



Automation in Systematic, Scoping and Rapid Reviews by an NLP Toolkit: A Case Study in Enhanced Living Environments

Eftim Zdravevski¹(✉), Petre Lameski¹, Vladimir Trajkovik¹, Ivan Chorbev¹,
Rossitza Goleva², Nuno Pombo³, and Nuno M. Garcia³

¹ Faculty of Computer Science and Engineering,
University Sts. Cyril and Methodius, Skopje, Macedonia
eftim.zdravevski@finki.ukim.mk

² New Bulgarian University, Sofia, Bulgaria

³ Instituto de Telecomunicaes, Universidade da Beira Interior, Covilh, Portugal

Abstract. With the increasing number of scientific publications, the analysis of the trends and the state-of-the-art in a certain scientific field is becoming very time-consuming and tedious task. In response to urgent needs of information, for which the existing systematic review model does not well, several other review types have emerged, namely the rapid review and scoping reviews. In this paper, we propose an NLP powered tool that automates most of the review process by automatic analysis of articles indexed in the IEEE Xplore, PubMed, and Springer digital libraries. We demonstrate the applicability of the toolkit by analyzing articles related to Enhanced Living Environments and Ambient Assisted Living, in accordance with the PRISMA surveying methodology. The relevant articles were processed by the NLP toolkit to identify articles that contain up to 20 properties clustered into 4 logical groups. The analysis showed increasing attention from the scientific communities towards Enhanced and Assisted living environments over the last 10 years and showed several trends in the specific research topics that fall into this scope. The case study demonstrates that the NLP toolkit can ease and speed up the review process and show valuable insights from the surveyed articles even without manually reading of most of the articles. Moreover, it pinpoints the most relevant articles which contain more properties and therefore, significantly reduces the manual work, while also generating informative tables, charts and graphs.

Keywords: Enhanced living environments · Ambient assisted living
NLP toolkit · Automated surveys · Scoping review · Rapid review
Systematic review

1 Introduction

Enhanced and Assisted living environments (ELE/ALE) have been in focus of the researches for more than decade [8]. Adaptation of novel technologies in healthcare has taken a slow but steady pace, from the first wearable sensors for chronic disease conditions and activity detection with offline processing towards implantable or non-invasive sensors supported by advanced data analytics for pervasive and preventive monitoring. The ELE/ALE progress is driven by the rapid advances in key technologies in several complementary scientific areas over the last decade: sensor design and material science; wireless communications and data processing; as well as machine learning, cloud, edge, and fog technologies [18, 19, 21].

The integration of novel sensors into consumer electronics increases gathering of personal health data. The place and importance of different sensors for healthcare, well-being, and fitness among consumer devices can be tracked by their increasing share on Consumer Electronics Shows promoting self-care and self-regulation. This creates enormous possibility in both healthcare and healthy lifestyle. The availability of data in vast amounts can lead to: cost-effective, personalized, and real-time monitoring, detection and recommendations, both for the end users and healthcare providers [21]. These services (monitoring, detection, recommendation) are significant research topic in ALE/ELE domain. Thus, a large percent of typical ALE/ELE systems aim to monitor daily activities, detect specific events (e.g. falls, or false alarms), automate assistance, and decrease caregiver burden [22]. Continuous vital signs monitoring is an important application area and various sensors have been developed for this purpose. Sensor devices are supported by various algorithms and computational techniques, context modeling, location identification, and anomaly detection [19].

Human activity recognition stands for recognizing human activity patterns from various types of low-level sensor data usually presented as time series data. The activity itself can be represented and recognized at different resolutions, such as a single movement, action, activity, group activity, and crowd activity. Recognizing such activities can be useful in many applications, for example: detecting physical activity level [25], promoting health and fitness [28], and monitoring hazardous events such as falling [2, 20].

The current trends in ALE/ELE systems research can be perceived from different perspectives [5]. In this work, we are investigating research topics in the ALE/ELE systems and services domain applied to healthcare and well-being. We identified potentially relevant articles with the following keywords: identification and sensing technologies, activity recognition, risks and accidents detection, tele-monitoring, diet and exercise monitoring, drugs monitoring, vital signs supervision, identification of daily activities, and user concerns like privacy and security.

Systematic reviews, use formal explicit methods, of what exactly was the question to be answered, how evidence was searched for and assessed, and how it was synthesized in order to reach the conclusion. The “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement” [13, 14] is one

of the most widely used methodologies for achieving this. Recently, new forms of reviews have emerged in response to urgent needs for information, for which the existing systematic review model does not fit well [15]. The rapid review is used when time is of the essence. The scoping review is applied when what is needed is not detailed answers to specific questions but rather an overview of a broad field [17]. The evidence map is similar to scoping reviews but is focused on specific visual presentation of the evidence across a broad field. Finally, the realist review is used where the question of interest includes how and why complex social interventions work in certain situations, rather than assume they either do or do not work at all.

Performing any of these reviews types is usually manual and very labor-intensive work. Therefore, we have identified the opportunity to use Natural Language Processing (NLP) and other software engineering methods to automate the analysis, identify relevant articles, generate visualizations of trends and relationships, etc. We have implemented an NLP-based toolkit that performs this, and, in this paper, we show our findings in the AAL/ELE domain.

By exploring the publications over the last decade, we have summarized the state-of-the-art technologies, future research focus and publication statistics related to the following key issues: enabling technology, typical applications and services of ALE/ELE in healthcare and well-being.

The remainder of this article is organized as follows. Section 2 will elaborate the different Natural Language Processing techniques (NLP) we are using, while also describing the processing the collected data. Section 3 presents the results of our analysis in the AAL/ELE use case and discusses them. Finally, in the last section we conclude the paper and point directions for future research.

2 Methodology

This work is an extension of our previous work presented in [3]. Namely, the architecture was reworked for better reusability of intermediate results per the architecture presented in [26], while ensuring compliance with the terms of use of the digital libraries, in regard to the number of requests per unit time. Additionally, the plotting of aggregate results was integrated and streamlined using the Matplotlib library [7] and Networkx [6].

2.1 Search Input Taxonomy

The user input is a collection of keywords that are used to identify potentially relevant articles and a set of properties, which define what are we looking for in the identified articles. In particular, this input is defined with the following parameters, which are further enhanced by proposing synonyms to the search keywords and properties by the NLP toolkit, as described in the following Subsect. 2.4:

Keywords. Search terms or phrases that are used to query a digital library (e.g. ambient assisted living, enhanced living environments, etc.). See example of

searched keywords in Figs. 6 and 7. Note that keywords are being searched for independently of each other and duplicates are being removed in a later phase.

Properties. The properties are words or phrases that are being searched in the title, abstract or keywords section of the identified articles. Exemplary properties used in this study can be seen in Figs. 8, 9, 10 and 11.

Property synonyms. In addition to the original form of the properties, also their synonyms or words with similar meaning in the domain terminology, are being searched for in the article’s abstract, title and keywords. For each property, only one original form appears in the results for brevity, while the synonyms are omitted. Note that a synonym can be a completely different word, or another form of the same word, such as a verb in another tense or an adjective (e.g. synonyms of Recognition: identification, identify, recognize, recognise (intentionally misspelled), discern, discover, distinguish, etc.). Therefore, instead of showing all those words, only one word per synonym set is being displayed in the results. Synonyms can be provided by the user, or proposed by the toolkit, with a possibility of fine-tuning the proposals. For the considered use case, the list of used properties and property groups is shown in Fig. 1.

Property groups. The property groups are thematically, semantically or otherwise grouped properties for the purpose of more comprehensive presentation of the results. Properties within property groups are being displayed together in charts or tables. The property group has a name (e.g. Topics, Technology, Concerns, etc.), and within a group, there are sets of properties, including their synonyms, such as within the Concerns propriety group: privacy, security and acceptance. Exemplary summary results per property group are presented in 7, while exemplary results per property within groups are shown in Figs. 8, 9, 10 and 11.

Start year. The start year (inclusive) of the articles that we are interested in.
Default: current year - 9.

End year. The end year (inclusive) of the articles that we are interested in.
Default: current year.

Minimum relevant properties. A number denoting the minimum number of properties that an article has to contain in order to be considered as relevant.
Default: 2.

2.2 Enhanced Search Capabilities with WordNet

Before the actual searching starts, the user provided input in the form of keywords and properties is enhanced by proposing synonyms from WordNet [1, 12, 16], using the NLTK library [4] for Python. In most cases, this increases the robustness of the searched properties by including synonyms that the user might have neglected. However, considering that Word Net is a general-purpose database, some of the proposed synonyms might not be appropriate or relevant.



Fig. 1. List of property groups and properties (main and the synonyms)

In such a case, the user can manually choose which of the proposed synonyms to be included before the actual processing starts.

The toolkit also performs stemming of the properties and the abstract, for a more robust searching. If none of the properties of interest are identified within the abstract, then those articles are removed from the result set, which corresponds to the eligibility step in the PRISMA statement. In addition to this, we can specify the minimum number of properties that need to be identified within an article for it to be considered eligible and potentially relevant.

2.3 Indexed Digital Libraries

As of this moment, the NLP toolkit indexes the following digital libraries (i.e. sources): IEEE Xplore, Springer and PubMed. From PubMed all articles that match the given search criteria (i.e. a keyword) are analyzed. IEEE Xplore results include the top 2000 articles that match given criteria, sorted by relevance determined by IEEE Xplore. For the Springer digital library, the search for each keyword separately is limited to 1000 articles or 50 pages with results, whichever comes first, sorted by relevance determined by Springer.

2.4 Survey Methodology

The methodology used for the selection and processing of the research articles in this section is based on “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement” [13,14], as shown in Fig. 2. The goal of PRISMA is to standardize surveys. The first part is gathering articles based on certain criteria, in our case using the search keywords. After the articles are collected, the duplicates are removed and some of the articles are discarded for various reasons, such as relevance, missing meta-data, invalid publication period, etc. Finally, from the selected subset of articles, a qualitative analysis is performed and from those articles, only a certain number is selected for more thorough screening. With this toolkit, we automate most of the steps in the PRISMA approach to significantly reduce the number of articles that need to be manually screened.

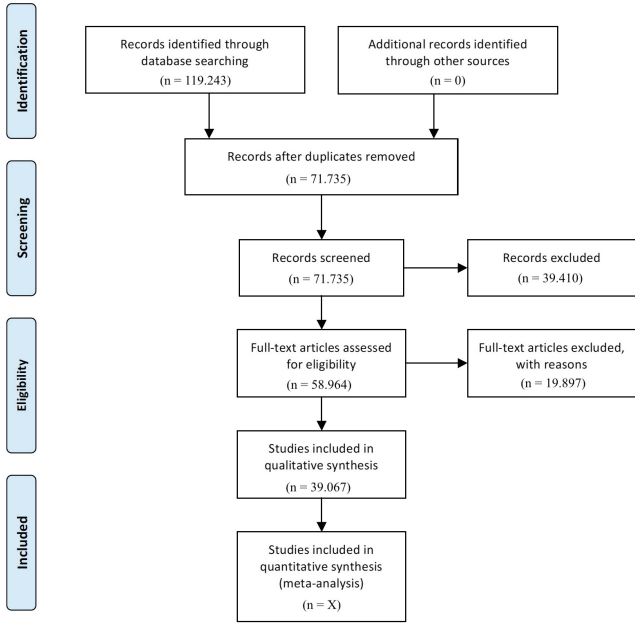


Fig. 2. PRISMA statement workflow with total number of articles for the current survey

Identification and Duplicate Removal. The proposed NLP toolkit performs the identification automatically. First, the possible article candidates are identified by querying the integrated libraries with the same search terms (i.e. keywords). While integrating the results from multiple sources (i.e. digital libraries), duplicate removal is also performed by using the article DOI as their unique identifier. Articles that were already found in another source or because they were identified by another search term, considering that an article can be found by multiple search terms, are not processed again, but still, are counted towards the number of identified articles per source. This means that the same article can be considered to exist in more than one source, therefore the sets of articles per source are not disjoint. After the candidate articles are identified, they are processed, and the properties of the texts are used for selection of the relevant articles. The process of article selection is the same as the one presented in [13,14], except for the last part where articles are manually processed by several researchers.

Augmented Screening and Eligibility Analysis by NLP. After the duplicates were removed, during the screening process discards articles which were not published in the required time period (e.g. last ten years) or for which the title or abstract could not be analyzed due to parsing errors, unavailability or other reasons.

Afterwards, the eligibility analysis is performed, which involves tokenization of sentences [10,23], English stop words removal, stemming and lemmatization [10] using the NLTK library [4] for Python. At the beginning, this is applied to each property, based on which a reverse lookup is created from each stemmed word and phrase to the original property. The same process is also applied to the title, keywords and abstract of each article. As a result of the stemming, for each property, the noun, verb and other forms are also considered. As a result of the lemmatization and the initial synonym proposal, the synonyms of properties are also considered. This results in a more robust analysis. Then, stemmed and lemmatized properties are searched in the cleaned abstract and title and the article is tagged with the properties it contains.

The identified articles are labeled as *relevant* only if they contain at least the minimum relevant properties, defined as an input, in its title or abstract (considering the above NLP-enhanced searching capabilities, thus performing a rough screening. To help in the eligibility analysis, the remaining relevant articles are sorted by number of identified property groups, number of identified properties, number of citations (if available) and year of publication, all in descending order. For the relevant articles the toolkit automatically generates a *Bibtex* file with most important fields that can be included in an article for simplified citations. An *Excel* file is also generated with the following fields: **DOI, link, title, authors, publication date, publication year, number of citations, abstract, keyword, source, publication title, affiliations, number of different affiliations, countries, number of different countries, number of authors, bibtex cite key, number of found property groups, and number of found properties.** The researcher can use this file to drill down and find specific articles by more advanced filtering criteria (e.g. by importing it in Excel). This can facilitate deciding which articles need to be retrieved from their publisher and manually analyzed in more detail in order to determine whether it should be included in the qualitative and quantitative synthesis.

Visualization of Aggregate Results. The results of the processing and retained relevant articles are aggregated by several criteria. The output contains CSV files and charts in vector PDF files for each of the following aggregate metrics:

- By source (digital library) and relevance selection criteria (see Fig. 3).
- By publication year (see Fig. 4a).
- By source and year (see Fig. 4b).
- By search keyword and source (see Fig. 5).
- By search keyword and year (see Fig. 6).
- By property group and year (see Fig. 7).
- By property and year, generating separate charts for each property group (see Figs. 8, 9, 10 and 11).
- By number of countries, number of distinct affiliations and authors, aiming to simplify identification of multidisciplinary articles (e.g. written by multiple authors with different affiliations) (See Fig. 14).

In addition to that, the toolkit also generates graph visualization of the results, where nodes are the properties and the edges are the number of articles that contain the two properties it connects. Articles which do not contain at least two properties and properties that are not present in at least two articles are excluded. An example of this is presented in Figs. 12 and 13. For a clearer visualization, only the top 25% property pairs by number of occurrence are shown (i.e. ones above the 75-th percentile).

A similar graph for the countries of the author affiliations is also generated (see Fig. 14). The top 50 countries by number of collaborations are considered for this graph. Additionally, we show only countries and an edge between them if the number of bilateral or multilateral collaborations between them in the top 5% (above 95-th percentile) within the top 50 countries.

3 Results

In this use case, we used the NLP toolkit with the keywords shown in Fig. 6. We searched for these keywords and automatically identified and screened the articles, as shown in Fig. 2. A more detailed analysis was performed using the properties that were clustered into four groups of properties, each containing at least three property synonyms, as shown in Fig. 1.

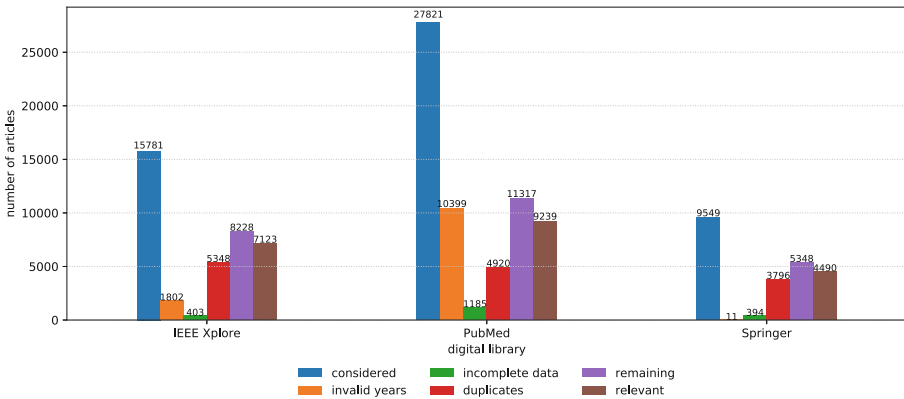
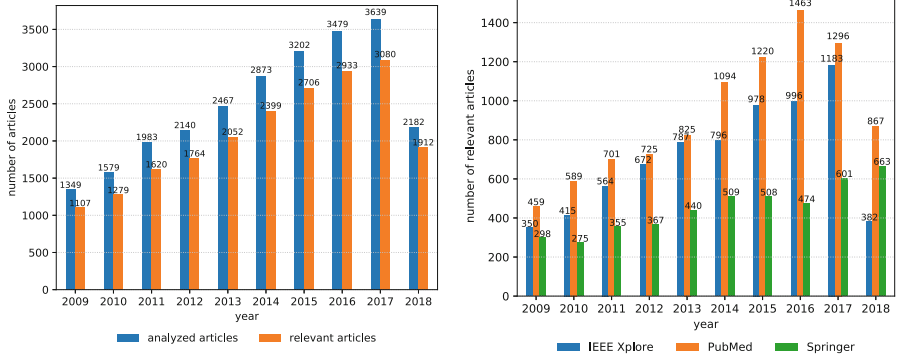


Fig. 3. Number of articles per relevance selection criteria

In Fig. 3, we show the selection process based on the adopted methodology. From all identified articles based on the keywords, first, the system eliminates the ones with incomplete or invalid meta-data. Next, the duplicate entries are eliminated and finally, from the remaining ones, the relevant articles are selected if they contain the minimum number of properties (in this case 1). In Fig. 4a, we present the number of remaining and searched for articles from each year, and in Fig. 4b, the number of relevant articles from each source.



(a) Number of remaining and relevant articles per year (b) Number of relevant articles from each digital library per year

Fig. 4. Number of articles per year and source

The number of relevant articles grouped by keywords from each source can be seen in Fig. 5. The top 3 keywords by the number of relevant articles are “assistive engineering”, “enhanced life environment” and “enhanced support environment”. It is interesting to see that they vary in frequency between different sources, which can be expected, considering that for PubMed the number of analyzed articles is unlimited, unlike the other sources.

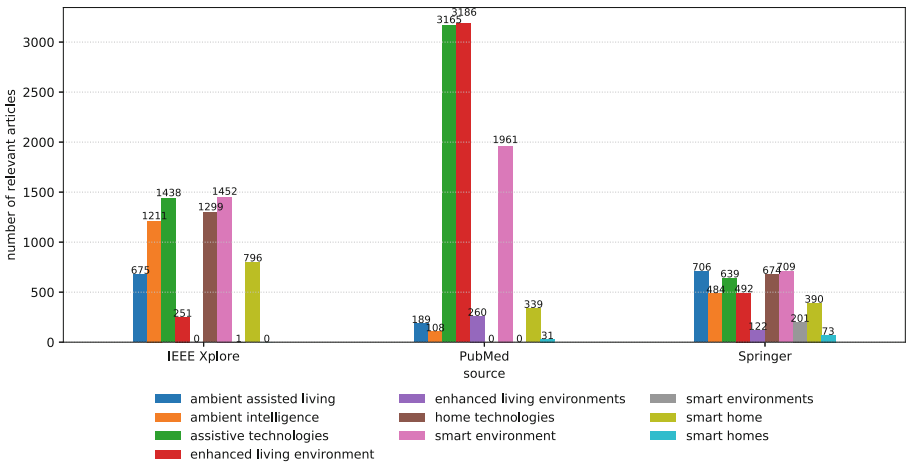


Fig. 5. Number of relevant articles for each keyword from each source

On Fig. 6, the distribution of articles per keyword for each year is shown. Notably, the number of papers for some of the keywords is increasing through the years, while for others it is relatively small.

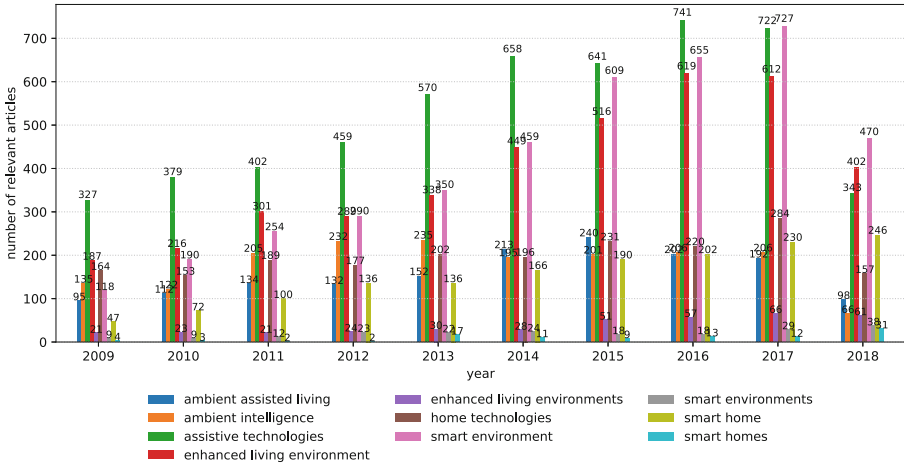


Fig. 6. Number of articles for searched keyword per year

Next, in Fig. 7 we can see the trends of articles mentioning at least one property from each property group, and evidently, all property groups are becoming more relevant. Apparently, the articles are not covering data management as often as the other themes (i.e. technology, topics and information delivery and prescriptive insight).

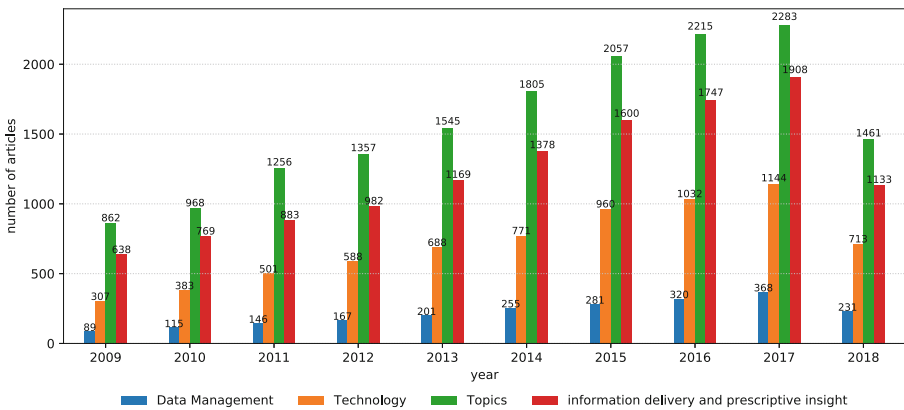


Fig. 7. Number of articles mentioning each property group per year

Properties and keywords follow a similar trend in the number of articles, with most of them reaching the highest number in 2015 and 2016. However, some terms, such as “smart environments”, is still on the rise. Note that the numbers from 2018 are inconclusive because, at the time of this analysis, 2018

is not yet finished. Also, the number of articles is increasing in IEEE Xplore and Springer and the in PubMed the number of articles starts decreasing after 2016.

After the initial property analysis, for each property group, we analyze the articles based on each property. In Fig. 8 the results about the Data Management property group is shown. Here, we consider the properties Cloud, Fog and Edge, and their synonyms. The observable trend is that all of the terms are increasing in popularity in the respective research communities. The most popular term in the articles is Edge followed by Cloud and finally Fog computing, which slowly and steadily increases in popularity.

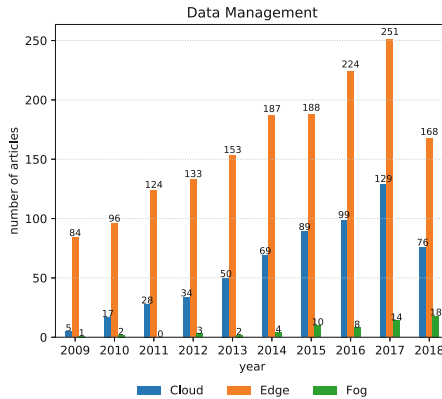


Fig. 8. Article distribution per year and properties in Data Management property group

The second property group is the “Technology”, which is consisted of the properties: Battery, Deep Learning, Machine Learning, Protocol and Sensor. These properties cover different technology groups within the surveyed articles. It can be observed that most of the published articles include Sensors and give observation regarding the power consumption, thus include the word battery. Communication is also one of the most popular topics, the word protocol is also often mentioned, while Machine Learning and Deep Learning are encountered sparsely, but are slowly increasing in popularity.

The third property group “Topics” includes the properties: Activities, Accidents, Diet, Exercises, Mobile and Vital Signs. The topics show increasing trends in all of these properties, except for vital signs and accidents. We reason that this is due to the fact that most of the studies that are intended for Enhanced living environments are more interested in prevention and well-being instead of treatment. Accidents and vital signs measurements are also much harder to simulate and need specific hospitals environments to be treated. This does not mean that they are less relevant, rather that it is simply a less attractive research topic.

The final group of properties, “Information delivery and prescriptive insight”, contains Sensing, Recognition, Monitoring and Supervision. It can be observed

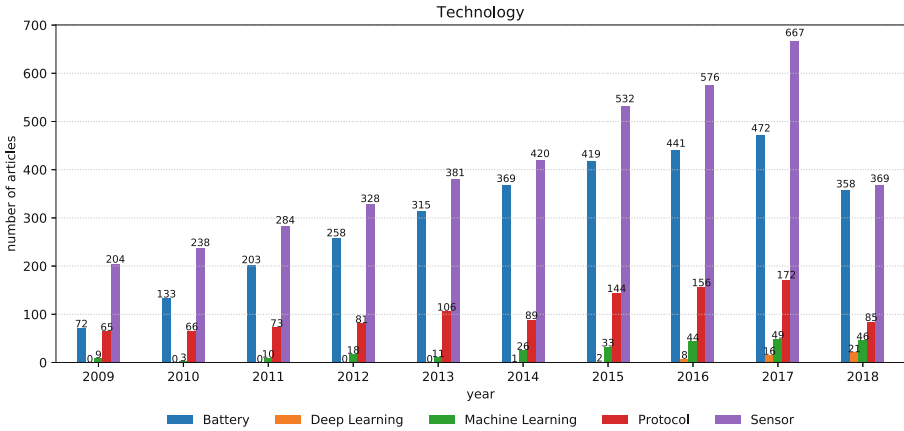


Fig. 9. Article distribution per year and properties in technology property group

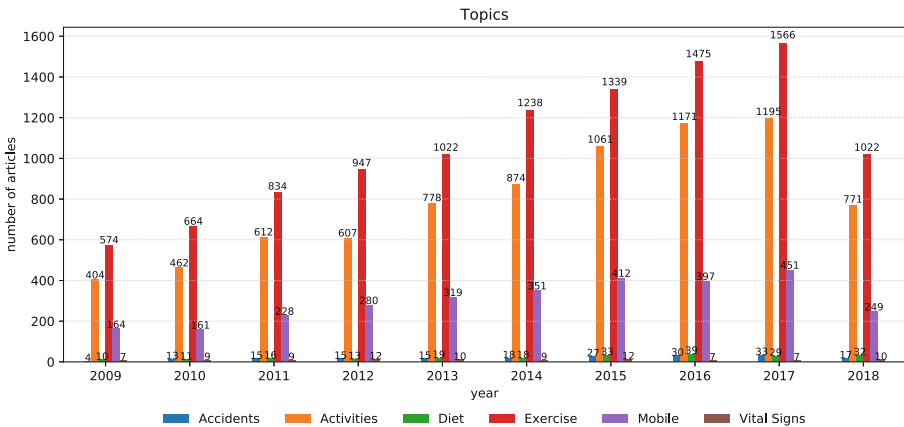


Fig. 10. Article distribution per year and properties in topics property group

that most of the publications are treating Sensing and Recognition and much less Monitoring and Supervision. The latter are much harder to study because of the special regulations related to ethical and processing of human data. The first two, Sensing and Recognition are much easier to simulate and there are many available datasets.

Next, Figs. 12 and 13 show how different properties are related between each other in terms of how often they occur together in the same article. These graphs can be used for guiding the drilling down process and selection of articles that need to be analyzed manually. The darker an edge is, the more articles there are that have the connected keywords. Also it shows that some properties are not often encountered with others (e.g. Cloud and Supervision on Fig. 13).

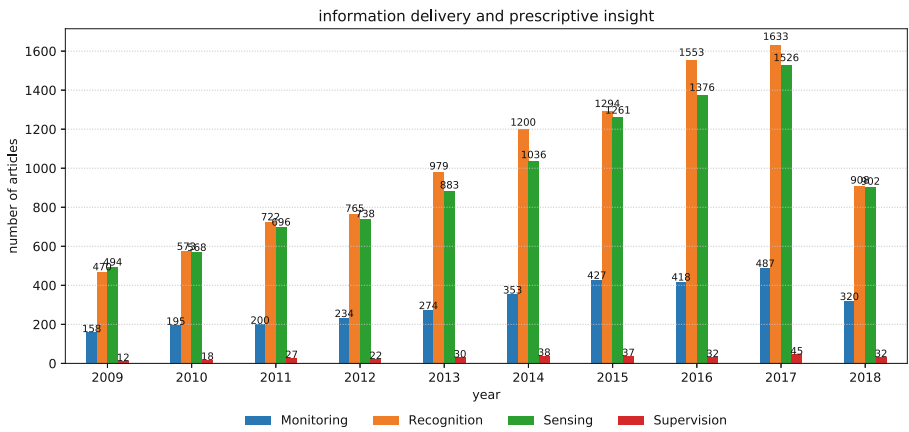


Fig. 11. Article distribution per year and properties in the information delivery and prescriptive insight property group

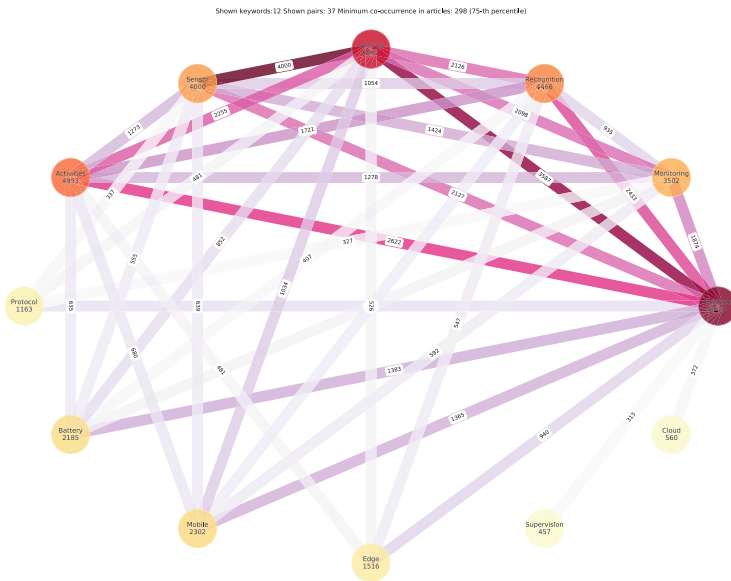


Fig. 12. Graph visualization with circular layout relevant articles by properties. Node labels show the property and number of articles that contain it and edge label shows the number of papers that have the properties it connects.

Finally, Fig. 14 shows how authors from different countries collaborated. This graphs clearly shows that communities exist between some countries. In most cases, we attribute this to geographical location, smaller language barriers, or both.

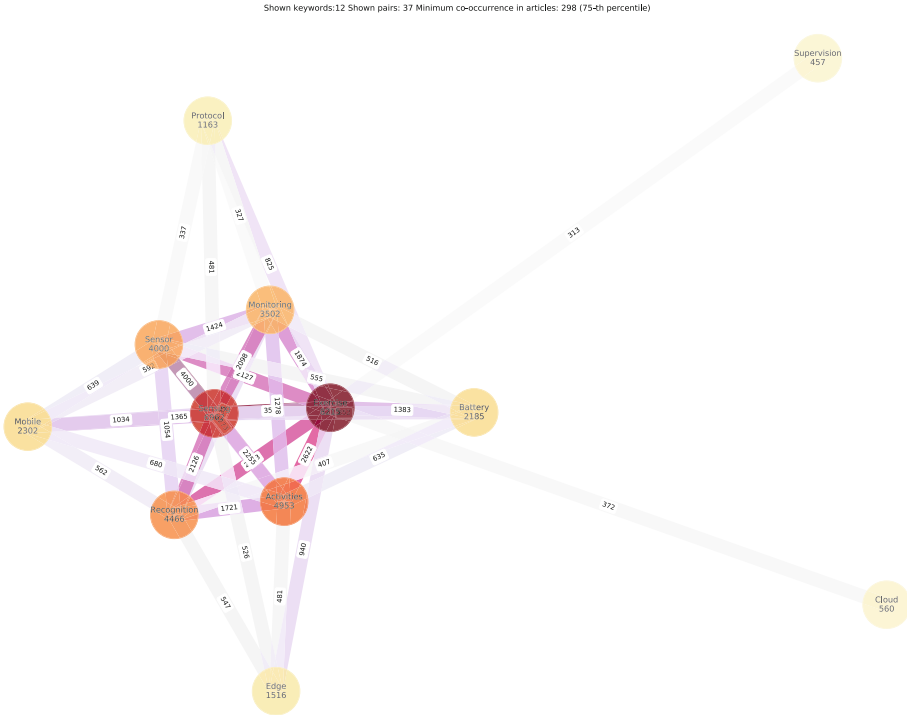


Fig. 13. Graph visualization with circular (i.e. Fruchterman-Reingold) layout relevant articles by properties.

4 Discussion

From this scoping review we can notice some increasing trends over different search keywords over the last decade (see Fig. 6). However, some keywords, such as “ambient assisted living” and “ambient intelligence” this trend is in a declining in the last 5 years. On the contrary, the trend for “assistive technologies” is in an even more rapid increase in the last 5 years, compared to its trend in the last 10 years. Interestingly, the singular form of “enhanced living environment”, “smart environment” and “smart home” consistently results in finding more relevant papers than their plural form. From the properties, “deep learning” started to gain attention only in the last few years.

The proposed NLP toolkit was demonstrated through the AAL/ELE use case in this paper. It was also applied to simplify the review process in several previous works [9,11]. Its continued improvement is owed to the constructive feedback obtained from multiple researchers that had tested it. By being able to reuse intermediate results and allowing tweaking and fine-tuning of keywords and properties, the researcher can test different alternatives of keywords and properties very quickly. The toolkit also provides ability to fine-tune the graph plotting thresholds, so the they can show appropriate number of edges. These

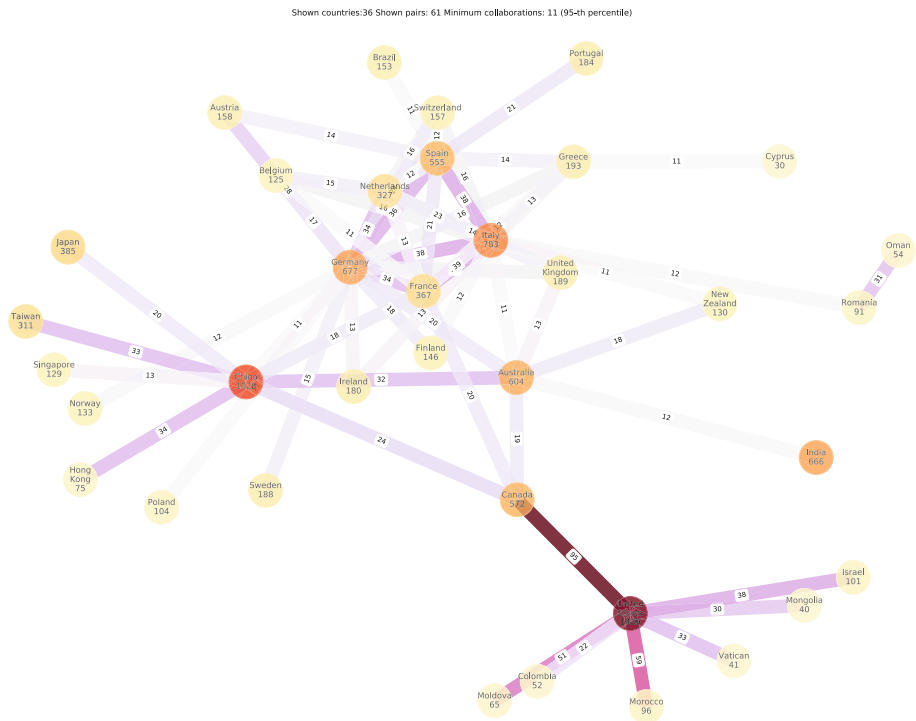


Fig. 14. Graph visualization relevant articles by countries. Node labels show the country and number of publications from it, while edge labels show the number of papers that were published by authors with affiliations from the countries it connects.

default parameters were empirically determined based on extensive analysis with over dozens of different use-cases.

Even though the results of the processing are automatically emailed to the researcher that started the analysis, the toolkit is still lacking a user interface. Right now, we are working on implementing a web-based user interface that will make the toolkit easily available for other researchers. Meanwhile, interested readers are encouraged to contact us for providing the source code or jointly performing a systematic or scoping review.

Another upcoming issue, as the number of users is increasing, is the scalability of the system. Even though we use a Microsoft Azure hosted instance for the toolkit, in order the system to be able to process multiple requests at once we need a more scalable solution, such as one based on Hadoop [24,27].

5 Conclusion

In this paper, we presented an NLP toolkit for speeding up the process of surveying scientific articles and trend analysis meta-studies. By leveraging NLP,

it facilitates a robust and comprehensive eligibility and relevance analysis of articles, so the user can focus on reading a small number of potentially relevant articles.

We have presented a use-case of the proposed framework that proves that the framework is able to analyze the abstracts of over 70000 articles automatically and visualize different trends of interest.

For this use case, we can conclude that almost all of the searched keywords and properties have an increasing trend over the years. The aggregate results show that the research community is more interested in Enhanced living environments that sense and recognize activities and aid exercising, thus helping the well-being of people. Monitoring and supervision, and also more serious health issues, such as accidents and vital signs have received less attention from the scientific community. Furthermore, regarding the way the data is processed, Edge computing and Cloud computing receive fairly large attention. Sensors and power consumption are more interesting for researchers than communication protocols and machine/deep learning.

Acknowledgment. This work was partially financed by the Faculty of Computer Science and Engineering at the Ss. Cyril and Methodius University, Skopje, Macedonia and is supported by the networking activities provided by the ICT COST Actions IC1303 AAPELE and CA16226 SHELD-ON. We also acknowledge the support of Microsoft Azure for Research through a grant providing computational resources for this work.

References

1. Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Paşca, M., Soroa, A.: A study on similarity and relatedness using distributional and wordnet-based approaches. In: *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 19–27. Association for Computational Linguistics (2009)
2. Alam, M.M., Hamida, E.B.: Surveying wearable human assistive technology for life and safety critical applications: standards, challenges and opportunities. *Sensors* **14**(5), 9153–9209 (2014)
3. Alla, A., Zdravevski, E., Trajkovik, V.: Framework for aiding surveys by natural language processing. In: *Web Proceedings of the ICT Innovations 2017 Conference, IKT-AKT* (2017)
4. Bird, S.: NLTK: the natural language toolkit. In: *Proceedings of the COLING/ACL on Interactive Presentation Sessions*, pp. 69–72. Association for Computational Linguistics (2006)
5. Dimitrievski, A., Zdravevski, E., Lameski, P., Trajkovik, V.: A survey of ambient assisted living systems: challenges and opportunities. In: *2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 49–53. IEEE (2016)
6. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using networkx. In: Varoquaux, G., Vaught, T., Millman, J. (eds.) *Proceedings of the 7th Python in Science Conference, Pasadena, CA USA*, pp. 11–15 (2008)

7. Hunter, J.D.: Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**(3), 90–95 (2007)
8. Kotevska, O., Vlahu-Gjorgievska, E., Trajkovik, V., Koceski, S.: Towards a patient-centered collaborative health care system model. In: 4th IEEE International Conference on Computer Science and Information Technology (IEEE ICCSIT 2011) (2011)
9. Lameski, P., Zdravevski, E., Kulakov, A.: Review of automated weed control approaches: an environmental impact perspective. In: Kalajdziski, S., Ackovska, N. (eds.) *ICT 2018. CCIS*, vol. 940, pp. 132–147. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00825-3_12
10. Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60 (2014)
11. Maresova, P., et al.: Technological solutions for older people with alzheimer’s disease. *Current Alzheimer research* (2018)
12. Miller, G.A.: WordNet: a lexical database for English. *Commun. ACM* **38**(11), 39–41 (1995). <https://doi.org/10.1145/219717.219748>
13. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Med.* **6**(7), 1–6 (2009). <https://doi.org/10.1371/journal.pmed.1000097>
14. Moher, D., et al.: PRISMA-P group: preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **4**(1), 1 (2015). <https://doi.org/10.1186/2046-4053-4-1>
15. Moher, D., Stewart, L., Shekelle, P.: All in the family: systematic reviews, rapid reviews, scoping reviews, realist reviews, and more. *Syst. Rev.* **4**(1), 183 (2015). <https://doi.org/10.1186/s13643-015-0163-7>
16. Pedersen, T., Patwardhan, S., Michelizzi, J.: WordNet: similarity: measuring the relatedness of concepts. In: *Demonstration papers at HLT-NAACL 2004*, pp. 38–41. Association for Computational Linguistics (2004)
17. Peters, M.D., Godfrey, C.M., Khalil, H., McInerney, P., Parker, D., Soares, C.B.: Guidance for conducting systematic scoping reviews. *Int. J. Evid.-Based Healthc.* **13**(3), 141–146 (2015)
18. Pombo, N., Garcia, N., Bousson, K.: Machine learning approaches to automated medical decision support systems. In: Pandian, V. (ed.) *Handbook of Research on Artificial Intelligence Techniques and Algorithms*, pp. 183–203. IGI Global, Hershey (2015)
19. Poon, C.C., Lo, B.P., Yuce, M.R., Alomainy, A., Hao, Y.: Body sensor networks: in the era of big data and beyond. *IEEE Rev. Biomed. Eng.* **8**, 4–16 (2015)
20. Rashidi, P., Mihailidis, A.: A survey on ambient-assisted living tools for older adults. *IEEE J. Biomed. Health Inform.* **17**(3), 579–590 (2013)
21. Suci, G., et al.: Big data, internet of things and cloud convergence-an architecture for secure e-health applications. *J. Med. Syst.* **39**(11), 141 (2015)
22. Trajkovik, V., Vlahu-Gjorgievska, E., Koceski, S., Kulev, I.: General assisted living system architecture model. In: Agüero, R., Zinner, T., Goleva, R., Timm-Giel, A., Tran-Gia, P. (eds.) *MONAMI 2014. LNICST*, vol. 141, pp. 329–343. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16292-8_24
23. Webster, J.J., Kit, C.: Tokenization as the initial phase in NLP. In: *Proceedings of the 14th Conference on Computational Linguistics*, vol. 4, pp. 1106–1110. Association for Computational Linguistics (1992)

24. Zdravevski, E., Lameski, P., Kulakov, A., Jakimovski, B., Filiposka, S., Trajanov, D.: Feature ranking based on information gain for large classification problems with mapreduce. In: Proceedings of the 9th IEEE International Conference on Big Data Science and Engineering, pp. 186–191. IEEE Computer Society Conference Publishing, August 2015. <https://doi.org/10.1109/Trustcom-BigDataSe-ISPA.2015.580>
25. Zdravevski, E., et al.: Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering. *IEEE Access* **5**, 5262–5280 (2017). <https://doi.org/10.1109/ACCESS.2017.2684913>
26. Zdravevski, E., Kulakov, A.: System for prediction of the winner in a sports game. In: Davcev, D., Gómez, J.M. (eds.) *ICT Innovations 2009*, pp. 55–63. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-10781-8_7
27. Zdravevski, E., Lameski, P., Kulakov, A., Filiposka, S., Trajanov, D., Jakimovski, B.: Parallel computation of information gain using hadoop and mapreduce. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*. *Annals of Computer Science and Information Systems*, vol. 5, pp. 181–192. IEEE (2015). <https://doi.org/10.15439/2015F89>
28. Zdravevski, E., Risteska Stojkoska, B., Standl, M., Schulz, H.: Automatic machine-learning based identification of jogging periods from accelerometer measurements of adolescents under field conditions. *PLOS ONE* **12**(9), 1–28 (2017). <https://doi.org/10.1371/journal.pone.0184216>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

