SPORENLP: A Spatial Recommender System for Scientific Literature

Johannes Wirth [a], Daniel Roßner [b], René Peinl [c] and Claus Atzenbeck [d]
Institute of Information Systems, Hof University, Alfons-Goppel-Platz 1, Hof, Germany

Keywords: Recommendation Systems, Scientific Literature, Spatial Hypertext, Natural Language Processing.

Abstract: *SPORENLP* is a recommendation system designed to review scientific literature. It operates on a sub-dataset comprising 15,359 publications, with a total of 117,941,761 pairwise comparisons. This dataset includes both metadata comparisons and text-based similarity aspects obtained using natural language processing (NLP) techniques.Unlike other recommendation systems, *SPORENLP* does not rely on specific aspect features. Instead, it identifies the top k candidates based on shared keywords and embedding-related similarities between publications, enabling content-based, intuitive, and adjustable recommendations without excluding possible candidates through classification. To provide users with an intuitive interface for interacting with the dataset, we developed a web-based front-end that takes advantage of the principles of spatial hypertext. A qualitative expert evaluation was conducted on the dataset. The dataset creation pipeline and the source code for *SPORENLP* will be made freely available to the research community, allowing further exploration and improvement of the system.

1 INTRODUCTION

Recommender systems have become an indispensable aspect of our daily lives, as they facilitate the discovery of new products, movies, and music, among other things. An area that has received increasing attention in recent years is the utilization of paper recommendation systems (Kreutz and Schenkel, 2022). Typically, such systems support exploratory search by providing appropriate papers or a collection of literature based on given papers or keywords that may be of interest to the user. Connections between articles are established using a wide range of sources of information, including basic factors such as shared authors or keywords, as well as more sophisticated techniques such as computation and comparison of textual content embeddings (Collins and Beel, 2019).

An inherent challenge is designing a user-friendly system that does not necessitate intricate input while simultaneously furnishing highly pertinent paper recommendations that align with the users' requirements. These requirements may differ, as junior researchers, senior researchers, and students may possess varying needs (Bai et al., 2019). When researchers incorporate a topic with which they are not yet familiar, their information needs typically include obtaining an overview of the topic, exploring specific aspects of the topic, and summarizing, as well as organizing their findings. To address this issue, we propose a spatial hypertext interface that simplifies the expression of contextual information associated with search queries, while offering the ability to spatially organize the information retrieved.

Hypertext is a crucial technology for linking information and enabling access to related pieces of information. In the realm of recommender systems, users typically have multifaceted and diverse preferences that are not easily captured by simple linear models. Spatial hypertext offers an alternative by allowing users to navigate and explore recommendation results non-linearly, thereby facilitating the discovery of unexpected and serendipitous recommendations. Our approach is rooted in a knowledge graph constructed with data obtained from *Semantic Scholar*[1]. The dataset comprises publication titles, authors, identifiers, abstracts, and relationship metrics that we derived from metadata and text features. The dataset itself as well as a pipeline for the creation of more data is publicly available at OSD[2], thereby increasing

[a] https://orcid.org/0009-0002-0666-7693
[b] https://orcid.org/0000-0002-2539-569X
[c] https://orcid.org/0000-0001-8457-1801
[d] https://orcid.org/0000-0002-7216-9820

[1] https://www.semanticscholar.org/
[2] https://opendata.iisys.de/opendata/Datasets/publication_similarity.zip

Wirth, J., Roßner, D., Peinl, R. and Atzenbeck, C.
SPORENLP: A Spatial Recommender System for Scientific Literature.
DOI: 10.5220/0012210400003584
In *Proceedings of the 19th International Conference on Web Information Systems and Technologies (WEBIST 2023)*, pages 429-436
ISBN: 978-989-758-672-9; ISSN: 2184-3252

the transparency of our findings and benefiting the research community engaged in this area of study.

2 RELATED WORK

This research integrates elements from various fields, such as knowledge graph construction for scientific literature, hypertext research, and natural language processing. We expand on existing work to create semantically meaningful text embeddings for calculating distances between papers. The interface we propose is rooted in spatial hypertext research, which centers on conveying implicit relationships through visual elements in a primarily 2D space. To the best of our knowledge, we identified shortcomings in existing systems that we leveraged to justify our approach.

2.1 Embeddings as a Similarity Measure

Xiaofei et al. (Ma et al., 2019) are one of the first to analyze BERT embeddings as universal text representations. They find that top and bottom layer embeddings are most useful and outperform a strong BM25 (Amati, 2009) baseline and already deliver reasonable results for question answering, text classification, and semantic similarity without any finetuning. It is therefore reasonable to use them for enhancing the evaluation of results from generative AI models on tasks like machine translation or open-ended question answering. The predominant measures for these tasks are BLEU, METEOR and ROUGE. Zhang et al. (Zhang et al., 2019) show that BERTscore, a similarity measure based on BERT embeddings and cosine similarity, outperforms all these measures with respect to semantic meaning in machine translation and image captioning tasks.

Based on these earlier advances, SPECTER (Cohan et al., 2020) was developed to learn general vector representations of scientific documents that additionally takes into account document-level relatedness derived from the citation graph during the training process. At the time of publication, this language model performed best on SciDocs, a benchmark for scientific document representations, while its successor, SPECTER2, outperformed even newer approaches on the now state-of-the-art and recommended SciRepEval benchmark (Singh et al., 2022).

2.2 Spatial Hypertext

Spatial hypertext uses a (mostly 2D) space on which informational units are organized spatially. Their spatial proximities or alignments implicitly represent the associations among the objects. Furthermore, the nodes' visual appearances suggest associations between similar-looking objects. The so created structure is *implicit* by nature (Shipman III et al., 1995) and appears by interpretation. Machines that can interpret the space to reach a similar "understanding" as the user are called *spatial parsers*. Some early attempts existed around the 2000s (Reinert et al., 1999). Schedel (Schedel, 2017) further developed this approach by introducing specialized parsers, each designed to focus on a specific attribute such as color, proximity, or shape. The so-earned workspace awareness, common between user and machine, enables an iterative process of creating a context of information by the user that is augmented by the machine's recommendations derived from its computational knowledge. It is the foundation and necessary requirement for spatial hypertext-based recommendation systems (Atzenbeck et al., 2023).

The visualization of recommendations is not only about providing high quality results, but also about explaining them in a reasonable context, which is essential in information retrieval and visualization (Beckmann and Gross, 2010). To go beyond conventional list representations, researchers in the field of information visualization have explored solutions that use two or even three dimensions to encode additional information (Parra, 2012; Bitton, 2009). Mostly, applications use the proximity of objects in a space to express the strength of relationships. In simple words, the closer two objects are to each other, the stronger the relationship. An important finding is that information visualization amplifies cognition. The perceptual system can reduce the cognitive load because it is well trained to observe changes, even if they are not in focus, and to recognize patterns.

2.3 Scientific Paper Recommendation Systems

The most recent literature review (Kreutz and Schenkel, 2022) of published recommender systems for scientific publications provides a detailed overview of different systems, algorithms, and use cases and lists concrete aspects of the challenges that still need to be overcome for a productive system or have not been solved across the board. These include technical and qualitative features such as scalability and accuracy, as well as many user-dependent aspects such as explainability by confidence and adaptability by user preferences. Although not all challenges are discussed in detail, an effort was made to address as many issues as possible.

3 SYSTEM

The following subsections describe the overall system presented and the underlying basic concepts for the dataset used along with the user interface for the recommendation system. In addition, the implemented approaches are put into relation with those of other systems and mechanisms for the identification of document similarity.

3.1 Dataset

The dataset incorporated into the presented recommender system consists of data from 15,359 publications, the majority (15,000) of which are listed on Semantic Scholar[3] and have full text published under an open access license. Metadata for these publications were obtained using the publicly available API of Semantic Scholar. In addition, the corresponding full text documents were downloaded. To include at least loose context references to a large extent, papers were randomly selected whose titles are related to a small collection of keywords related to deep learning and NLP. The remaining papers (359) were presented at the ACM Conference on Hypertext and Social Media and are additionally included to possibly identify similarities in the scope of a conference or its tracks to explore the recommendation of fitting venues for new papers. Since full texts for publications presented at this conference are subject to a more restrictive license, the published dataset does not contain their entire content, but only publicly available metadata and final calculated metrics, which are not indicative of the actual content of these works. The original full texts were obtained from the authors' records.

Using *GROBID* (GRO, 2023), full texts and additional metadata were extracted from text documents and publication title, keywords, keyphrases (2- to 4-grams), text embeddings, references, direct citations, authors, document identifiers (doi, arxiv, etc.) as well as chapter names were determined for subsequent pairwise comparisons. Similarly to the approach of (Renuka et al., 2021), keywords and keyphrases were extracted from abstracts and fulltexts, but were not further processed into word vectors using term frequency-inverse document frequency (TF-IDF), but are included as-is in the dataset to allow for keyword-based queries. Keywords were retrieved and lemmatized with *SpaCy* (Honnibal et al., 2020), keyphrases were extracted using Rapid automatic keyword extraction (RAKE) as demonstrated by (Rose et al., 2010).

[3]https://www.semanticscholar.org/product/api

As shown in (Cohan et al., 2020) and (Singh et al., 2022), the language model SPECTER2 achieves state-of-the-art performance in several benchmarks for the representation of scientific documents. The use of a language model instead of conventional methods such as TF-IDF also overcomes the challenge of dealing with synonyms, as described in (Kreutz and Schenkel, 2022). The unmodified, pretrained retrieval model was used to generate embeddings of the titles and abstracts from publications to subsequently determine cosine similarity among publication pairs. The publicly available, pretrained checkpoint "allenai/specter2_proximity" was used.

Based on the previously extracted metadata and text properties, all publications were compared pairwise with respect to matching components. The dataset consists of a total of 117,941,761 publication pairs (combination without repetition of publication pairs with $n = 15359$, $r = 2$) for which matches between abstract keywords and keyphrases, fulltext keywords and keyphrases, chapter names, authors and referenced works were determined as well as the cosine similarity between abstract embeddings was calculated.

Pairwise comparisons based on these aspects were cleaned up, tagged with the previously described metadata, and finally converted to TSV format to allow easy import into the associated recommender system. All comparisons performed can be used in the recommendation system described hereafter for user-specific queries to create recommendations not only based on multiple metrics, but also to strongly weight certain keywords and phrases of interest and to directly compare chapter sections (e.g. comparison of the chapter "Methodology" between several papers dealing with the same topic). These aspects not only increase the explainability of recommendations to users, but also allow for fine-grained customization. The implemented dataset creation pipeline will be made publicly available so that users can extend the data according to their needs.

3.2 Interface

To serve as a front-end for the dataset presented in this paper, we have developed a web-based interface that leverages the principles of spatial hypertext, as described in Section 2.2. By representing the relationships between entities visually, the interface enables users to establish a contextual framework that can be utilized to formulate queries for the underlying knowledge graph and to organize their knowledge. In the context of exploratory search, we want to foster an iterative approach whereby the system recom-

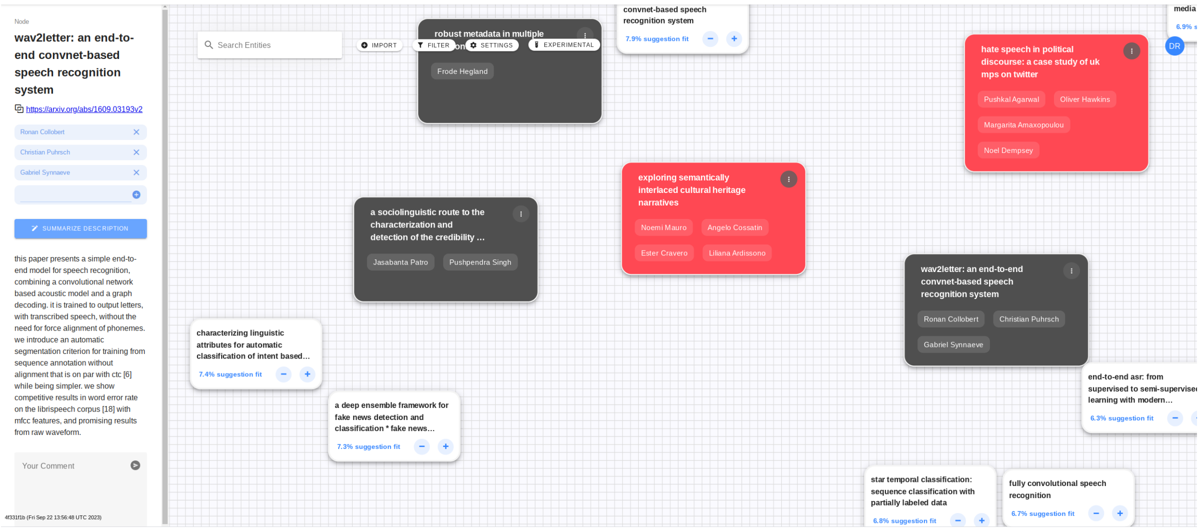

Figure 1: Screenshot of the Web user interface, with user entities and suggestions (white boxes).

mends relevant literature to the user, who can then react by modifying the query context or incorporating the recommended materials into the interface. This process is designed to be highly interactive, and users can use visual cues to guide the exploration and organization of their findings. This iterative approach can be repeated until the user's information needs are met, allowing for a customized and effective research experience.

Following the infrastructure proposed in *Mother* (Atzenbeck et al., 2018), our application is divided into three distinct layers that operate independently of each other. The first layer comprises the knowledge component, which in our case is a Neo4J graph database responsible for managing the dataset described in Section 3.1. The second layer is responsible for managing the structure and associated services, primarily related to the spatial structure within the user interface, but may also encompass link structure services (Carr et al., 1995) or hierarchies and other types of structures. The third and final layer includes the user interface, which is complemented by additional software components, such as a web server and an API, in the case of a web application.

The fundamental concept driving the system is that of entities which can represent any kind of data without any specific constraints imposed by the implementation. Generally, an entity can contain arbitrary text, a file, or a URI. Since entities are also employed as vertices in the knowledge graph, they are intended to be unique. Our generated dataset is used to create individual entities for each publication, encompassing the title, abstract, and an identifier to link the publication (e.g. the DOI).

The UI layer houses the Web interface, which comprises 2D spaces that facilitate entity organization and recommendation retrieval. Multiple *workspaces* can be created and may be used concurrently by numerous users. Figure 1 illustrates a workspace screenshot, with five entities organized by the user (three dark gray and two red boxes) along with six recommendations (white boxes). The left-hand side of the workspace displays an information bar that provides additional information, such as tags, abstract, and comments, for the selected entities. At the top, a toolbar is available, allowing the addition of entities and customization of various settings to personalize the recommendations provided. The interface is designed to provide an optimized user experience for both desktop and tablet devices.

The organization and management of the behavior of the suggested entities is done using an algorithm described in (Roßner et al., 2019), cf. Section 2.2. The basic idea is to exploit the proximity of objects to encode their meaning and relation within the space (Chalmers and Chitson, 1992). To achieve this, the algorithm uses Planck.js[4] and a spring metaphor. When provided with accurate weights to parameterize the springs, the algorithm renders recommendations at positions that make sense to users and controls their position to react to user interactions with smooth transitions. The structure layer, with its capabilities to manage and interpret spatial layouts, is of particular interest. Spatial parsers within this layer are used to monitor the space that users are working on and provide an interpretation based on the size, position, color, shape, temporal interactions and how these are related to properties of other objects in the space (e.g.

[4] *Typescript* re-write of Box2D, a physics engine for rigid bodies: https://box2d.org/

lists or lists of lists). As a result, the parsers calculate a weighted graph that is used to infer objects that are visually and temporally related on the basis of interactions. The weights range from 0 to 1 and denote the visual relation between two objects. An algorithm for detecting groups removes edges below a certain threshold (Schedel, 2017), leading to a sparse graph with interconnected groups/clusters of entities. This interpretation is used to achieve two goals: *Query generation* – as the structure service is aware of related objects, it can transform this information into multiple queries, one for each *visual group*. The result of each query is a set of suggestions that are relevant to the combination of entities within the respective group. *Refine the knowledge base* – The visual memory of humans fosters the organization of objects in visual groups (Brady et al., 2011). Therefore, the system is designed to utilize the visual grouping and organization of objects to refine the knowledge base. The system adapts to user behavior by monitoring the creation, alteration, or dissolution of object groups, adjusting the relationship weights in the knowledge base accordingly. This enables the system to refine its recommendation accuracy over time.

A typical user session begins with an empty workspace, using default settings such as 3 suggestions, group detection enabled, and the same weight of 0.7 for all metrics. If group detection is disabled, suggestions pertain to the entire workspace, ignoring any visual structure set by the user. From here, the user starts with searching publications in the search field. Currently, the system indexes authors, titles and tags for search. Search results are presented in a list format, with additional information, such as the abstract, being revealed by hovering over a specific entry. Selecting an entry adds the publication to the center of the current viewport and updates the set of shown *suggestion nodes*.

More nodes can be added by *accepting* suggestions by pressing on the '+' button of a suggestion or by using the search field again. Suggestions may be replaced with others by pressing '-'. The filter section facilitates fine-tuning the influence of each metric (ranging from 0 to 1) on the recommendation calculation. By manipulating one or more filters, a new weighted average is calculated, and an adjusted set of suggestions is being integrated into the workspace. Users can customize filters according to their specific needs, and the system promotes exploration of these settings. Interactions like adding new publications or tweaking filters dynamically update suggestions. Irrelevant suggestions fade out, while new and existing ones adjust to the current context. This organization helps in thought modeling and allows context-aware publication suggestions. Users can add tags to publications for better discoverability and leave comments for shared insights (*cf.* left side of Figure 1). A quick summary of each publication's abstract is generated on demand. New publications, URLs, and images can be imported, and PDF metadata is automatically extracted using GROBID. Live updates facilitate real-time collaboration, although suggestion settings remain user-specific.

4 EVALUATION

To automatically evaluate the similarity of publication pairs, the absolute number of shared references is defined as an objective similarity value, and all publication pairs are set in proportion to shared references using the metrics described in Section 3.1 to determine at which metric in this regard a correlation can be observed.

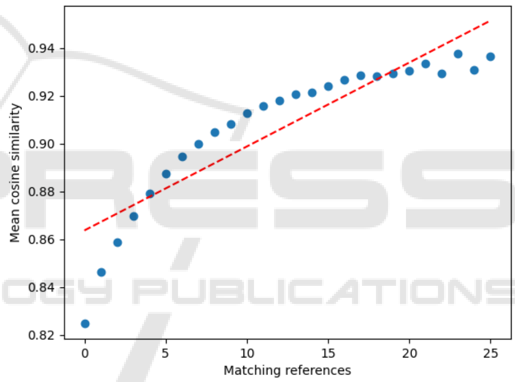

Figure 2: Mean cosine similarity for abstract embeddings (y-axis) in relation to common references (x-axis).

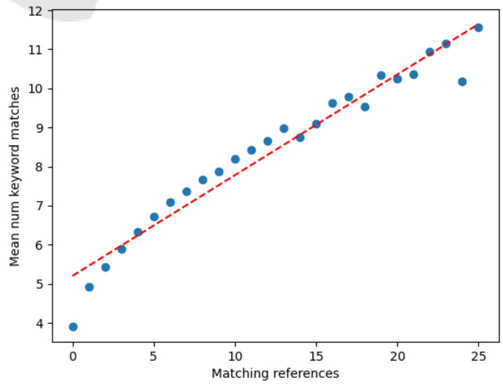

Figure 3: Mean number of matching keywords in abstracts (y-axis) in relation to common references (x-axis).

As shown in Figure 2, the mean cosine similarity between abstract embeddings and the number of shared references correlate well, confirming the us-ability of cosine similarity as a base metric for the recommendation system, as already indicated by the benchmarks shown in (Singh et al., 2022). A similar correlation can also be seen in mean matching keywords for abstracts, which is unsurprising, as text embeddings created with SPECTER2 also embed the keywords themselves quite well. However, as can also be observed in Figure 3, publications with few or no matching references may still share multiple keywords of potential interest for users. When searching for specific shared keywords within the recommender system, a lower weighting of cosine similarity could also lead to the recommendation of publications that do not have relevant similarity, but, for instance, use the same methodologies in different application domains, which can be very useful for literature reviews. Furthermore, statistics on correlations between the number of shared references with matching keywords of full texts as well as matching keyphrases in abstracts and full texts were also conducted, which show a similar correlation as the previously described relations, but significantly less prominent. In the case of keyphrases for abstracts, this can be attributed to the fact that N-grams with $N > 1$ have a naturally lower occurence probability than single keywords. In terms of the statistics of keywords and keyphrases found within the full texts, it is worth noting that there may be numerous less relevant words and phrases present. These may not directly correlate with the main objective of a published work and, as a result, are unlikely to be frequently found in other publications, despite overall similarities.

A qualitative user study was also conducted that involved four participants (two junior researchers, two senior researchers) for a more detailed evaluation of the metrics. All participants were asked to use the complete system for an example literature search by following the following instructions.

1. Choose a publication stored in the system as a starting point.
2. Are the suggested publications similar to the root publication in one or more of the following aspects? Background, Objective, Method
3. Do the proposed publications complement or expand upon the root publication in one or more of the following aspects? Objective, Method, Results
4. Select a proposed publication to which at least one aspect from question (3) applies and add it as a new node.
5. Repeat the previous instruction for two and three related nodes that reside in your workspace.

To determine similarity with respect to the different aspects in the task description, a simple rating scale was introduced, consisting of zero points for no similarity, one point for loose similarity, and two points for strong similarity. All participants had to document this process and rate all proposals based on this scale. It was up to the users to decide how much weight the system should assign to different metrics. This process was carried out at a total of 20 starting points.

Table 1: Result scores of the expert study with average relevance scores (0-2) of proposed publications for 1-3 base nodes.

	Number of Base Nodes		
	1	2	3
Similar Background	1.16	1.14	0.98
Similar Method	0.79	0.84	0.76
Similar Objective	0.64	0.57	0.49
Complement Method	0.85	0.88	0.84
Complement Objective	0.52	0.47	0.49
Complement Results	0.52	0.49	0.51

As shown in Table 1, recommended publications are considered most similar to base nodes in terms of background, making the system primarily useful for topic-oriented recommendations and secondarily problem- and approach-oriented in the context of this dataset and user study. For a meaningful, quantitative evaluation of how the quality of recommended publications relates to the number of existing base nodes, the amount of data recorded and the number of participants were too low.

Participants were also asked for their subjective assessment of metrics that produced the most promising suggestions and what general aspects of the system were noticeable. Cosine similarity based on embeddings was unanimously mentioned as the most useful metric. This assessment is also consistent with the automated evaluation described above. Furthermore, for publications with a special field of application or technology, significantly more relevant papers are recommended than for general scientific works.

5 DISCUSSION

As both the objective evaluation and qualitative user study suggest, efficient literature research can be carried out using the presented dataset and metrics, but allows only an explorative and no aspect-based way of operation and thus cannot be evaluated in a strictly objective manner. However, the system could be extended to support aspects such as from (Ostendorff et al., 2020), include them as comparison parameters, and also assign weights to them. This would also allow for finer-grained control of the influence of these classifications. Furthermore, algorithms for the calculation of similarity have a runtime complexity of $O(n^2)$ (pairwise comparison of each publication), which makes it highly computationally expensive to cover publication data beyond the dataset. There are several approaches that can reduce the number of necessary pairwise comparisons to continuously extend the dataset at low costs in terms of processing. Kanakia et al. (Kanakia et al., 2019) describe a clustering approach based on k-means, which could also be incorporated into our system to eliminate the need for pairwise comparisons. In addition, the results of the expert user study conducted cannot be considered representative, since, firstly, too little data could be collected, and secondly, questions deliberately allowed for subjective views, which significantly complicates an evaluation. While conducting the study, it also became apparent that identical recommendations are rated differently by individual persons. A more useful and detailed evaluation of the system requires a significantly more complex evaluation scheme and rating scale, as well as a substantially larger group of participants with expert knowledge (active researchers) to directly gather feedback from the intended user audience.

Data from a preliminary user study offered some early insight for future system improvements. Participants, who were not extensively briefed on the interface, used various features beyond task requirements. Features like colorization, tags, and comments, while useful for workspace organization, were not essential for the task, but indicate user engagement. This is particularly noteworthy for future collaborative functionalities, an aspect not covered in this study. Participants also sought more customization options, highlighting mixed satisfaction levels that can evolve over time with increased complexity of the interface and recommendation quality.

Despite positive initial impressions, the study did not yield detailed feedback on recommendation experience. Future research should allow users to rate suggestions directly and offer evaluation criteria, as in Section 4. Overall, the study confirmed the benefit of a visual interface for organizing and exploring publications, but suggests that a more extensive study is needed for a full understanding of the user experience.

6 CONCLUSION AND FUTURE WORK

In this work, we presented a content-based, spatial recommender system for scientific papers using a dataset consisting of pairwise comparisons based on metadata and relationships extracted with help of NLP techniques. The proposed system has shown promising results in both objective evaluation and user study, but to fully leverage its capabilities, both the mechanisms for creating the dataset and the interface itself can be further developed in a variety of ways. The current system enables users to assign custom weights to different similarity measures. However, it may be beneficial to incorporate an algorithm, which includes a personalized bias in identifying similar publications, based on the weight settings for previously recommended publications, which were added to the workspace. An algorithm of this kind is a way to personalize the system, which recognizes whether users prefer to receive recommendations for publications based on matching keywords and phrases or contextually similar work calculated by cosine similarity of text embeddings. One possibility would be to enhance the capabilities of the spatial parsers to infer the current needs of the user, regarding the customization. A limitation of the present dataset lies in the computationally intensive generation. As a countermeasure, clustering and subsequent dimensionality reduction techniques are investigated to reduce the number of comparisons to publications within the same cluster and to reduce the feature space in general before computing cosine similarities between text embeddings of abstracts. This would improve the efficiency of our system and allow for ongoing dataset expansion to dramatically increase the practicality. Another restriction of the current system is the need for users to input a specific publication as a starting point for recommendations. To address this limitation, we propose exploring approaches for suggesting base publication nodes based on natural language queries from users. This could make our system more user-friendly and create easier conditions for discovering new research areas.

In summary, our research has laid a solid basis for the design of a spatial recommender system for scientific publications. We are confident that the future work outlined above has the potential to significantly improve our current system and make it even more useful to researchers and practitioners in the scientific community.

REFERENCES

(2008–2023). Grobid. https://github.com/kermitt2/grobid.

Amati, G. (2009). *BM25*, pages 257–260. Springer US, Boston, MA.

Atzenbeck, C., Herder, E., and Roßner, D. (2023). Breaking the routine: spatial hypertext concepts for active decision making in recommender systems. *New Review of Hypermedia and Multimedia*, pages 1–35.

Atzenbeck, C., Roßner, D., and Tzagarakis, M. (2018). Mother - An integrated approach to hypertext domains. In *HT 2018 - Proceedings of the 29th ACM Conference on Hypertext and Social Media*, pages 145–149, New York, New York, USA. ACM Press.

Bai, X., Wang, M., Lee, I., Yang, Z., Kong, X., and Xia, F. (2019). Scientific paper recommendation: A survey. *IEEE Access*, 7:9324–9339.

Beckmann, C. and Gross, T. (2010). Towards a group recommender process model for ad-hoc groups and on-demand recommendations. In *Proceedings of the 16th ACM international conference on Supporting group work - GROUP '10*, page 329, New York, New York, USA. ACM Press.

Bitton, E. (2009). A spatial model for collaborative filtering of comments in an online discussion forum. *RecSys'09 - Proceedings of the 3rd ACM Conference on Recommender Systems*, pages 393–396.

Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision*, 11(5):4–4.

Carr, L. A., DeRoure, D. C., Hall, W., and Hill, G. J. (1995). The Distributed Link Service: A Tool for Publishers, Authors and Readers. In *Proceedings of the Fourth International World Wide Web Conference*, pages 647–656, Boston, MA.

Chalmers, M. and Chitson, P. (1992). Bead: Explorations in Information Visualization. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '92, pages 330–337, New York, NY, USA. ACM.

Cohan, A., Feldman, S., Beltagy, I., Downey, D., and Weld, D. S. (2020). SPECTER: Document-level Representation Learning using Citation-informed Transformers. In *ACL*.

Collins, A. and Beel, J. (2019). Document embeddings vs. keyphrases vs. terms for recommender systems: A large-scale online evaluation. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 2019-June:130–133.

Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python.

Kanakia, A., Shen, Z., Eide, D., and Wang, K. (2019). A scalable hybrid research paper recommender system for microsoft academic. In *The World Wide Web Conference*. ACM.

Kreutz, C. K. and Schenkel, R. (2022). Scientific paper recommendation systems: a literature review of recent publications. *International Journal on Digital Libraries 2022 23:4*, 23(4):335–369.

Ma, X., Wang, Z., Ng, P., Nallapati, R., and Xiang, B. (2019). Universal text representation from BERT: an empirical study. *CoRR*, abs/1910.07973.

Ostendorff, M., Ruas, T., Blume, T., Gipp, B., and Rehm, G. (2020). Aspect-based document similarity for research papers. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6194–6206, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Parra, D. (2012). Beyond lists: Studying the effect of different recommendation visualizations. *RecSys'12 - Proceedings of the 6th ACM Conference on Recommender Systems*, pages 333–336.

Reinert, O., Bucka-Lassen, D., Pedersen, C. A., and Nürnberg, P. J. (1999). CAOS: A Collaborative and Open Spatial Structure Service Component with Incremental Spatial Parsing. In *Proceedings of the Tenth ACM Conference on Hypertext and Hypermedia : Returning to Our Diverse Roots: Returning to Our Diverse Roots*, HYPERTEXT '99, pages 49–50, New York, NY, USA. ACM.

Renuka, S., Raj Kiran, G. S. S., and Rohit, P. (2021). An Unsupervised Content-Based Article Recommendation System Using Natural Language Processing. In Jeena Jacob, I., Kolandapalayam Shanmugam, S., Piramuthu, S., and Falkowski-Gilski, P., editors, *Data Intelligence and Cognitive Informatics*, pages 165–180, Singapore. Springer Singapore.

Rose, S., Engel, D., Cramer, N., and Cowley, W. (2010). *Automatic Keyword Extraction from Individual Documents*, pages 1 – 20.

Roßner, D., Atzenbeck, C., and Gross, T. (2019). Visualization of the relevance: Using physics simulations for encoding context. In *HT 2019 - Proceedings of the 30th ACM Conference on Hypertext and Social Media*, pages 67–76, New York, New York, USA. ACM Press.

Schedel, T. (2017). Spatio-Temporal parsing in spatial hypermedia. *ACM SIGWEB Newsletter*, (Winter):1–5.

Shipman III, F. M., Marshall, C. C., and Moran, T. P. (1995). Finding and Using Implicit Structure in Human-organized Spatial Layouts of Information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 346–353, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Singh, A., D'Arcy, M., Cohan, A., Downey, D., and Feldman, S. (2022). Scirepeval: A multi-format benchmark for scientific document representations. *ArXiv*, abs/2211.13308.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2019). Bertscore: Evaluating text generation with BERT. *CoRR*, abs/1904.09675.