Review of the NCD 45 (2024), 94–106

Ümit Atabek, Gülseren Şendur Atabek
Faculty of Communication, Akdeniz University (R), Antalya/Turkey

NEWS AS DIGITAL DATA: TEXT MINING AND ANALYSIS OF ONLINE NEWS WITH KNIME

Abstract. This paper considers news as a text corpus. The computational analysis of the news from this perspective requires data mining and analysis techniques that can handle large amounts of unstructured dataprocessing. KNIME, a free andopen-source software (FOSS) that is developed for data science applications, is very suitable for news analysis from this perspective. KNIME offers advanced text mining and analytics capabilities. Here, we develop an example workflow for online news content analysis, as well as a text-network analysis. Furthermore, we explicatein detail the workflow and the KNIME nodes used for these analyses. Our proposed workflow is reasonably versatileand flexible to be applied to other journalistic textual analyses such as similarity, sentiment, frame, discourse, and thematic analyses. This workflow also exemplifies how no/low codedata processing and computing could be effectivelyemployed in journalism and media studies.

Keywords: Computer assisted content analysis, online news content analysis, text-network analysis, KNIME, no/low code computational analysis

1. Introduction

A newspaper can be consideredas a textual corpus. The first newspapers of the $17^{th.}$ Century were almost entirely composed of textual content. The use of visual materials in newspapers was introduced later and especially increased after the invention of photographic techniques in the 19^{th}. Century[1]. The $20^{th.}$ Century had witnessed the divergence of tabloid/yellow press versus broadsheet/quality press [2, 3] where the latter is believed to be more professional,better quality and serious while the former is pejoratively labelled as lower quality for its higher usage ratio of visuals over text.

Today, text is still the most important information source in a newspaper. This is true also for online newspapers, which use multimedia materials extensively. Obviously, visuals enhance the meaning and improve the communication effectiveness of text in newspapers. Barthes [4] uses the concept of *anchorage* to explain the text-image relationship so that a particular denotation is selected out of the interpretationspossible for an image. This relationship is particularly important in newspaper captions. Without an appropriate caption, the meaning of a picture wouldn't be clear [5]. Therefore, the texts are one of the main communicative elements to inform the meaning. This fact is alsovalid for online news outlets, which typically tend to use more multimedia content as well as hyper-textuality and interactivity. Therefore, news texts are of great importance for journalism, and textual analysis is undoubtedly an appropriate methodology for journalism studies. This article revisits the issue of analysis of news as text, from a computer assisted perspective, and proposes an alternative approach with the help of a freeand open-source data analysis software. It also aims to exemplify how no and lowcode computational techniques could be employed in journalism and media studies.

In the age of digital journalism, more computational techniques for online news analysis are required [6, 7]. As demonstrated by Lansdall-Welfare, Lewis, and Cristianini [8], computational techniques are very useful for content, framing, sentiment, narration or discourse analysis of large samples of news content. Considering the facts that applications of computational approaches in social scienceare relatively limited, and computational social science is still an emerging scientific field [9], "no code" or "low code" computational techniques are needed in order to encourage the use of more computational techniques in the journalism and media studies field. The main aim of this paper is to suggest an alternative for the use of no/low code computational techniques in text analysis. Our paper proposes such no/low code techniques with an example KNIME workflow, and we accordingly hope to contribute to the recently developing computational journalism and media research. Although there is a growing literature on computational text analysis, the literature on no/low code text analysis for the field of journalism and media studies is quite limited. We hope that our study will fill this gap and familiarize the scholars with an easy-to-use tool for complex data science analyses.

2. Textual analysis

Textual analysis is a broad term that encompasses a variety of methodologies. Textual analysis includes many forms of analysis such as content, similarity, sentiment, frame, discourse, rhetorical and thematic analyses, all being centred on the text. A broader term, artefact analysis, is proposed to include the analysis of all artefacts including the texts. Thomas [10] designates and praises content analysis as a prime application of cultural analysis in artefact studies. However, primarily focusing on text, textual analysis is more elaborative and wide-ranging than other types of analyses of communication artefacts. Text as artifact can be in diverse genres, journalistic, literary, political, or legal. Newspapers may include multiple forms of journalistic texts such as news, columns, interviews, magazine etc. As Mckee suggests [11], a text is something that we make meaning from. In a text,meaning is created by using several textual elements such as words, sentences, punctuations, modalitiesetc. Therefore, the main purpose of textual analysis is to reveal the meaning in the text. In search of meaning, Fairclough [12] accepts a relational method and distinguishesbetween external and internal levels. External relations of a text are the relations with the social events, social practices and social structures, as well as the relations with other texts. Internal relations are semantic, grammatical, lexical, and phonological relations within a text.Relations within the text and relations with other texts constitute networks within the text and network of texts. All these relations contribute to the meanings in and of a text.Accordingly, these networks require the network analysis of the texts [13, 14]. Hence, the textual analysis is a complex and multilevel task.

Mckee [11] asserts that meaning can be created not only from texts but also from books, television programs, films, magazines, or even kilts or furniture. Therefore, the scope of textual analysis can be expanded. Frey, Botan and Kreps [15] widen the definition of textual analysis further to include visual messages.Therefore, textual analysis can be conceived as the analysis of all artefacts, textual, visual or material. However, textual analysis even in the form of "text-only" analysis has great potential for the analysis of communication messages. After carefully showing the misunderstandings on the goal and reach of textual analysis, Fürsich [16] (2009) clearly explains why journalism scholars would benefit from "text-only" analysis if they were aware of the limitations and strengths of this type of methodology. Particularly, when it comes to larger amounts of data, the analysis of "text" has a great analytical value.

Digital and online forms of texts can be analysed comprehensively with computer assisted textual analysis tools, generally known as computer assisted text analysis (CATA) software.

3. Content analysis

Content analysis is a particular form of textual analysis.It is an established research methodology for systematic examination of all kinds of textual materials. In his seminal work, Holsti [17] defined content analysis as “any technique for making inferences by objectively and systematically identifying specified characteristics of messages” (p. 14). His wide definition involves two imperative elements: objectivity and systematicity, which Berelson [18] previously underlined. On the other hand, Krippendorff [19] challenges Berelson’s argument on the manifest attribute of content analysis. He definesthe content analysis as the research technique for making replicable and valid inferences from texts, and expands its scope to include latent content as well. However, the analysis of the latent element in content makes that analysis somehow closer to the discourse analysis, in which the hidden intertextual meanings are searched. On the other hand, content analysis can be performed quantitatively or qualitatively, or a mix of both. The aim and the scope of the research are the main determinants of the content analysis style. Nevertheless, the overall aim of content analysis is to quantify the findings. Content analysis is also a valuable tool to trace the social change through time. The methodological worth of content analysis of newspapers for exploringthe social change was first suggested by Max Weber [20] for his 'Survey of the Press’ project at the beginning of the 20th. Century. Today, content analysis is successfully applied in many fields other than journalism and media studies, such as law, psychology, sociology, political science etc.

Simply put, content analysis is the study of recorded all human communications [21]. It is the counting and the quantification of the systematically coded words, sentences, idioms, themes, paragraphs, articles etc. in order to draw valid statistical inferences. Remembering that all texts are unstructured data by nature, cautious coding is particularly important. Computer Assisted Content Analysis (CACA) software help the analysts in coding and designing more transparent and statistically advanced research with ease.Computational approachcan be beneficial to content analysis forprocessing large volumes of data, providing uniformity, and time and cost reduction [22]. Thus, computational thinking is typically appropriate for automated content analysisin a wide range of academic disciplines, including journalism and media studies.

4. Computer Assisted Content Analysis

History of the use of computers for scholarly research in textual analysis goes back to the early 1960s. The General Inquireris the earliest content analysis softwarefor an IBM 7090 program system that was developed at Harvard in the spring of 1961 for content analysis research [23]. Early computer programs for content analysis assigned words and phrases to predefined categories to count and infer statisticallyreliable results. However, many programs came equipped with ready-made dictionaries that had been developed and used with some success by other researchers. Dictionary vocabulary can be derived either inductively from a text or deductively from more general constructs [24].

This lexical (dictionary-based) approach to content analysis is quite common in many free and open-source software [25]. For instance, Yoshikoder, a free and open-source multilingual content analysis software, developed as part of the Identity Project at Harvard, allows predefined or user developed dictionaries [26]. Yoshikoder is also able to deliver key-word-in-context (KWIC) lists in the form of concordances. Quite similar to Yoshikoder is AntConc which is a freeware toolkit for text analysis by AntLab [27]. Other free and open-sourcesoftware alternatives, such as KH Coder [28] provide co-occurrence analysis to specify the maximum distance between two words that will be judged to co-occur. KH Coder also provides clustering of authorship based on the measure of the distance of texts as dendrograms.Another free and open-source software General Architecture for Text Engineering (GATE), developed by The University of Sheffield since 1995, is a more comprehensive alternative to multilingual text analysis, providing information extraction, semantic annotation, fake news analysis, and many other natural language processing (NLP) applications [29].GATE is somehow similar to Konstanz Information Miner (KNIME) [30] in many respects. In addition to lexical content analysis tools, supervised and unsupervised machine learning tools are also available for content analysis [6]. Such machine learning analyses generally require some coding knowledge in Python. Nevertheless, both GATE and KNIME can handle supervised and unsupervised machine learning content analyses such as topic extraction/classification and clustering in an easier visual-based fashion. All these new computational developments and improvements contribute to the scholarly attention of content analysis. Among all different methodological positions, content analysis is the most commonly used method in current digital journalism studies [7].

The domain of qualitative textual analysis is dominated by proprietary software mostly. NVivo [31], MaxQDA [32] and Atlas-ti [33], the top three proprietary software that arewell-known among academics, position themselves as qualitative data analysis gears. However, these threesoftware are also well known for their textual analysis capabilities. They provide many content analysis tools for mapping word frequencies, co-occurrence lists, key-word-in-context lists, word clouds, concept maps, etc. for unstructured data such as transcriptions of focus group interviews. These three software can also process audio-visual data and exhibit certain capabilities for network analysis. SPSS [34] and SAS [35], otherwell-known proprietary data analysis software that have been essentially developed for quantitative data analysis, also include certain NLP modules for preliminary analysis of unstructured data (SAS Text Miner and SPSS Text Analytics*)*. Likewise, RapidMiner [36], T-Lab [37], Leximancer [38] and Dedoose [39] are other proprietary software with quantitative and qualitive data analysis capabilities. All these proprietary software claim to have multilingual properties; however, this is not valid for agglutinative languages such as Turkish. One of the few free and open-source alternatives for qualitative text analysis tools is RQDA, an R based software package [40]. Although its development had been formerlystopped, RQDA can handle many qualitative data analysis tasks with the use of R language.

5. Konstanz Information Miner (KNIME)

KNIME is developed by a research group in Konstanz University, Germany in 2004 (www.knime.org). Written in JAVA, KNIME is a modular free and open-source data manipulation and visualization software. It runs on Linux, MacOS and Windows operating systems. The KNIME user can create workflows, which consist of connected nodes that process data [30]. The workflows are in a visual-based programming fashion, which makes the software package easy to use. KNIME, formerly known as Hades, is a versatile data mining, cleaning, and analysis package. KNIME is somehow similar to RapidMiner, a proprietary software with a free version option (RapidMiner Basic Edition). KNIME allows users to utilize several computing languages such as C, C++, R, Python, Java, JavaScript and R through snippets. It also integrates almost all WEKA algorithms and may use many properties of Tableau.

KNIME can retrieve data from multiple online and offline sources. Text can be input in many formats including flat txt files, pdf, word format, csv, etc. KNIME offers several parser nodes for flat, doc and pdf formats, as well as aTIKA general document parser. It can also process audio and video for future extraction and image processing applications. Parallel and distributed data pipelining is also possible with KNIME Server, which also provides additional functionalities for organizational usage.

KNIME offers many nodes for machine learning (ML) algorithms, which can be used for many interesting solutions for big data analytics. Certain machine learning nodes can be used for text recognition and classification solutions. KNIME has strong text mining and textprocessing features that enableit to retrieve, mine, process and visualize unstructured textual data. It also provides many functionalities for natural language processing (NLP). KNIME has strong capabilities for sentiment analysis with specialized algorithms. Network analysis is also possible with KNIME. It can be used for many social network analysis (SNA) tasks. In order to run KNIME for text and network analyses, "Text Processing" and "Network Mining" extensions packages must be installed[41]. The fact that KNIME is a modular free and open-source software, it is very convenient for analysts. The workflows can be shared and modified easily.

A large amount of information is available from KNIME Forum. Users and companies may develop nodes for certain data analysis problems. Some companies have also contributed to specialized nodes that made KNIME a very strong data processor for bioinformatics and chemistry. Users may connect several nodes to each other and create loops so that a workflow can easily be edited for different tasks. A group of nodes may also be created and saved as meta-nodes, which can be used later in different tasks. CACA opens new challenges in online news content analysis. Although it is not possible to capture all dimensions of online content fully with one methodological tool only [42], KNIME is a good alternativeas flexibly adaptable tool for most content analysis demands. In any case, cautions must be taken for decreasing the pitfalls of any automated content analysis. In their inspiring overview paper in the context of political communication, Grimmer and Stewart [43] had argued that due to the complexity of language, automated content analysis methods would never fully replace careful and close reading of texts. Nevertheless, day after day, new cutting-edge computational tools appear to provide more sophisticated solutions to the problems of automated content analysis.

6. Online News Content Analysis with KNIME

Online news requires some additional precautions for content analysis, as opposed to traditional news. Interactivity and immediacy are the two important characteristics of online news. These make it imperative to freeze the flow of news by capturing and saving for a systematic content analysis [44]. Interactivity and immediacy are related to the liquidity of the online news, which is the result of the liquid journalism [45]. On the other hand, some web sites offer archived content that could be retrieved byweb crawling software. However, not all online news archival systems offer the full content of what was actually published on the web. Hence, the reliability of content analysis of online news is questioned due to the inconsistent and *ad hoc* nature of the online publishing environments [46]. To avoid such problematic issues, the example workflow in this paper offers a text-based content analysis from Really Simple Syndication (RSS) feeds, which is concise but robust. RSS feeds are based on Extensible Markup Language (XML) technology to publish the summaries of updated news stories periodically from a web site. RSS feeds enhance automated web publishing by combining traditional pull-oriented techniques with push-oriented protocols [47].

Figure 1 shows the example KNIME workflow that is developed for online news content analysis. As can be seen from the figure, the workflow is composed of 7 meta-nodes and 7 nodes that are connected sequentially. It has two parts: the first part is for news text content analysis and the second part is for network analysis of news text.In addition, a simple similarity measure meta-node is also included. Many of the procedures that are described here in this workflow are explained in detail in [48, 49, 50]. The workflow is fully accessible and downloadable from KNIME hub (https://hub.knime.com/), and it is also available from the corresponding author upon request.

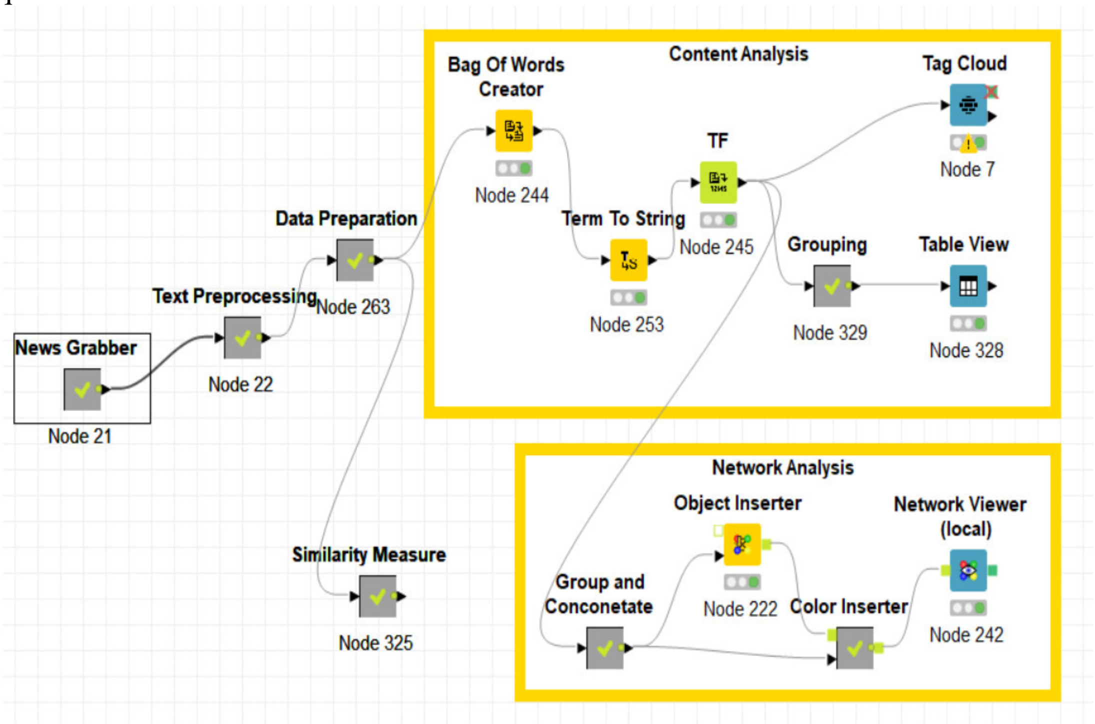

Figure 1. KNIME Workflow Example

The workflow starts with the first meta-node for news grabbing from online sources. This meta-node starts with a Table Creator Node, which forms a table of rows of URL addresses of four international online media outlets' RSS feeds.In order to make a comparative analysis our sample included RSS outlets from Russian Federation, Turkey, UK and USA. These four RSS feed sources are Russia Today (www.rt.com/rss/), Anatolian Agency (www.aa.com.tr/en/rss/default?cat=live), The Guardian (www.theguardian.com/world/rss) and The New York Times (rss.nytimes.com/services/xml/rss/nyt/HomePage.xml). The output of this node is connected to the RSS Feed Reader node, from which Document Data Extractor extracts the news text. Then the meta-node ends with a Column Filter node to clean unnecessary output. KNIME RSS feed nodes are very handy for retrieving the RSS feeds easily. The same procedure can alternatively be achieved also by using KNIME's Http Retriever node together with the Feed Parser node and the String to Document node. At the end point of this meta-node, a table of the source, the document, the title and the text of the news from four RSS feed providers are created.

The second meta-node of the workflow is for the pre-processing of textual data in order to prepare it for the analysis. Text pre-processing includes several cleaning and enriching procedures that help the textual analysis more precise and effective. There are several filters in KNIME for cleaning the text: Mark-up Tag Filter node cleans possible mark-up language tags in the news text; Number Filternode cleans unnecessary numbers and Punctuation Erasure node cleans punctuations. For further text pre-processing stemmers may be needed for stemming the words down to their roots, so that for example, the words “book” and “books” are not counted as different words. For example, KNIME has Snowball Stemmer nodethat is a very versatile with built-in stemmers for many different languages. Nevertheless, stemming can also be achieved with dictionary replacer nodes, and this solution is especially appropriate for agglutinative languages. Domination of English language in software for comparative research of languages other than English is an obstacle. In this respect, KNIME offers robust solutions to multilingual research problems, including agglutinative languages.

Our meta-node starts with the Stop Word Filter node. Stop word filtering is for cleaning unnecessary words whichhave less information value for content analysis, such as, ‘the’, ‘of’, ‘are’ etc. [51]. This node has built-in stop word lists for many languages, as well as a connection for an additional user definable list ofother unwanted words. In our example workflow, we did not use the built-in English stop word list, and used only the additional list to filter out the words that were wrongly tagged in the tagging process. Tagging is needed for text enrichment processes that can be achieved in two ways: entity recognition such as the identification of names or other domain-specific items, and tagging parts of speech (POS) to identify nouns, verbs, and so on [41, 52]. KNIME has StanfordNLP NE Tagger node thatcan both tokenize and tag the words. Our workflow used NE “person” tag option in order to enrich the text bytagging the names of the political actors in the news. The last node in the text pre-processing meta-node is a Tag Filter node which is configured to retain only the chosen name entity “person” for the analysis so that the content analysis will be limited to the political actors (persons) in the news texts. Tag Filter is a very versatile node with many built-in alternatives for tag filtering. Tagging can also be achieved by using dictionary-based tagger nodes that are provided in the KNIME node repository.

The next meta-node is configured for data preparations in order to improve the final analyses. A Dictionary Replacer node is used in order to replace similar terms with proper ones. For instance, the terms “Joseph Biden”, “Joe Biden” and “Biden” are grouped under “Biden”. These name entities and their proper counterparts are included in a Table Creator node which is connected to the Dictionary Replacer node.

The content analysis part of the workflow starts with the Bag of Words node, which extracts terms from documents. In order to remove the tag signs from the terms, a Term to String node is applied. TF node then calculates the relative or absolute term frequencies.In the example workflow, absolute frequencies are calculated with TF node. It is also possible to calculate inverse category frequency (ICF) and inverse document frequency (IDF) when needed. Finally, the term frequency list is fed to Tag Cloud node for visual representation as a word cloud, which is shown in Figure 2, for the analysis of 22$^{nd.}$ November, 2024. Other than this node, it is also possible to utilize more sophisticated R Language graphic potentials as well as Tableau visualisation solutions within KNIME.A second track in our workflow includes the Grouping meta-node which is composed of Group by, Sorter and Row Filter nodes in order to provide a table showing the frequency list for each news feed.

Figure 2. Word Cloud Output of Tag Cloud Node

The second part of the workflow is a network analysis of the news texts. Every text can be analysed as a network;the relations between the words, sentences, themes etc. within texts can be conceived as a network. Similarly, the relations between the texts in a corpus can also be analysed as networks. KNIME offers a set of appropriate tools for text-network analyses. Here, our workflow handles each news RSS feed as a separate document, in which words together with RSS sources create the nodes and edges of a network. After a document preparation, meta-node data is fed into the Bag of Word and the Term to String nodes followed by TF and Frequency Filter nodes. The output is then fed into an Object Inserter node in order to build a network of words and RSS sources. A Group and Concatenate meta-node is formed to limit the number of persons to be shown for each RSS feed to 10 persons maximum. A Network Viewer node finally visualizes the output of this network.

The output of the Network Viewer node for the analysis of 22$^{nd.}$ November, 2024is shown in Figure 3. Other than this node, it is also possible to utilize R Language network graphic potentials as well as Tableau visualisation solutions. By using Vis Output Connector node, it is also possible to direct the output of workflow to Visone, a free and open-source software for analysis and visualization of social networks [53]. Furthermore, the output from Visone can be exported to the network file formats that can be imported by Gephi, a more sophisticated free and open-source software for analysis and visualization of networks [54]. As can be seen from the figure, the name entity(person) “Netanyahu” is common in all four RSS news sources. On the other hand, “Putin” is common in NY Times, RT and AA. “Biden” and “Trump” is common in NY Times and RT. Finally, “Guatam Adani” is common in NY Times and Guardian, and “Yoav Gallant” is common in AA and Guardian. From this network graph, we may also observe the level of similarities among of the online news sources. In order to quantify this visually observable similarity, we employed a simple similarity check with Similarity Measure meta-node which is composed of Document Vector and Similarity Search nodes, and six other nodes. The results show that Anatolian Agency had a higher cosine similarity value for Russia Today (0.306), while it had relatively lower values for Guardian (0.288) and for NY Times (0.227).

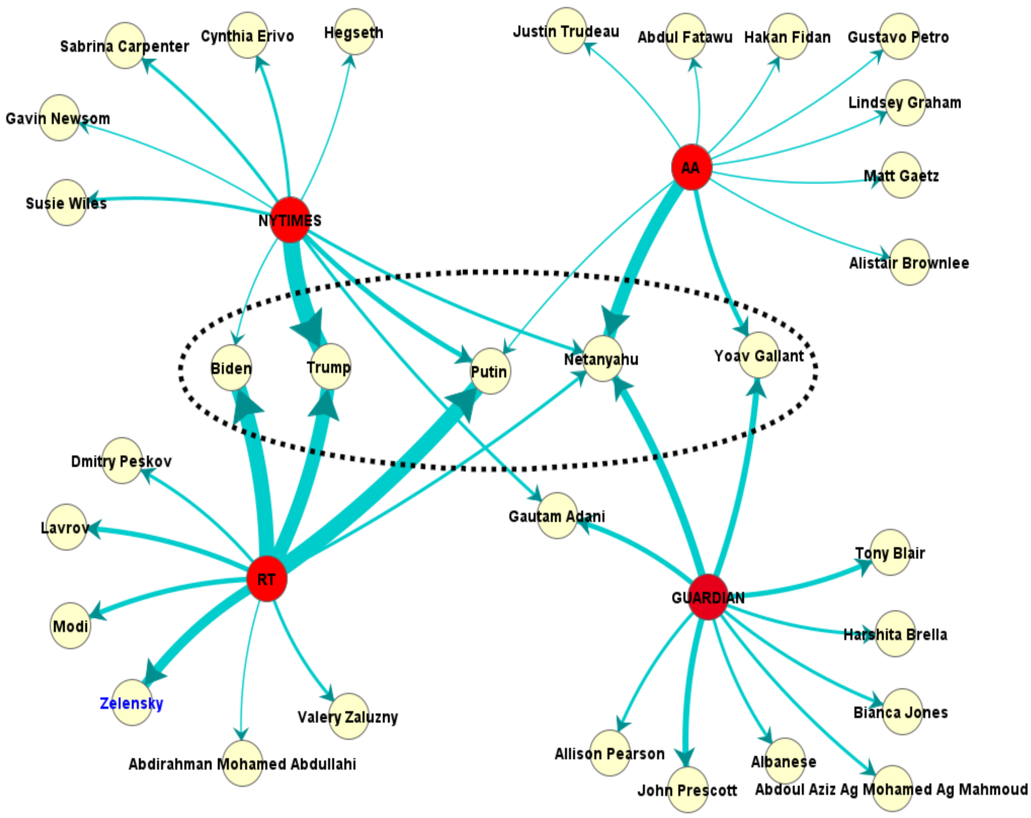

Figure 3. KNIME Network Viewer Node Output

7. Discussions and Conclusion

This paper explained a KNIME workflow that is developed for the analysis of online news content. This workflow exemplified how no/low code computing can be utilized in journalism and media analyses. Thus, we believe that our paper fulfils the aim of contributing to the no/low code computational text analysis literature. The workflow has two parts: the first part is for online news text content analysis and the second part is for network analysis of online news text. It demonstrates the content and network analyses of RSS news text from four international sources. KNIME is a versatile software for all data science applications and has strong textual analysis tools. Its text mining and analytics capabilities can easily be configured to the demands of scholars with average computing skills. Several other alternative analyses are also possible with the modified versions of this workflow. KNIME workflows can be grouped and connected for more complex systems. Even a real-time news agenda alert system can be built for practical purposes, similarly based on the systems proposed by Lamba, Yadav and Lele [55], and Tirea and Negru [56]. All these automated systems analyses require only no/low code computing techniques with KNIME's visual-based interface.

Our proposed workflow can easily be modified in order to be applied to sentiment, frame, discourse, and thematic analyses as well. For example, as proposed by Sebestyén, Domokos & Abonyi [57], it is possible to scrutinize the semantic and other qualitative differences between text bodies, by using KNIME's N-gram Extractor and Term Co-occurrence Counter nodes. Furthermore, similarity check presented in our workflow can be improved to detect churnalism. Davies [58] defined churnalism as the excessive recycling of pre-packaged public relations and press agency copy as news. As suggested by Van Houtand Van Leuven [59], churnalism can be studied in static frozen and flowing real-time news, and KNIME can be configured to both types. Other than the dictionary-based analyses, KNIME also offers unsupervised (UML) and supervised (SML) machine learning procedures to implement advanced content analysis with innovativetopic detection and classification capabilities for large amounts of data. As Van Dijk [60] suggests the derivation of topics is closely related to the discourse analysis of news as textual structures. Therefore, it is possible to utilize the topic detection capabilities of KNIME within the context of discourse analysis. KNIME includesa Topic Detector node with Parallel Latent Dirichlet Allocation (LDA) algorithm capability in order to implement UML topic detections, as exemplified by Jia [61]. On the other hand, if proper sampling methods are used, SML also contributes to the epistemological advancements of the field [62]. Robust supervised machine learning models can be built by using KNIME's SML capabilities with various learner nodes, as exemplified by Sangkaew& Zhu [63].

Because of its cross-sectional character, the workflow proposed here is obviously limited to a short period of time. Alternatively, a longitudinal design is also conceivable for a time-series analysis in order to examine the changes of the content through time. Recursive data mining is possible by using KNIME's loop nodes to repeat the whole online text retrieval within certain time windows. Additionally, for longer periods of online news, this workflow's news grabber meta-node can be modified to extract news texts from other sources such as online web portal pages and even from the print newspapers. Text retrieval from PDF versions of the print newspapers is also possible by using PDF Parser and TIKA nodes provided in KNIME's node repository. The possibilities are almost unlimited. Sjøvaag and Karlsson [64] point out that digital journalism research methods have two main dimensions: the consequences of digitalisation and the digital objects themselves. KNIME is capable of fulfilling the necessary operations for both of these dimensions that open wide epistemological opportunities. As a free and open-source modular software, KNIME may provide vast methodological potentials for journalism and media scholars. In today's complex digital media landscape, collaboration between journalism and information technology scholars are encouraged to achieve multi-perspective approaches [65]. In the recently emerging computational social sciencegrounds, journalism and media scholars are also expected to have certain computational skills to be able to use advanced analysis techniques. Software like KNIME, with visual-based interfaces and no/low code capabilities, are expected to ease such fussy requirements, and contribute to the development of journalism and media scholars' computationalskills.

References

[1] Gervais, T. & Morel, G. (2017). *The Making of Visual News: A History of Photography in the Press*. London: Bloomsbury Academic.
[2] Chapman, J. L. & Nuttall, N. (2011). *Journalism Today: A Themed History*. Oxford: Wiley-Blackwell.
[3] Popović, V., & Popović, P. (2014). The Twenty-First Century, the Reign of Tabloid Journalism. *Procedia - Social and Behavioral Sciences*, 163, 12-18. https://doi.org/10.1016/j.sbspro.2014.12.280
[4] Barthes, R. (1977). *Image Music Text*. London: Fontana Press.
[5] Bateman, J. (2014). *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. Oxon: Routledge
[6] Boumans, J. W. & Trilling D. (2016). Taking Stock of the Toolkit. *Digital Journalism*, 4(1), 8-23, https://doi.org/10.1080/21670811.2015.1096598
[7] Steensen, S. & Westlund, O. (2021). *What is Digital Journalism Studies?* Oxon: Routledge.

[8] Lansdall-Welfare, T., Lewis, J. and Cristianini, N. (2017). Automating massive-scale analysis of news content. In Franklin, Bob & Eldridge II, Scott (Eds.). *The Routledge Companion to Digital Journalism Studies*. London: Routledge. pp. 96-107.
[9] Engel, U., Quan-Haase, A., Liu, S. X. & Lyberg, L., (2021). Introduction to the Handbook of Computational Social Science. In Uwe Engel, et al. (Eds.). *Handbook of Computational Social Science*. Vol. 1. London: Routledge, pp: 1-15.
[10] Thomas, S. (1994). Artifactual Study in the Analysis of Culture: A Defense of Content Analysis in a Post-Modern Age. *Communication Research*. 21(6), 683-697. https://doi.org/10.1177/009365094021006002
[11] Mckee, A. (2003). *Textual Analysis: A Beginner's Guide*. London: Sage.
[12] Fairclough N. (2003). *Analysing Discourse: Textual Analysis for Social Research*. London: Taylor & Francis Routledge.
[13] Carley, K. M. (1997). Network Text Analysis: the network position of concepts. In Carl W. Roberts (Ed.), *Text Analysis for The Social Sciences*. Mahwah: Lawrence Erlbaum Associates. pp. 79-102.
[14] Popping, R. (2000). *Computer-assisted Text Analysis*. London: Sage Publications.
[15] Frey, L., Botan, C., & Kreps, G. (1999). I*nvestigating communication: An introduction to research methods*.(2nd ed.). Boston: Allyn & Bacon.
[16] Fürsich, E. (2009). In defense of textual analysis: Restoring a challenged method for journalism and media studies. *Journalism Studies*. 10(2), 238-252. https://doi.org/10.1080/14616700802374050
[17] Holsti, O. R. (1969). *Content Analysis for the Social Sciences and Humanities*. Reading, MA: Addison-Wesley.
[18] Berelson, B. (1952). *Content Analysis in Communication Research*. Glencoe, Ill: Free Press.
[19] Krippendorf, K. (1980). *Content Analysis: An Introduction to Its Methodology*. 2nd. Edition, Beverly Hills, CA: Sage.
[20] Weber, M. (1998). Preliminary report on a proposed survey for a sociology of the press. *History of the Human Sciences*. 11(2), 111-120. https://doi.org/10.1177/095269519801100207
[21] Babbie, E. (2004). *The Practice of Social Research* (15th ed.). Boston, MA: Cengage Learning.
[22] Cheng, A.-S., Fleischmann, K.R., Wang, P. & Oard, D.W. (2008), Advancing social science research by applying computational linguistics. *Proceedings of the American Society for Information Science and Technology*. 45, 1-12. https://doi.org/10.1002/meet.2008.1450450355
[23] Stone, P. J. & Hunt, E. B. (1963). A computer approach to content analysis: studies using the General Inquirer system. In *Proceedings of the May 21-23, 1963, AFIPS '63 Spring*. ACM, New York, NY, USA, 241-256.
[24] Mergenthaler, E. (1996). Computer-Assisted Content Analysis. In C. Züll, J. Harkness, & J. H. P. Hoffmeyer-Zlotnik (Eds.), *Text analysis and computers*. Mannheim: ZUMA. pp. 3-32.
[25] Evans, W. (1996). Computer-Supported Content Analysis: Trends, Tools, and Techniques. *Social Science Computer Review*. 14(3), 269-279. https://doi.org/10.1177/089443939601400302
[26] Lowe, W. (2015). Yoshikoder: Cross-platform multilingual content analysis. Java software version 0.6.5, http://conjugateprior.org/software/yoshikoder/
[27] Anthony, L. (2022). AntConc (Version 4.0.5) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software
[28] Higuchi, K.(2016). A Two-Step ApproachtoQuantitative Content Analysis: KH CoderTutorialusing Anne of GreenGables (Part I). *Ritsumeikansocialsciencesreview*. 52(3), 77-91.
[29] Cunningham, H., Maynard, D., Bontcheva, K. &Tablan, V. (2002). GATE: an Architecture for Development of Robust HLT Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*. Vol. 1, pp. 54–62. https://doi.org/10.3115/1073083.1073112
[30] Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T. &Wiswedel, B. (2009). KNIME - the Konstanz information miner. *ACM SIGKDD Explorations Newsletter*. 11(1), 26-31.

[31] Jackson, K. andBazeley, P. (2019) *Qualitative Data Analysis withNVivo*. (3rd Edition). Sage Publications.
[32] VERBI Software (2021). *MAXQDA 2022*. Berlin, Germany: VERBI Software.
[33] Scientific Software Development (2023). *ATLAS.ti*. Scientific Software Development GmbH.
[34] IBM (2023). *IBM SPSS Statisticsfor Windows*.Armonk, NY: IBM Corp.
[35] SAS (2023). *SAS Help and Documentation*. Cary, NC: SAS Institute Inc.
[36] RapidMiner (2024). *Altair® RapidMiner*. Troy, MI: Altair EngineeringInc.
[37] T-LAB (2024). *T-LAB*. Roccasecca (FR): T-LAB di Lancia Franco.
[38] Leximacer (2024). *Leximancer*. https://www.leximancer.com/
[39] Dedoose (2021). *DedooseVersion 9.0.17*.Los Angeles, CA: SocioCulturalResearchConsultants, LLC www.dedoose.com
[40] Huang, R. (2016). RQDA: R-based qualitative data analysis (R Package Version 0.2–8) [computer software].http://rqda.r-forge.r-project.org/
[41] Thiel, K., Kötter, T., Silipo, R. & Winters, P. (2016). Social Media Analysis: Text Mining Meets Network Mining. In Hofmann & Chisholm (Eds.) *Text Mining and Visualization: Case Studies Using Open-Source Tools*. Boca Raton, FL: CRC Press. pp. 81-92.
[42] Karlsson, M. &Sjøvaag, H. (2015). Content Analysis and Online News: Epistemologies of analysing the ephemeral Web. *Digital Journalism*. 4(1), 177-192, https://doi.org/10.1080/21670811.2015.1096619
[43] Grimmer, J. & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*. 21(3), 267–297. http://dx.doi.org/10.1016/j.pubrev.2016.09.001
[44] Karlsson, M., &Strömbäck, J. (2010). Freezing the flow of online news: Exploring approaches to the study of the liquidity of online news. *Journalism Studies*. 11(1), 2–19. https://doi.org/10.1080/14616700903119784
[45] Deuze, Mark. 2008. The Changing Context of News Work: Liquid journalism and monitorial citizenship. *International Journal of Communication*. 2, 848–65.
[46] Blatchford, A. (2020) Searching for online news content: the challenges and decisions.*Communication Research and Practice*. 6(2), 143-156, https://doi.org/10.1080/22041451.2019.1676864
[47] Hmedeh, H., Travers, N., Vouzoukidou, N., Christophides, V., Du Mouza, C., et al. (2011). Everything you would like to know about RSS feeds and you are afraid to ask. *BDA'11, Base de Donn´eesAvanc´ees*, Oct 2011, Morocco. pp.1-20.
[48] Bakos, G. (2013). *KNIME Essentials: Perform accurate data analysis using the power of KNIME*. Birmingham: Pact Publishing.
[49] Thiel, K. (2016). Introduction to the KNIME Text Processing Extension. In Hofmann & Chisholm (Eds.) *Text Mining and Visualization: Case Studies Using Open-Source Tools*. Boca Raton, FL: CRC Press. pp. 61-80.
[50] Tursi, V. & Silipo, R. (2018). *From words to wisdom: An introduction to text mining with KNIME*. Zurich: KNIME Press.
[51] Silva, C. and Ribeiro, B. (2003). The importance of stop word removal on recall values in text categorization. In *IEEE Proceedings of the International Joint Conference on Neural Networks*. Volume 3. pp. 1661–1666.
[52] Maseda, B. (2021). Text as Data: Finding Stories in Text Collections. In Bounegru& Gray (Eds.), *The DataJournalism Handbook Towards a Critical Data Practice*. Amsterdam University Press. pp. 116-123.
[53] Baur, M., Benkert, M., Brandes, U., Cornelsen, S., Gaertler, M., Köpf, B., Lerner, J. & Wagner, D. (2002). Visone Software for Visual Social Network Analysis. In: Mutzel, P., Jünger, M., Leipert, S. (Eds) *Graph Drawing. GD 2001. Lecture Notes in Computer Science*, Vol. 2265. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-45848-4_47
[54] Bastian M., Heymann S. &Jacomy M. (2009). Gephi: an open-source software for exploring and manipulating networks. In *International AAAI Conference on weblogs and social media*. Association for the Advancement of Artificial Intelligence, Vol. 3, No. 1. pp. 361-362.

[55] Lamba, A., Yadav, D. & Lele, A. (2016). CitizenPulse: A text analytics framework for proactive e-governance-a case study of Mygov.in. In *Proceedings of the 3rd IKDD Conference on Data Science*. https://doi.org/10.1145/2888451.2888463
[56] Tirea, M. & Negru, V. (2015). Text Mining News System - Quantifying Certain Phenomena Effect on the Stock Market Behavior. *17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. 391-398. https://doi.org/10.1109/SYNASC.2015.65
[57] Sebestyén, V., Domokos, E. & Abonyi, J. (2020). Focal points for sustainable development strategies—Text mining-based comparative analysis of voluntary national reviews. *Journal of Environmental Management*. 263. https://doi.org/10.1016/j.jenvman.2020.110414
[58] Davies, N. (2008). *Flat Earth News: An Award-Winning Reporter Exposes Falsehood, Distortion, and Propaganda in the Global Media*. London: Vintage Books.
[59] Van Hout, T. & Van Leuven, S. (2017). Investigating 'churnalism' in real-time news. In Franklin, Bob & Eldridge II, Scott (Eds.). *The Routledge Companion to Digital Journalism Studies*. London: Routledge. pp. 117-125.
[60] Van Dijk, T. A. (1988). *News as discourse*. New Jersey: Lawrence Erlbaum.
[61] Jia, S.S. (2018). Leisure Motivation and Satisfaction: A Text Mining of Yoga Centres, Yoga Consumers, and Their Interactions. *Sustainability*. 10. https://doi.org/10.3390/su10124458
[62] De Grove, F., Boghe, K. & De Marez, L. (2020) (What) Can Journalism Studies Learn from Supervised Machine Learning?,*Journalism Studies*. 21(7), 912-927, https://doi.org/10.1080/1461670X.2020.1743737
[63] Sangkaew,N. & Zhu, H. (2022) Understanding Tourists' Experiences at Local Markets in Phuket: An Analysis of TripAdvisor Reviews. *Journal of Quality Assurance in Hospitality & Tourism*. 23(1), 89-114, https://doi.org/10.1080/1528008X.2020.1848747
[64] Sjøvaag, H. & Karlsson, M. (2017). Rethinking Research Methods for Digital Journalism Studies. In Franklin, Bob & Eldridge II, Scott (Eds.). *The Routledge Companion to Digital Journalism Studies*. London: Routledge. pp. 87-95.
[65] Lecheler, S., de Haan, Y. &Kruikemeier, S. (2020). Introduction to the Special Issue: Meeting the Digital Demand through a Multi-Perspective Methodological Approach. *Journalism Studies*. 21(7), 857-862, https://doi.org/10.1080/1461670X.2020.1752290

uatabek@akdemiz.edu.tr