An Unconventional Look at a Historical Monograph. Analysis with Artificial Intelligence (AI) Tools

Katarzyna Jarzyńska

University of Warsaw
k.jarzynska2@uw.edu.pl

Abstract. Goal and theses: *The article aims to check the applicability of methods based on processing large sets of information in research in social sciences.*

Conception/research methods: *The dynamic development of new research methods based on the automated processing of large data sets using artificial intelligence (AI) means that they are used in an increasingly wide range of disciplines, going beyond the field of exact and natural sciences. Text mining was combined with available CLARIN web applications and keyword extraction and analysis strategy, a combination of the YAKE! written in Python with the VOSViewer program for the visualisation of bibliometric networks.*

Results and conclusions: *The study showed how automatic keyword extraction creates opportunities in social science research. The use of CLARIN and Google Pinpoint web tools in the analysis significantly facilitates working with a large body of texts and accelerates its analysis.*

Cognitive value/originality: *The study indicates new research methods that can contribute to the development of social sciences. The perspectives for the implementation of the ways of dealing with large data sets are presented in work in research on society, and conclusions regarding the development of digital social sciences are formulated.*

Keywords: *artificial intelligence, big data, CLARIN, December crisis of 1970, Natural Language Processing*

**Neįprastas žvilgsnis į istorinę monografiją.
Analizė su dirbtinio intelekto (AI) įrankiais**

Santrauka. Straipsnio tikslas ir tezės: *straipsniu siekiama patikrinti metodų, grįstų didelių informacijos rinkinių apdorojimu, pritaikomumą socialinių mokslų tyrimuose.*

Koncepcija/tyrimo metodai: *dinamiška naujų tyrimų metodų, grįstų automatizuotu didelių duomenų rinkinių apdorojimu naudojant dirbtinį intelektą (DI), plėtra reiškia, kad jie naudojami vis platesniuose disciplinų laukuose, peržengiant tiksliųjų ir gamtos mokslų sritis. Teksto gavyba buvo derinama su turimomis CLARIN žiniatinklio programomis ir raktinių žodžių ištraukimo bei analizės strategija, YAKE! parašyta Python kalba su VOSViewer programa, skirta bibliometriniams tinklams vizualizuoti.*

Esminiai žodžiai: *dirbtinis intelektas, didieji duomenys, CLARIN, 1970 m. gruodžio mėn. krizė Lenkijoje, natūralios kalbos apdorojimas*

1. Introduction

As part of the research, an attempt was made to take an unconventional look at the historical monograph, the publication prepared as part of the series „Polska mniej znana 1944–1989."[1], edited by Marek Jabłonowski, Włodzimierz Janowski and Grzegorz Sołtysiak, entitled *Kryzys grudniowy 1970 r. w świetle dokumentów i materiałów Komisji Władysława Kruczka i relacji obozu władzy*[2]. Its text was subjected to statistical and quantitative research using significant data processing technology and artificial intelligence. This article aims to present the applicability of such methods in everyday research in areas beyond the natural and exact sciences.

---

[1]   It can be translated as "Poland less known 1944–1989".

[2]   *The December crisis of 1970 in the light of the documents and materials of the Commission of Władysław Kruczek and the reports of the government camp.*

The detailed objectives of the study undertaken include:
- checking the possibility of using automatic keyword extraction
- examining the use of CLARIN tools: CompCorp, Cat, Topic, LEM, Pelcra
- indication of the possibilities of the Google Pinpoint tool in research in the field of social sciences.

2. Methodology

The analyses were performed using a corpus of documents from the book *Kryzys grudniowy 1970 r. w świetle dokumentów i materiałów Komisji Władysława Kruczka i relacji obozu władzy*. The study also covered off-source material, i.e. editorial comments and historical explanations. At an early stage of text processing, footnotes, bibliographies, indexes and tables of contents were removed.

After collecting the book excerpts in pdf format, a particular application was converted to text format, which is more relevant for the needs of natural language processing tools. The simplification of the text was achieved thanks to lemmatisation, i.e. the process of reducing words to their basic lexical form using the LEM (Literary Machine Explorer) – a tool created by the CLARIN (Common Language Resources & Technology Infrastructure) consortium, which is a pan-European scientific infrastructure that allows researchers in the field of humanities and social sciences to work conveniently with large corpora of texts. It was also used to remove the so-called "stopwords", that is, the most common words in Polish that do not have a significant meaning for the content of the text and may distort the final result. The Tokeniser tool was used for text tokenisation, and the essence is to break it down into smaller lexical units that are easier for tools to analyse.

Among the methods for analysing texts (text mining) based on artificial intelligence (AI), it is possible to distinguish several that best meet the research needs of scientists dealing with scientific work within the disciplines included in the social sciences.

Keywords are "natural language words (expressions) used to describe the texts of the documents for their later retrieval."[3] (Babik, 2010, s. 77). A human can give them, most often the author of the work, manually or by a machine (artificial intelligence) in an automated manner. Automatic keyword extraction is done with YAKE! (Yet Another Keyword Extractor!), which is an unsupervised keyword extractor. It works on any text, extracting the essential terms from the studied corpus based on their occurrence frequency and coexistence.

VOSViewer is used to visualise automatically obtained keywords for individual documents in the form of a network of interrelationships. Thanks to bibliographic mapping and clustering, it allows generating graphs of dependencies between keywords obtained thanks to YAKE!.

The study also used several of the many web applications available under CLARIN. For this study, the CompCorp tool was used – it is used to compare the frequency of occurrence of lemmas in two corpora; Cat – enables the categorisation of texts; LEM – a multitasking application that enables the processing of literary texts; Topic – software that allows finding dominant topics in a given set of texts; WiKNN – an application that searches for expressions in the text that are classified in Wikipedia and gives them appropriate categories.

The possibility of using a free solution for document collections analysis – Google Pinpoint – available as part of the Google Journalist Studio service was also mentioned. Pinpoint converts text files of any format – docx, pdf, and photos into readable form by artificial intelligence. Based on sources uploaded in various formats, the machine designates people, places, and institutions and counts their occurrences in the created collection. It is an easy-to-use tool that can be successfully used even by people with low IT competencies.

---

[3] Originally „wyrazy (wyrażenia) języka naturalnego, użyte do opisu tekstów dokumentów w celu ich późniejszego wyszukiwania". Translated by the author of the article.

3. Results

3.1. Keywords analysis

Automatic keyword extraction was possible thanks to the YAKE! Tool, written in the Python programming language. They were selected from text collections based on their frequency and co-occurrence statistics.

The visualisation of the keywords extracted from the 1st part of the documents collected in the book was prepared in the VOSViewer program. Spheres mark individual words, and their size increases with the number of times that the word is found. The colours indicate clusters or thematically related groups of words. The lines between words define the connections among them, while their thickness determines the strength of the bonds linking the two lemmas.

Figure 1. A network of keyword associations
Source: VOSViewer.

The first red cluster focuses on the figures of Edward Babiuch, Stefan Misiaszek, Władysław Kruczek, the matters of the eighth plenum and the commission (Fig. 1). The second green cluster includes the figures of Lucjan Czubiński and Kociołek, conversations with workers in shipyards, obtaining information, and working in the field. The third blue cluster represents topics related to Gdańsk, Gdynia, decision making, the situation in the shipyard, the figures of Korczyński, Log-Sowiński and Cyrankiewicz. The last smallest yellow cluster refers to the highest authority residing in Warsaw – Gomułka and the Polish United Workers' Party (PZPR).

The connections of words within one cluster indicate the thematic closeness of the designated keywords. Those whose visualisation in the form of a sphere has the most significant size represent the most dominant topics in a given cluster. The thickness of the lines shows the number of connections between specific lemmas. Numerous connections between different sets indicate the closeness of the issues included in all groups of words.
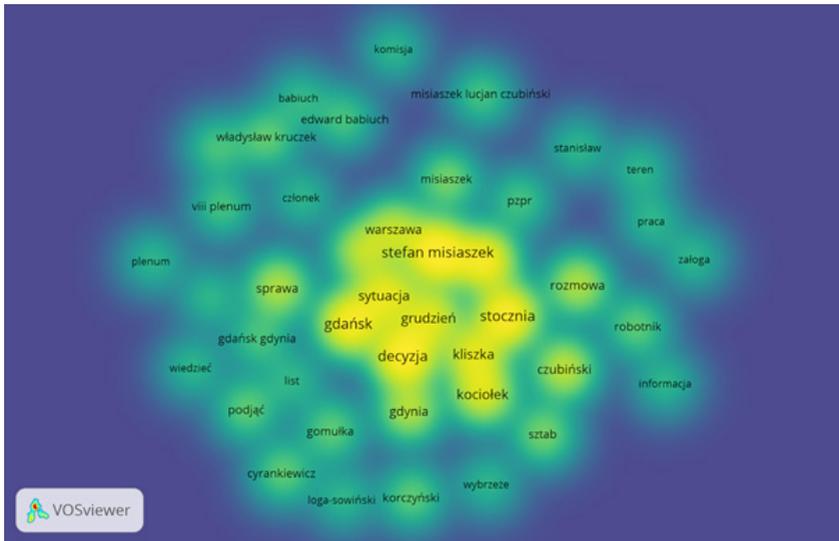
Figure 2. Density mapping
Source: VOSViewer.

Density mapping provides additional insight into the associations' intensity (Fig. 2). The more intense the sphere's colour with the word, the more often it appears and becomes associated with other words. In the analysed documents, the fundamental topic is the situation in Gdańsk and Gdynia in the shipyards. It is necessary to make some decisions, and there are also talks between the staff. The most important people participating in these events are Misiaszek, Kliszka, Kociołek and Czubiński. Gomułka is somewhat on the side of the situation, and it is indicated by the remoteness of the lemma representing him from the centre of the graph.

Figure 3. Network of connections with the lemma "stefan misiaszek"

Source: VOSViewer.

It is also possible to check with what words a given lemma represented as a specific "ball" is connected by clicking on it with the mou-

se cursor. Stefan Misiaszek is involved in the shipyards in Gdańsk and Gdynia, workers, matters of the eighth plenum, and committee deliberations (Fig. 3). On the other hand, the decision concerns a situation that is taking place in Gdynia and Gdańsk, there is a conversation about it, and some information is needed to make it (Fig. 4).
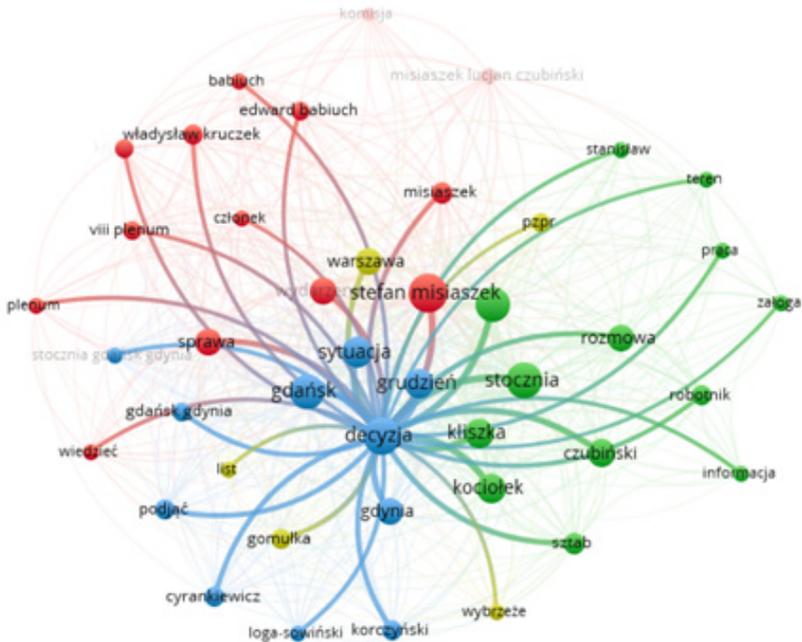
Figure 4. Network of connections with the lemma "decision"
Source: VOSViewer.

3.2. Corpora comparison

CompCorp is a tool offered by CLARIN for comparing linguistic features of corpora. It allows uploading any two text sets previously packed to the .zip archive format. Then, it compares them in terms of features such as the presence of characteristic (specific) multiword units or the company of grammatical tags (according to the NKJP tagset). It also considers the presence of particular vocabulary for given corpora,

the presence of language that differentiates corpora, the presence of proper names, verb characteristics, and statistical features of corpora. Thus, it enables quick identification of linguistic features common to and differentiation between any two sets of texts.

The possibilities of the application were examined by comparing the two corpora of documents from the book *Kryzys grudniowy 1970 r. w świetle dokumentów i materiałów Komisji Władysława Kruczka i relacji obozu władzy*. Corpus A consists of documents from the first part of the book, and corpus B – documents from the 2nd part (Fig. 5). In corp A, the words *decyzja, stocznia, sytuacja, Kliszka, Gdańsk, grudzień*[4] were the most common ones. The most common words in the B corpus were the words *godzina, stocznia, grudzień, decyzja, Gdańsk, Kliszka, praca, sytuacja*[5]. The convergence of lemmas allows for the formulation of a conclusion that the examined corpora contain documents on a similar subject.
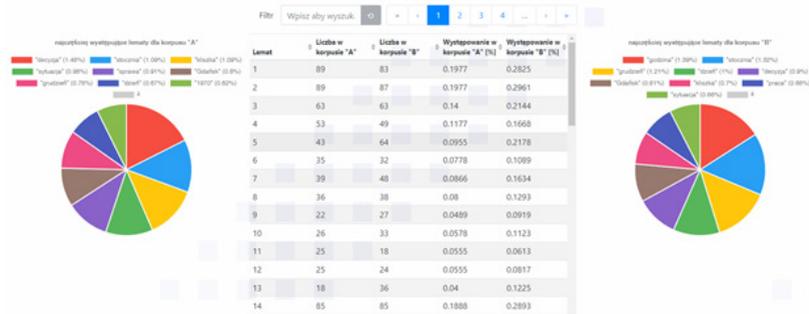
Figure 5. Comparison of the most common lemmas in the A and B corps

Source: CompCorp.

CompCorp enables the detection of proper names in the text and divides them into categories: *przymiotniki pochodzące od nazw własnych,*

---

[4]   It can be translated as decision, shipyard, situation, Kliszka, Gdańsk, December.
[5]   Which refers to an hour, shipyard, December, decision, Gdańsk, Kliszka, work, situation.

budynki, istoty żywe, produkty (wytwory ludzkie), lokalizacje, pozostałe, organizacje i ich składowe, wydarzenia[6]. The number of occurrences of individual proper names in lemmatised form is measured statistically, their type, number of occurrences in the A and B corpus are determined, and the coefficient of differentiation of the occurrences of a given lemma in the analysed sets.

In corpus A (Fig. 6), the most common were living creatures (1800 appearances), locations (600 appearances), as well as organisations and their components (over 500 appearances). In the case of Corpus B, products predominated, i.e. human creations (almost 600 appearances), events (400 appearances) and other lemmas (over 300 appearances).

On this basis, we can initially determine that the A-corpus documents mainly relate to people, organisations associated with them, and the locations in which they are located. On the other hand, the B corps concerns with some events with which human creations are connected and other lemmas not defined by automation.

Figure 6. A and B corpus proper names statistics

Source: CompCorp.

---

[6]   In English: proper names adjectives, buildings, living creatures, products (human products), locations, others, organisations and their components, events.

3.3. Text classification

Cat is a simple text classification tool. Any document corpus packed into a .zip archive can be analysed. The web application allows to group documents according to one of the following criteria:

1. thematic classification according to the model learned on the five Wikipedia clustering categories,
2. thematic classification according to the model learned on the press subject,
3. classification according to the similarity of the grammatical style to the style of one of the famous writers of the 20th century,
4. detection of language participation in the entire body.
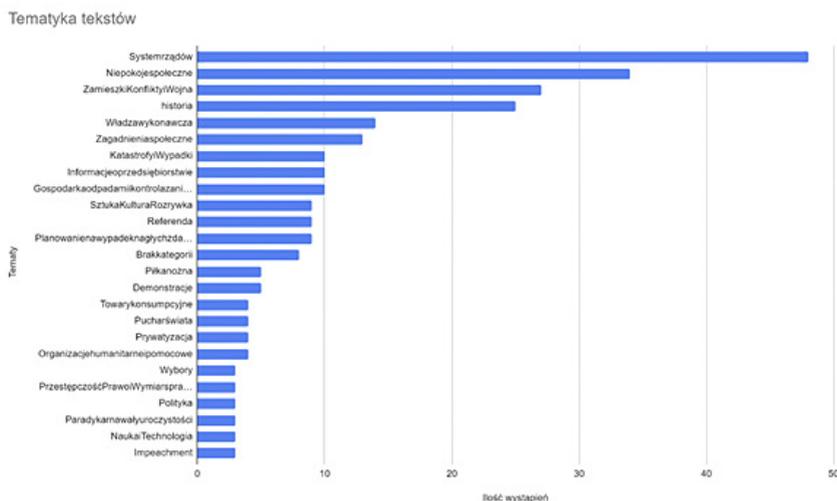
Figure 7. The most common topics in the studied documents

Source: Cat.

For each document from the book under study, using the Cat tool, several topics were selected that dominate it (Fig. 7). Governance, social

unrest, riots, conflicts and war, history and the executive power[7] are the most frequently represented issues. Based on the categories obtained for the analysed texts, VOSViewer made a graph visualising the most frequently repeated topics and their most robust connections (Fig. 8). Each colour in the graph represents a separate cluster of related topics. The larger the circle, the more frequently the topic was represented, and the lines connecting the circles indicate thematic links.

The subject of the governance system represented by the red cluster is associated with social unrest, riots, conflicts and war, and social issues. The green cluster devoted to history shows its links with executive power, impeachment, art, culture and entertainment. The blue cluster focuses on topics related to people's everyday life: football, insurance, crime, choices and consumer goods.
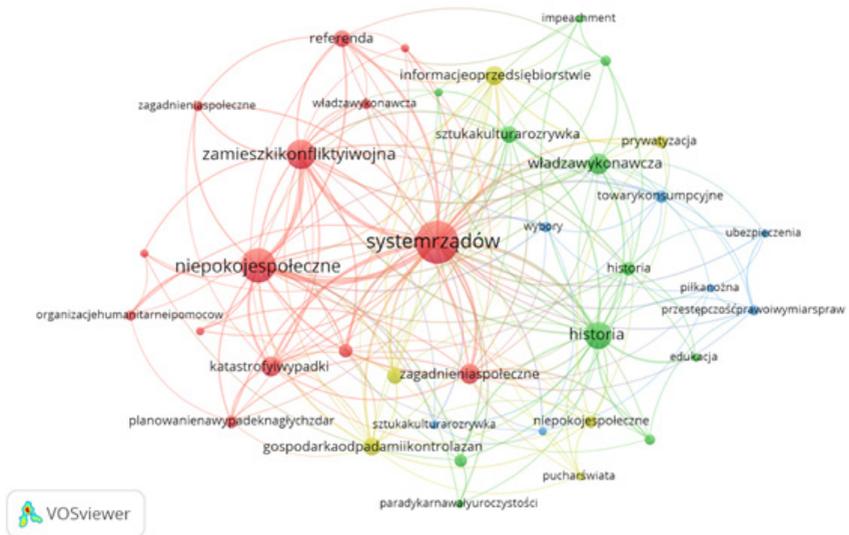
Figure 8. Visualization of the network of thematic connections

Source: Cat.

---

[7]   System rządów, niepokoje społeczne, zamieszki, konflikty i wojna, historia oraz władza wykonawcza.

3.4. Thematic modelling

Topic is a tool that enables thematic modelling of collections of texts, thanks to their meticulous analysis. It allows the designation and extraction of a user-defined number of topics, i.e. topics from the analyzed text corpus. Thematic interpretation is based on activities on a group of texts, the result of which is the determination of distinct groups of words that determine their mutual occurrence in the entire set of documents, which are often semantically coherent.

The application allows entering the corpus of the examined texts in the form of a zip archive. The analysis results are available for review in visualizations and as files with data for download in JSON and xlsx formats. It is also possible to visualize individual topics in the form of graphic files with word clouds.
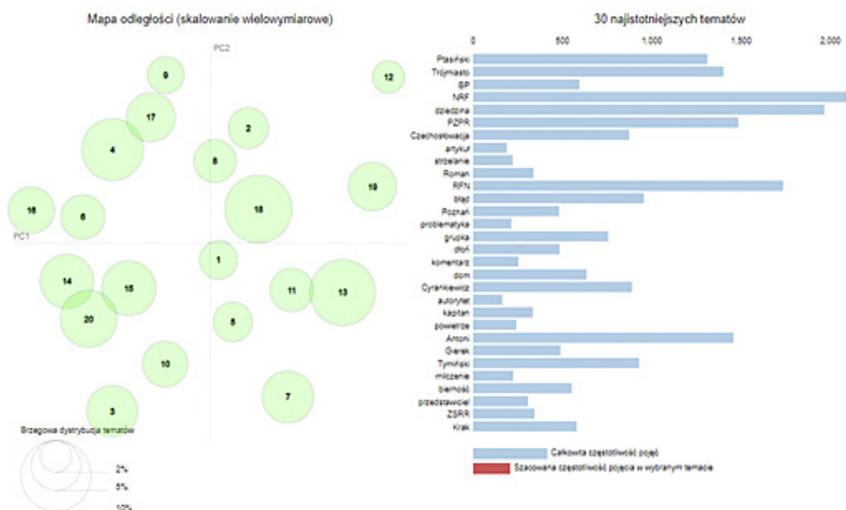
Figure 9. Thematic distance map for the documents from 1st part

Source: Topic.

For the documents from the 1st part of the book, a thematic distance map was made using the multidimensional scaling method

(Fig. 9). Each circle represents a separate topic, and the distances be-
tween them – the issues – indicate how related they are. The object's
size determines its importance, and the numbers placed on the wheels
are the numbers of successive topics chosen by the machine.

After clicking on the circle, it is possible to see what words make up
a given topic and how often they appear in it and throughout the docu-
ment (Fig. 10). When hovering the cursor over it, the tool generates a
more graphically accessible word cloud for a given topic (Fig. 11). In
the analyzed set of documents, topic 18, located in the centre of the
distance map, was the most strongly represented. Among the key terms
appearing there are lemmas: *grudzień, Gdańsk, Gdynia, Kociołek, egze-
kutywa* and *stocznia*.[8]
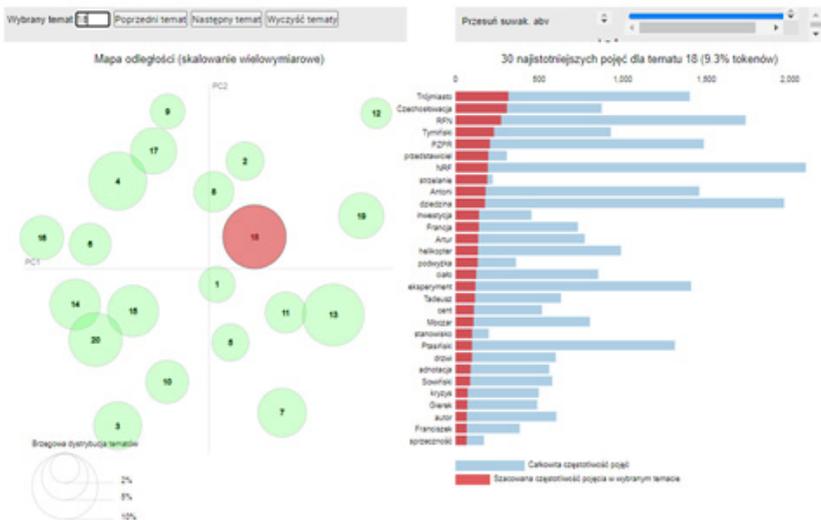
Figure 10. Lemmas that make up topic 18 in the analyzed set
Source: Topic.

Comparing the most strongly represented topics with each other
can give a complete picture of the subject matter of the studied docu-

---

[8]   December, Gdańsk, Gdynia, Kociołek, executive and shipyard.

ments. The analysis of the issues whose graphical representation in the form of circles overlapping each other will allow us to isolate the related problems and draw appropriate conclusions in this regard.

Figure 11. Word cloud for the 18th theme

Source: Topic.

3.5. Proper names statistics

LEM, or Literary Machine Explorer, processes literary texts in Polish to extract statistical information from them. It enables the processing of text data from many files saved in various formats. Subjecting them to lemmatization and determining the parts of speech they contain is the central part of LEM. It also allows for characterizing the verbs used in the text, extracting statistics from the corpus, unifying the linguistic meaning, determining hypernyms and hyponyms, performing a stylometric analysis, and creating a sorted list of proper names.

LEM, or Literary Machine Explorer, is used to process literary texts in Polish to extract them from them. The last of these applications was

used to examine the frequency of proper names in several selected documents from the analyzed book. After uploading the corpus of texts, LEM found the terms appearing in them, and it also determined their type, e.g. nam_liv – a living creature, nam_org – organization, nam_loc – place, nam_adj – adjective. They were then counted for each document separately. Thanks to this, it is possible to compare how often a specific person, place, event or organization is repeated in a given document.

Word [no] Lemma		Frequency	Type	aneks%Grudzień „70 wersja II-571-577.txt.txt	aneks%Grudzień „70 wersja II-599-615.txt.txt	aneks%Grudzień „70 wersja II-515-570.txt.txt	aneks%Grudzień „70 wersja II-595-598.txt.txt
Gdańsk	Gdańsk	110	nam_loc	19	26	64	1
Wiesław	Wiesław	79	nam_liv	1		77	1
KC	KC	49	nam_org	6	19	24	
Warszawa	Warszawa	47	nam_loc	9	15	22	1
Polska	Polska	37	nam_loc	14	3	20	
Gomułka	Gomułka	32	nam_loc	12	5		15
Gdynia	Gdynia	31	nam_loc	2	16	13	
KC PZPR	KC PZPR	30	nam_org	10	14	3	3
Gierek	Gierek	30	nam_liv	7	3	19	1
Stanisław I	Stanisław I	24	nam_liv	3	15	6	
PZPR	PZPR	24	nam_org	11		10	3
Władysław	Władysław	21	nam_liv	2	10	8	1
Kociołek	Kociołek	20	nam_loc	3	1	15	1
KW	KW	20	nam_org	1	9	10	
polski	polski	18	nam_adj	3	4	10	1
MSW	MSW	18	nam_org	2	13	3	
BP	BP	15	nam_org	1		14	
Moskwa	Moskwa	15	nam_loc	11		2	2
Edward Gi	Edward Gi	15	nam_liv	3	5	3	4
Karkoszka	Karkoszka	15	nam_loc	1		14	
Cyrankiew	Cyrankiew	13	nam_liv	1	1	10	1
Bolesław J	Bolesław J	13	nam_liv	1	4	6	2
Ignał	Ignał	12	nam_liv			12	
radziecki	radziecki	12	nam_adj	6		6	
S.I	S.I	12	nam_org				12

Figure 12. Proper names statistics

Source: LEM.

In the analyzed fragments, the most common lemma was *Gdańsk*, with 110 occurrences (Fig. 12). There are many occurrences of this word in the third analyzed fragment, which may indicate that it describes the events in *Gdańsk*. Among the lemmas for living creatures, *Wiesław* is the most common, appearing with the same frequency as *Gdańsk* in the third of the analyzed fragments. We can conclude that the described person named Wiesław has excellent connections with Gdańsk in him. Among the organizations, the dominant lemma of the *KC* (Central Committee) is also the most frequent in the third analyzed fragment. The *polski* (Polish) lemma appears most often in the group of adjectives, also the dominant one in the third fragment. It may signal the connections between the words Gdańsk, Wiesław, KC and Polish, which are highly concentrated in the third document. Therefore it is worthwhile to refer to it for further analysis first.

3.6. WiKNN tool

WiKNN (Wikipedia K–Nearest Neighbors) is a web-based text classifier that allows processing texts in Polish and English. It checks them for the presence of Wikipedia objects in them and matches them with the appropriate category based on those existing in the online encyclopedia. Unique instances of the character are designated in the count column, and the category's weight is specified in the weight column. Without reading the text, the analysis results prove that in the analyzed fragment of the book, four characters appearing in its content were members of the National National Council, three – presidents, three – deputy prime ministers of the People's Republic of Poland (Fig. 13).

Show 10 entries					Search:
Name	Count	Weight	ScaledWeight	Skip	
Posłowie_do_Krajowej_Rady_Narodowej	4	708.63	-0.45	false	
Prezydenci_Polski	3	638.71	-0.57	false	
Wicepremierzy_PRL	3	635.84	-0.58	false	
Showing 1 to 3 of 3 entries				Previous 1 Next	

Figure 13. Categories designated for one of the examined documents

Source: WiKNN Classifier.

The tool also has a mechanism for generating keywords for the uploaded texts. It allows to sort them according to the calculated weight and frequency. Results for categories and keywords can be downloaded in JSON processable format. For the examined textx, the most important words are: *członek, komisja, posiedzenie, pzpr, władysław* and *gdańsk*[9] (Fig. 14).

---

[9]   member, committee, meeting, PPL, Władysław and Gdańsk

Term	keynessValue	ScaledKeynessValue	localFreq	localDocSize	refFreq	refCollectionSize
członek	41.53	1.95	7	222	168599	272536863
komisja	37.53	1.53	5	222	54017	272536863
posiedzenie	45.37	2.34	4	222	6285	272536863
pzpr	42.63	2.06	4	222	8856	272536863
władysław	28.81	0.64	4	222	50225	272536863
gdańsk	28.63	0.62	4	222	51418	272536863
egzekutywa	45.38	2.34	3	222	707	272536863
sekretariat	37.59	1.54	3	222	2596	272536863
oświadczenie	35.62	1.34	3	222	3606	272536863
mislaszek	40.02	1.79	2	222	40	272536863
kolczyński	38.84	1.67	2	222	54	272536863
karkoszka	34.40	1.21	2	222	166	272536863
kruczek	29.07	0.67	2	222	632	272536863
zapoznać	23.74	0.12	2	222	2402	272536863
bochnacki	25.27	0.28	1	222	1	272536863
słabczyk	23.04	0.05	1	222	4	272536863

Showing 1 to 16 of 16 entries

Figure 14. Keywords generated by the WiKNN Classifier for the examined document

Source: WiKNN Classifier.

3.7. Google Pinpoint

Google Pinpoint is a free tool available as part of Journalist Studio from Google. Processes text documents in any format – docx, pdf, and also in the form of photos. It can recognize handwriting in scans, even in the case of poor quality old documents. It is also possible to upload mp3 files from which artificial intelligence automatically creates an exact transcription.

In the collection based on the uploaded documents in various formats, artificial intelligence designates people, places, and institutions and counts their occurrences (Fig. 15). After selecting passwords identifying a given person and spot, it is possible to check which documents appear together without manually checking the documents and reading

them carefully. In addition to predetermined people, places or organizations, we can enter any word in the search bar and find it in the text.

Filter collection
Entities displayed in polski

By people	Documents
Zenon Kliszko	51
Ignacy Loga-Sowiński	44
Stanisław Kociołek	39
Grzegorz Korczyński	35
Władysław Gomułka	33

Show more

By organizations	Documents
Polska Zjednoczona Partia Rob...	62
Warszawa	62
Archiwum Akt Nowych	57
Biuro polityczne	31
Stocznia Gdańsk S.A.	31
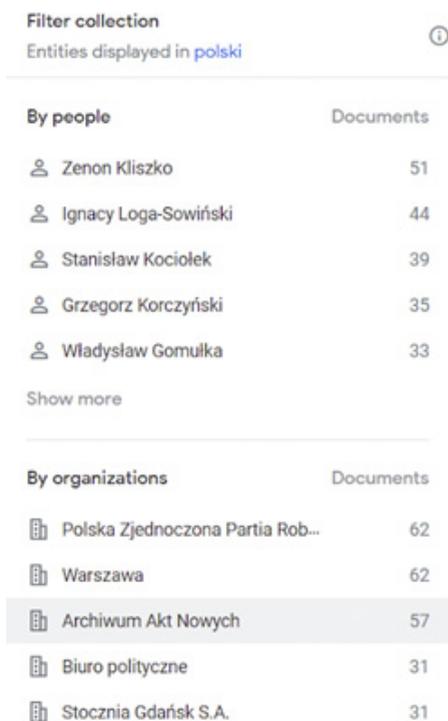
Figure 15. Appearances of people and organizations in a set of documents in Google Pinpoint

Source: Google Pinpoint.

4. Summary

The article shows the possibilities of using methods based on new technologies and artificial intelligence in research in social sciences. The applied methods of text analysis (text mining) with the use of available CLARIN web applications made it possible to demonstrate the usefulness of these tools at the initial stage of the research. They provide immediate insight into the contents of the examined corpus without

the necessity of traditional reading and browsing through hundreds of pages of documents. It makes it possible to significantly shorten the analysis of texts and select the most relevant fragments for in-depth analysis.

Keyword extraction and analysis method based mainly on the use of YAKE! written in the Python programming language can be used to establish critical topics for the documents under study. Visualization of the network of connections of the most common lemmas in VOSViewer gives an insight into the thematic connections between them. It allows us to conclude the relationships between specific people, institutions and places in the studied corpus.

The analysis of the book prepared as part of the series "Polska mniej znana 1944–1989" edited by Marek Jabłonowski, Włodzimierz Janowski and Grzegorz Sołtysiak *Kryzys grudniowy 1970 r. w świetle dokumentów i materiałów Komisji Władysława Kruczka i relacji obozu władzy* is an example how statistical-quantitative and artificial intelligence support the evaluation of large text corpora. Further studies in the use of modern research methods in social sciences are sure to discover more examples of their application and lead to the establishment of digital social sciences on a larger scale than is currently the case.

Bibliography

1. Babik, W. (2010) *Słowa kluczowe*, Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

2. Bengfort, B., Bilbro, R., Ojeda, T. (2018) *Applied text analysis with Python*, Sebastopol.

3. Campos, R., Mangaravite, V., Pasquali, A., Jatowt, A., Jorge, A., Nunes, C. and Jatowt, A. (2020) *YAKE! Keyword Extraction from Single Documents using Multiple Local Features*. In: "Information Sciences Journal". Elsevier, Vol 509, pp 257–289.

4. Goutte, C., Gaussier, E. (2005) *A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation*. „Lecture Notes in Computer Science", vol. 3408, pp. 345–359.

5. Lane, H., Howard, C., Hapke, H. M. (2019) *Natural language process-*

ing in action. Understanding, analyzing, and generating text with Python, Shelter Island.

6. Maryl, M., Piasecki, M., Walkowiak, T., (2018) *Literary Exploration Machine A Web-Based Application for Textual Scholars* In: *Linköping Electronic Conference Proceedings* 147:11, s. 128–144.

7. Pęzik, P. (2016) *WiKNN Text Classifier,* CLARIN-PL digital repository, http://hdl.handle.net/11321/277.

8. Sarkar, D. (2019) *Text Analytics with Python. A practical real-world approach to gaining actionable insights from your data*, Bangalore.

9. Walkowiak T., Datko S., Maciejewski H. (2019) *Distance metrics in Open-Set Classification of Text Documents by Local Outlier Factor and Doc2Vec.* In: Wotawa F., Friedrich G., Pill I., Koitz-Hristov R., Ali M. (eds) *Advances and Trends in Artificial Intelligence. From Theory to Practice.* Lecture Notes in Computer Science, vol 11606. Springer, Cham.

10. Walkowiak, T. (2017) *Language Processing Modelling Notation – Orchestration of NLP Microservices*. In: *Advances in Dependability Engineering of Complex Systems: Proceedings of the Twelfth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX*, Springer International Publishing, pp. 464–473.

Netography

1. Cat (N/D). Downloaded 1.02.2022 from https://clarin-pl.eu/index.php/cat/

2. CompCorp (N/D). Downloaded 1.02.2022 from https://clarin-pl.eu/index.php/comcorp/

3. LEM – Literacki Eksplorator Maszynowy (N/D). Downloaded 1.02.2022 from https://ws.clarin-pl.eu/lem.shtml

4. Tokenizer (N/D). Downloaded 1.02.2022 from https://ws.clarin-pl.eu/tokenizer.shtml

Lietuvos žiniasklaidos dešimtmečio transformacijos bruožai

Darija Ustilaitė

Vilniaus universiteto Komunikacijos fakultetas
Žurnalistikos bakalauro studijų programos komitetas
Maironio g. 7, LT-01124 Vilnius
El. paštas: u.darija2013@gmail.com

Santrauka. *Pasauliniai įvykiai, informacinių technologijų plėtra ir kiti veiksniai per pastaruosius 10 metų išaugino informacijos srautus bei pakeitė žurnalistų darbo specifiką. Tačiau žurnalistai, nepaisant permainų, išaugusių galimybių plėsti naujienų spektrą, vis dar gali aprėpti tik tam tikrą informacijos dalį. Žiniasklaidos darbuotojai dėl ribotos vietos, laiko ar žmogiškųjų išteklių atrenka dalį įvykių, kurie tampa naujienomis. Žiniasklaidoje skelbiamos atrinktos naujienos daro įtaką tam, kaip visuomenė suvokia ir interpretuoja tikrovę.*

Šiame straipsnyje, taikant turinio analizės ir giluminių interviu metodus, buvo tiriama, kokias transformacijas per 10 metų patyrė Lietuvos žiniasklaidos priemonės, buvo siekta nustatyti, kokioms naujienoms žiniasklaida skyrė ir skiria daugiausia dėmesio, kaip ir kodėl keitėsi naujienų specifika.

Esminiai žodžiai: *transformacija, radijo stotys, interneto dienraščiai, televizijos žinių tarnybos, naujienos, žurnalistai.*