Popular culture
Mass media	Canterbury scene	Concepts in aesthetics	Music genres	Popular music

Progressive rock (shortened as prog; sometimes called art rock, classical rock or symphonic rock) is a broad genre of rock music[8] that developed in the United Kingdom and United States throughout the mid to late 1960s. Initially termed "progressive pop", the style was an outgrowth of psychedelic bands who abandoned standard pop traditions in favour of instrumentation and compositional techniques more frequently associated with jazz, folk, or classical music. Additional elements contributed to its "progressive" label: lyrics were more poetic, technology was harnessed for new sounds, music approached the condition of "art" and the studio, rather than the stage, became the focus of musical activity, which often

Fig. 28. View the result of categorization of text from a document

As a result of a practical implementation, a resource has been created. The main proposes is solving the problem of automating the categorization of text content. A description of the created software product was made. Using the ER-diagrams for the Chen note, the database was described. In this section, there were also illustrated user roles in the system and their main types of interaction with the system. Test case of the main functional of the system was carried out and analyzed.

10 Conclusion

During the implementation of work, an analysis and review of literary sources were conducted, in which key concepts, recent research and ready-made software solutions of the problem were described and analyzed. As a result of this analysis, it was determined that the developed system would be successful among users. It was also determined which deficiencies of competitive systems need to be corrected in the system being developed. System was analyzed after the analysis of literary sources. In this analysis, an objective tree, a UML diagram and a system hierarchy were developed. During this stage of the system development, it was determined which goals are necessary to create the system. The next step at this stage was to determine the ways of moving data in the system. The next stage was the choice of software solutions. The MERN stack was selected as the backbone of the system. This stack was chosen due to the ability to create an isomorphic JavaScript application. After the successful determination of software products, the hardware requirements of the system under development were determined. After the previous stages, the practical implementation of the system was developed. The finished product has been described in detail. Also, during this stage of development has been described user's manual. The system was tested on a test case. The final stage of development was the analysis of the economic feasibility of the system. It was considered economically feasible and competitive.

As a result of the implementation of the graduation paper, we received a software product that provides a convenient text categorization. Anyone can access this web site and to categorize their own text through a website. The system is open to improvements through the expansion of supported languages, as well as improving the speed and quality of the definition of the categories.

References

1. Alipanah, N., Parveen, P., Khan, L., Thuraisingham, B.: Ontology-driven query expansion using map/reduce framework to facilitate federated queries. In: Proc. of the International Conference on Web Services (ICWS), 712-713. (2011)
2. Euzenat, J., Shvaiko P.: Ontology Matching. In: Springer, Heidelberg, Germany, (2007)
3. Maedche, A., Staab, S.: Measuring Similarity between Ontologies. In: Knowledge Engineering and Knowledge Management, 251-263. (2002)
4. Xue, X., Wang, Y., Hao, W.: Optimizing Ontology Alignments by using NSGA-II. In: The International Arab Journal of Information Technology, 12(2), 176-182. (2015)
5. Martinez-Gil, J., Alba, E., Aldana-Montes, J.F.: Optimizing ontology alignments by using genetic algorithms. In: The workshop on nature based reasoning for the semantic Web, Karlsruhe, Germany. (2008)
6. Calvaneze, D.: Optimizing ontology-based data access. KRDB Research Centre for Knowledge and Data. In: Free University of Bozen-Bolzano, Italy. (2013)
7. Gottlob, G., Orsi, G., Pieris, A.: Ontological queries: Rewriting and optimization. In: Data Engineering, 2-13. (2011)
8. Li, Y., Heflin, J.: Query optimization for ontology-based information integration. In: Information and knowledge management, 1369-1372. (2010)
9. Keet, C.M., Ławrynowicz, A., d'Amato, C., Hilario, M.: Modeling issues & choices in the data mining optimization ontology. (2013)

10. Keet, C.M., Ławrynowicz, A., d'Amato, C., Kalousis, A., Nguyen, P., Palma, R., Stevens, R., Hilario, M.: The data mining Optimization ontology. In: Web Semantics: Science, Services and Agents on the World Wide Web, 32, 43-53. (2015)
11. Montes-y-Gómez, M., Gelbukh, A., López-López, A.: Comparison of Conceptual Graphs. In: Artificial Intelligence, 1793. (2000)
12. Biggs, N., Lloyd, E., Wilson, R.: Graph Theory. In: Oxford UP, 1736-1936. (1986)
13. Bondy, J., Murty, U.: Graph Theory. In: Springer. (2008)
14. Kravets, P.: The control agent with fuzzy logic, Perspective Technologies and Methods in MEMS Design, MEMSTECH'2010, 40-41 (2010)
15. Davydov, M., Lozynska, O.: Information System for Translation into Ukrainian Sign Language on Mobile Devices. In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 48-51 (2017).
16. Davydov, M., Lozynska, O.: Linguistic Models of Assistive Computer Technologies for Cognition and Communication. In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 171-175 (2017)
17. Martin, D., del Toro, R., Haber, R., Dorronsoro, J.: Optimal tuning of a networked linear controller using a multi-objective genetic algorithm and its application to one complex electromechanical process. In: International Journal of Innovative Computing, Information and Control, Vol. 5/10(B), 3405-3414. (2009).
18. Precup, R.-E., David, R.-C., Petriu, E.M., Preitl, S., Rădac, M.-B.: Fuzzy logic-based adaptive gravitational search algorithm for optimal tuning of fuzzy controlled servo systems. In: IET Control Theory & Applications, Vol. 7(1), 99-107. (2013).
19. Ramirez-Ortegon, M.A., Margner, V., Cuevas, E., Rojas, R.: An optimization for binarization methods by removing binary artifacts. In: Pattern Recognition Letters, 34(11), 1299-1306. (2013).
20. Solos, I.P., Tassopoulos, I.X., Beligiannis, G. N.: Optimizing Shift Scheduling for Tank Trucks Using an Effective Stochastic Variable Neighbourhood Approach. In: International Journal of Artificial Intelligence, 14(1), 1-26. (2016).
21. Martin, D., del Toro, R., Haber, R., Dorronsoro, J.: Optimal tuning of a networked linear controller using a multi- In: objective genetic algorithm and its application to one complex electromechanical process. In: International Journal of Innovative Computing, Information and Control, Vol. 5/10(B), 3405-3414. (2009).
22. Nazarkevych, M., Klyujnyk, I., Nazarkevych, H.: Investigation the Ateb-Gabor Filter in Biometric Security Systems. In: Data Stream Mining & Processing, 580-583. (2018).
23. Nazarkevych, M., Klyujnyk, I., Maslanych, I., Havrysh, B., Nazarkevych, H.: Image filtration using the Ateb-Gabor filter in the biometric security systems. In: International Conference on Perspective Technologies and Methods in MEMS Design, 276-279. (2018).
24. Nazarkevych, M., Buriachok, V., Lotoshynska, N., Dmytryk, S.: Research of Ateb-Gabor Filter in Biometric Protection Systems. In: International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 310-313. (2018).
25. Nazarkevych, M., Oliarnyk, R., Dmytruk, S.: An images filtration using the Ateb-Gabor method. In: Computer Sciences and Information Technologies (CSIT), 208-211. (2017).
26. Nazarkevych, M., Kynash, Y., Oliarnyk, R., Klyujnyk, I., Nazarkevych, H.: Application perfected wave tracing algorithm. In: Ukraine Conference on Electrical and Computer Engineering (UKRCON), 1011-1014. (2017).
27. Nazarkevych, M., Oliarnyk, R., Troyan, O., Nazarkevych, H.: Data protection based on encryption using Ateb-functions. In: International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), 30-32. (2016).

28. Babichev, S., Gozhyj, A., Kornelyuk A., Litvinenko, V.: Objective clustering inductive technology of gene expression profiles based on SOTA clustering algorithm. In: Biopolymers and Cell, 33(5), 379–392. (2017)
29. Lytvyn, V., Sharonova, N., Hamon, T., Vysotska, V., Grabar, N., Kowalska-Styczen, A.: Computational linguistics and intelligent systems. In: CEUR Workshop Proceedings, Vol-2136 (2018)
30. Vysotska, V., Fernandes, V.B., Emmerich, M.: Web content support method in electronic business systems. In: CEUR Workshop Proceedings, Vol-2136, 20-41 (2018)
31. Gozhyj, A., Vysotska, V., Yevseyeva, I., Kalinina, I., Gozhyj, V.: Web Resources Management Method Based on Intelligent Technologies, Advances in Intelligent Systems and Computing, 871, 206-221 (2019)
32. Gozhyj, A., Chyrun, L., Kowalska-Styczen, A., Lozynska, O.: Uniform Method of Operative Content Management in Web Systems. In: CEUR Workshop Proceedings (Computational linguistics and intelligent systems, 2136, 62-77. (2018).
33. Korobchinsky, M., Vysotska, V., Chyrun, L., Chyrun, L.: Peculiarities of Content Forming and Analysis in Internet Newspaper Covering Music News, In: Computer Science and Information Technologies, Proc. of the Int. Conf. CSIT, 52-57 (2017)
34. Vysotska, V., Lytvyn, V., Burov, Y., Gozhyj, A., Makara, S.: The consolidated information web-resource about pharmacy networks in city. In: CEUR Workshop Proceedings (Computational linguistics and intelligent systems), 2255, 239-255. (2018)
35. Lytvyn, V., Vysotska, V., Burov, Y., Veres, O., Rishnyak, I.: The Contextual Search Method Based on Domain Thesaurus. In: Advances in Intelligent Systems and Computing, 689, 310-319 (2018)
36. Lytvyn, V., Vysotska, V.: Designing architecture of electronic content commerce system. In: Computer Science and Information Technologies, Proc. of the X-th Int. Conf. CSIT'2015, 115-119 (2015)
37. Lytvyn, V., Vysotska, V., Uhryn, D., Hrendus, M., Naum, O.: Analysis of statistical methods for stable combinations determination of keywords identification. In: Eastern-European Journal of Enterprise Technologies, 2/2(92), 23-37 (2018)
38. Rusyn, B., Lytvyn, V., Vysotska, V., Emmerich, M., Pohreliuk, L.: The Virtual Library System Design and Development, Advances in Intelligent Systems and Computing, 871, 328-349 (2019)
39. Gozhyj, A., Kalinina, I., Vysotska, V., Gozhyj, V.: The method of web-resources management under conditions of uncertainty based on fuzzy logic. In: International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT, 343-346 (2018)
40. Su, J., Sachenko, A., Lytvyn, V., Vysotska, V., Dosyn, D.: Model of Touristic Information Resources Integration According to User Needs. In: International Scientific and Technical Conference on Computer Sciences and Information Technologies, 113-116 (2018)
41. Lytvyn, V., Peleshchak, I., Vysotska, V., Peleshchak, R.: Satellite spectral information recognition based on the synthesis of modified dynamic neural networks and holographic data processing techniques, 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT, 330-334 (2018)
42. Vysotska, V., Hasko, R., Kuchkovskiy, V.: Process analysis in electronic content commerce system. In: 2015 Xth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), 120-123. (2015).