Exploring Neuroscience Literature and Understanding Relations Between Brain-Related Topics - Using Augmented Reality

Ghazaleh Tanhaei
Department of Information and Computing Sciences, Utrecht University
Utrecht, The Netherlands
g.tanhaei@uu.nl

ABSTRACT

Neuroscience researchers are interested in understanding relation between anatomical regions of the brain and disorders that affect them, for example. Using the topics themselves, rather than individual articles, to examine relation in a large body of literature, can provide a higher-level approach. I investigate the use of 3D representations in Augmented Reality to aid neuroscientists to explore literature and understand relations between brain-related topics, given the three-dimensional nature of the brain. Distant reading refers to comprehending the results of studies of a large number of articles, as opposed to the more common "close reading" of individual publications. For distant reading of neuroscience literature, I identify visualization and interaction design requirements. My assumption is that by providing overviews of the correlations among topics through the use of literature, these will allow neuroscientists to better understand the gaps in the literature and more quickly identify the most suitable experiments to carry out. The DatAR team at Utrecht University has created a prototype 3D AR implementation using which I have carried out two studies of a literature exploration interface. These studies showed that visualizing topics and their relation in an immersive AR environment is clear, understandable and helpful for exploring neuroscience literature. In the following, I will carry out a study that participants can make parallel query and compare the results. I will further investigate in finding indirect relations between brain regions and brain diseases. The last study will support neuroscience students to understand course material. Interface improvements are considering where necessary.

CCS CONCEPTS

• **Human-centered computing → Visualization systems and tools**; *Immersive Analytics*; **Interactive systems and tools**; *Augmented Reality*; • **Information systems → Users and interactive retrieval**; **Information retrieval query processing**.

KEYWORDS

neuroscience literature search, exploratory search, information visualization and visual analytics, measures of user experience and performance, augmented reality, virtual reality

CHIIR '22, March 14–18, 2022, Regensburg, Germany

ACM ISBN 978-1-4503-9186-3/22/03.
https://doi.org/10.1145/3498366.3505806

ACM Reference Format:
Ghazaleh Tanhaei. 2022. Exploring Neuroscience Literature and Understanding Relations Between Brain-Related Topics - Using Augmented Reality. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22), March 14–18, 2022, Regensburg, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3498366.3505806

1 MOTIVATION AND RELATED WORK

Literature exploration is an important part of any research, including in subjects like neuroscience. Personal communication with Cunqing Huangfu[1], neuroscientists need to understand relations between brain-related topics in order to identify promising areas for further research. The typical method of exploring literature is to look for individual papers. However, allowing researchers to search for literature by topic, such as anxiety as a brain disease or the Amygdala as a brain region, can save time and effort. Swanson has produced a number of medical breakthroughs since 1986 by examining novel relations from a bibliographic database [Ganiz et al. 2006]. Exploring literature is applicable to various scientific disciplines, not just biomedical domains [Crichton et al. 2020; Henry and McInnes 2017; Swanson 2008]. In biomedical research, literature exploration discovers new relations between genes and disorders, relationships between diseases, predict drug reactions, and locate new study areas [Banerjee et al. 2014; Zhang et al. 2014].

One of the major limitations of traditional literature exploration is the inability to carry out intricate relation exploration. Identifying these relations is time-consuming and not all of them are fascinating, whereas a neuroscientist's primary purpose is to find high-potential relations to experiment. Topic-based exploration, rather than article-based exploration, helps scholars to pick and uncover relations between themes from enormous amounts of literature more quickly. We build a prototype interactive environment including neuroscience literature to learn more about how topic-based literature discovery might benefit neuroscience researchers. We use the Linked Brain Data (LBD)[2] repository as an example of a topic-based analysis of neuroscience literature. LBD is a repository of tens of thousands of neuroscience papers that was created to aid in the investigation of relations between topics, including those found in PubMed. LBD comprises almost 6,000 connections between around 300 brain diseases and 500 brain regions. When the size and complexity of the available literature increases, analysis complexity increases accordingly and it is more likely that essential relations may be overlooked [Görg et al. 2010]. Through (semi-)automated analysis employing machine-learning approaches, literature exploration aids in the finding of hidden and unknown relations.

[1] A neuroscientist at the Institute of Automation of the Chinese Academy of Sciences group
[2] http://www.linked-brain-data.org/index.jsp?link=link

However, extracting topics from literature by neuroscientists requires technical query-making expertise. Exploring literature and extracting topics is only part of the problem. It is also important to visualize this information and the relations. Visualizing Linked Data (LD) can aid non-technical users in comprehending the meaning of the content [Po et al. 2020]. LD graphs created from data files or SPARQL endpoints can be presented interactively using a variety of tools. Authors show topics as nodes in a graph, including zooming in and out, as well as filtering and altering nodes and edges [Bellini et al. 2014; Graziosi et al. 2018; Micsik et al. 2014; Pietriga 2002].

Immersive analytics (IA), particularly Augmented Reality, provides a platform for binocularly studying 3D structures, such as brain regions. Given the importance of the brain in comprehending neuroscience, I hypothesis that offering an interactive 3D visualization that is incorporated into a researcher's current "2D workflow" will aid knowledge tasks. Users can identify possible relations in the literature in the context of the structures they are studying by using a 3D IA environment. I contribute design requirements and recommendations for the development of immersive analytics tools to support neuroscientists in performing topic-based literature exploration.

2 DATAR TERMS

During implementing the DAtAR prototype, we established some fundamental terms.

Literature Topics: The topics derived from the literature analysis were gathered from neurology publications from online resources such as PubMed and are stored as the LBD knowledge graph. LBD contains a set of topics identified and classified within the publication's title and sentences of the abstract. Brain diseases, brain regions, cognitive functions, and neurons are some of the topics in the LBD. For these studies, we investigate how participants are able to explore connections between brain regions and brain diseases.

Topic-Model: Colleagues at the Institute of Automation of the Chinese Academy of Sciences group applied the LDA topic model method and t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm in the abstract and title of neuroscientific publication, yielding 3D locations for around 300 diseases. To preserve the structure of low-dimensional data such as high-dimensional, I prefer 3D dimensions rather than 2D, because the distance between diseases gives an indication of their semantic similarity; the closer they are, the more semantically similar.

Neighboring Diseases: Brain diseases that are semantically similar may affect similar brain regions. Brain diseases that are unrelated to a brain region but they are close to the related one are "Neighboring diseases" and can be promising areas for investigation.

Number of Co-Occurrences: The number of co-occurrences of a topic pair indicates the number of times that a topic, such as a specific brain disease, appears in the same sentence as another topic, such as a brain region.

3D Brain Model: The 3D brain model is based on 3D region positions provided by the SBA. The tool highlights regions in the 3D brain model based on a user's query. Most regions have multiple names based on different conventions known by a neuroscientist. To ensure that these are correctly highlighted in the tool, we need to create a mapping from the different names to the name displayed. Two neuroscientists helped us perform the required mappings for the 561 regions.

3 WORK TO DATE

3.1 Research Approach, Methods and Rationale

The main goal of this research is to design a tool to find unexplored research areas and aid comprehension of relations between neuroscience topics. I follow a user-centered design approach by involving neuroscience and literature exploration experts at early stages in the design. A neuroscience researcher, Cunqing Huangfu, checked my initial user scenario and helped me determine a real neuroscientist's task: to find diseases related to a brain region (e.g. Amygdala). Two design requirements (DR1 and DR2) came from this neuroscientist's task and were used to introduce my first scenario. I identified DR3 and DR4 based on the scenario.

- DR1: Co-occurrences. Present a wide range of co-occurrences of brain diseases with a brain region.
- DR2: Neighboring diseases. Identify unexplored brain diseases that may affect a brain region.
- DR3: Select Topic. Select one or more topics in the topic model and brain visualization.
- DR4: Identify Related Topics.

I verified my design with data discovery experts and identified potential functionalities that an immersive environment could provide. In December, 2019, we organised a brainstorming session with three data discovery experts and three AR/VR developers and presented our designed system. I explained the scenario and asked for potential visualization options; then we demonstrated the system and gathered opinions for improvement. Resulting in the following requirement:

- DR5: Identify Sources. Provide access to the source literature.

Type of visualization of available objects such as brain topics (diseases and regions) or implementing features that help meet the requirements considered as our design rationale.

3.2 Study 1: Filtering Number of papers on Related Brain Diseases to Amygdala

We understand from Cunqing Huangfu that neuroscientists need to be able to explore the relations between brain-related topics before any real experiments and that topic-model based representation may be a useful visual representation on which to base an interactive interface.

- RQ1: How DatAR procedure to find relation between brain related topics support representative of topic-based literature exploration and how the functionality we provide in the tool useful for representative tasks?

Diseases having a small number of, or even zero, co-occurrences with a certain brain region in the literature can be interesting since they may point to undiscovered territory that requires more research. Researchers may want to look into diseases with a large number of co-occurrences because they need to double-check other people's findings. In this case, I provide a small number of co-occurrences to assist users in discovering previously unknown relationships. The challenge of discovering new topics is illustrated by the following scenario: a user is interested in disorders involving the Amygdala. I chose the Amygdala as an example since it is a brain region that is linked to approximately 107 disorders, which is neither very many nor very few.

To answer RQ1, in January 2020, the prototype was demonstrated in individual sessions with eight neuroscience students. Participants could explore the relation between the Amygdala and associated brain diseases. The aim of the session was to allow neuroscientists to test an early prototype tool rather than a polished one. Also, I explained my system's core functionalities and asked whether they were useful.

Participants had the opportunity to work with the Topic Model widget and explored semantically similar diseases in a 3D environment. Through interviews with participants, I found that the prototype tool is valuable for exploring literature based on topics since they did not have appropriate tools to find relations between brain regions and diseases.

3.3 Study 2: Relation-Finding and Sentences Exploration

I improved my preliminary user scenario based on the information that I gathered from these neuroscientists. For example, I added the ability that users can run a reverse scenario: asking for related brain regions when a disease is selected. Also, users can check the sentences that indicate the relation between selected region and disease. There was relatively little language analysis on the relation between two topics found in a single sentence throughout the creation of the LBD repository. When two topics appear in the same sentence and no negative words such as "no" or "not" appear in the sentence, a single instance of a co-occurrence is counted. There is no other way to verify if the sentence's meaning is correct. "The relation between region A and disease B has been questioned." or "The relation between region A and disease B has been fabricated." are two sentences in the dataset that show that region A and disease B are linked. As a result, users must be able to retrieve phrases from the original document, which are represented by the co-occurrences matching to a query, in order to assess the positive or negative contingency. As a result, users must be able to access the sentences in the original text, which are represented by the co-occurrences matching to a query, in order to evaluate the positive or negative contribution to the relationship.

- RQ2: How do DatAR visualization of semantically similar brain diseases support neuroscientists to understand and verified the relation between brain regions and brain diseases?
- RQ3: Does AR visualization help users to find unknown relations between diseases and Amygdala?

To answer RQ2 and RQ3, in June 2020, I evaluated the prototype by involving literature search experts. I asked about the user experience, their understanding of the prototype's functionality, their current work practices and their observations on the visualization. The result of the qualitative evaluation show that DatAR 3D visualization helped participants to understand relation between topics. It also navigated them through the work with the tool.

4 EXPECTED NEXT STEPS AND LONG TERM GOALS

4.1 Study 3: Providing Support for Investigating Brain Diseases Related to the Same Brain Regions

After running two studies, we understand that sometimes neuroscientists need to know how two diseases are discussed in the literature and how those impact the same regions. They need to see the results at the same time to compare the two findings. According to a study [Goodkind et al. 2015], many distinct mental-health illnesses, ranging from schizophrenia to depression, are caused by the same brain regions. They investigated whether common psychiatric disorders have a common structure in the brain. Poldrack, et al. [Poldrack et al. 2012] showed that how exploring topics in the literature and mapping them to the brain models can support researchers to discover new insights about the roles of particular brain systems in mental function. In our study, a neuroscientist can conclude that which of these two diseases has been discussed more often and whether both diseases affect the same parts of the brain or not. We investigate whether DatAR can be useful for discovering the conceptual structure of disorders and their mapping to brain systems by searching the literature for brain-related topics.

- RQ4: How can DatAR support users in parallel investigations of multiple queries and result comparison to discover whether some diseases affect a common regions in the brain?

For this study, we do not need any new implementation in DatAR. The 3D presentation of the brain and the possibility of comparing two models at the same time support neuroscientists to access more information. Participants of this study should be selected on their background in neuroscience or a different related biology field, since doing the task and evaluation need related background. In the task, we ask participants to select at least two diseases that they think those diseases influence same regions. Filters can help participants to limit the result and conclude their findings. Qualitative evaluation will be in terms of "Usability" and "Data interpretation". Usability questions target at obtaining information on the DatAR's HCI characteristics. The data interpretation questions will be designed to determine whether DatAR visualization is successful in supporting neuroscientists in discovering common regions of selected diseases. It is also vital to measure how end-users perceived the **usability** of DatAR on which they were working by the System Usability Scale (SUS) method.

4.2 Study 4: The Hidden Links in Relation-Finding

DatAR is designed to explore literature to find direct relations between brain regions and brain diseases, but sometimes there is not a direct link between a brain region and a brain disease. It is worth checking whether there are any other brain related topics in between that have relation with that brain region and disease (this procedure is followed in PubMed by neuroscientists of the first study). Brain diseases are related to cognitive function, cognitive function is related to neurons, neurons have neurotransmitters and receptors working in them. Among these relations, direct relation is not always available, usually different kinds of topics are interconnected. The lack of a direct relation cannot be a reason for the lack of a real relation. A neuroscientific researcher can initiate a new practical research to see whether the finding of an indirect relation is valuable. This study will answer RQ5.

- RQ5: How finding indirect relations for the brain regions and brain diseases without direct ones can support neuroscientists to initiate the new area for research?

This study needs some new features to implement in DatAR. For evaluation, I will follow the evaluation method of study 3.

4.3 Study 5: Support Educational Purposes

Technology is now ingrained in education, and the results show that it has a favorable impact on learning and teaching methods [Dübel et al. 2014]. AR/VR technology can provide simulated experiences that can replace some components of hands-on instruction, either temporarily or permanently [Bakharia et al. 2016]. AR provides an efficient way to represent a model that needs visualization, so a brain model can benefit the most from AR. Since we consider a 3D brain model in DatAR, a neuroscientist at UvA, Dr Harm Krugers suggested to consider a situation that neuroanatomy course students can use DatAR to learn some basic research techniques to explore brain structure. This study will answer RQ6, evaluate DatAR in supporting neuroscientific students in learning the neuroanatomy course material.

- RQ6: How DatAR can be used as an educational tool in neuroanatomy courses and provide students with a basic understanding of the structural organisation of the human central nervous system?

In this study, I will provide a task that students will follow to understand course material. Then I will ask about their experience in using AR device and its effect on understanding the material.

5 CONCLUSION

The DatAR tool provides a visual representation of the relations between brain diseases and brain regions. My design's purpose is to assist neuroscientists in searching the literature for the most appropriate experiments to do. Expert neuroscientists should be able to build an understanding through examination of the tool's representations of topics in the literature using our prototype tool. DatAR has the potential to be used as part of a real research project, according to feedback from neuroscientists on the first study. I'm about to begin a procedure. I've demonstrated that AR is an excellent fit for topic-based literature study, and I intend to continue to improve the environment's usability.

ACKNOWLEDGMENTS

I would like to thank my supervisors, Prof. Dr Lynda Hardman and Dr. Wolfgang Hürst for all of their support so far. Also, thanks to Ivar Troost for his cooperation with me in implementing this project and Cunqing Huangfu in Research Center for Brain-inspired Intelligence at the Institute of Automation, Chinese Academy of Sciences, for assisting us in surveying the problem domain. Tessa Heeroma and Anna van Harmelen helped to link various external data sources. This research is supported by a scholarship from the Iranian Ministry of Science, Research and Technology.

REFERENCES

Aneesha Bakharia, Peter Bruza, Jim Watters, Bhuva Narayan, and Laurianne Sitbon. 2016. Interactive Topic Modeling for Aiding Qualitative Content Analysis *(CHIIR '16)*. Association for Computing Machinery, New York, NY, USA, 213–222. https://doi.org/10.1145/2854946.2854960

Ritwik Banerjee, Yejin Choi, Gaurav Piyush, Ameya Naik, and I. V. Ramakrishnan. 2014. Automated Suggestion of Tests for Identifying Likelihood of Adverse Drug Events. In *Proceedings of the 2014 IEEE International Conference on Healthcare Informatics*. IEEE Computer Society, USA, 170–175. https://doi.org/10.5555/2761731.2762011

Pierfrancesco Bellini, Paolo Nesi, and Alessandro Venturi. 2014. Linked Open Graph. *Journal of Visual Languages and Computing* 25, 6 (December 2014), 703–716. https://doi.org/10.1016/j.jvlc.2014.10.003

Gamal Crichton, Simon Baker, Yufan Gue, and Anna Korhone. 2020. Neural networks for open and closed Literature-based Discovery. *Journal of PLoS ONE* (May 2020). https://doi.org/10.1371/journal.pone.0232891

Steve Dübel, Martin Röhlig, H. Schumann, and Matthias Trapp. 2014. 2D and 3D Presentation of Spatial Data: A Systematic Review. IEEE VIS International Workshop on 3DVis. https://doi.org/10.13140/2.1.1281.7602

Murat C. Ganiz, William M. Pottenger, and Christopher D. Janneck. 2006. Recent Advances in Literature Based Discovery. *Journal of the American Society for Information Science and Technology, JASIST* (2006).

Madeleine Goodkind, Simon B. Eickhoff, Desmond J. Oathes, Ying Jiang, Andrew Chang, Laura B. Jones-Hagata, Brissa N. Ortega, Yevgeniya V. Zaiko, Erika L. Roach, Mayuresh S. Korgaonkar, Stuart M. Grieve, Isaac Galatzer-Levy, Peter T. Fox, and Amit Etkin. 2015. Identification of a Common Neurobiological Substrate for Mental Illness. *JAMA Psychiatry* 72, 4 (04 2015), 305–315. https://doi.org/10.1001/jamapsychiatry.2014.2206 arXiv:https://jamanetwork.com/journals/jamapsychiatry/articlepdf/2108651/yoi140096.pdf

Carsten Görg, Hannah Tipney, Karin Verspoor, William A. Baumgartner, K. Bretonnel Cohen, John Stasko, editor="Setchi Rossitza Hunter, Lawrence E.", Ivan Jordanov, Robert J. Howlett, and Lakhmi C. Jain. 2010. Visualization and Language Processing for Supporting Analysis across the Biomedical Literature. In *Knowledge-Based and Intelligent Information and Engineering Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 420–429. https://doi.org/10.1007/978-3-642-15384-6_45

Alice Graziosi, Angelo Di Iorio, Francesco Poggi, Silvio Peroni, and Luca Bonini. 2018. Customising LOD Views: A Declarative Approach. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. Association for Computing Machinery, New York, NY, USA, 2185–2192. https://doi.org/10.1145/3167132.3167367

S Henry and B T McInnes. 2017. Literature based discovery: models, methods, and trends. 74 (October 2017), 20–32. https://doi.org/10.1016/j.jbi.2017.08.011

András Micsik, Zoltán Tóth, and Sándor Turbucz. 2014. LODmilla: Shared Visualization of Linked Open Data. In *Theory and Practice of Digital Libraries – TPDL 2013 Selected Workshops*. Springer International Publishing, Heidelberg, 89–100. https://doi.org/10.1007/978-3-319-08425-1_9

Emmanuel Pietriga. 2002. IsaViz: a Visual Environment for Browsing and Authoring RDF Models. In *WWW 2002, the 11th World Wide Web Conference*. World Wide Web Consortium.

Laura Po, Nikos Bikakis, Federico Desimoni, and George Papastefanatos. 2020. Linked data visualization: techniques, tools, and big data. *Synthesis Lectures on Semantic Web: Theory and Technology* 10, 1 (2020), 1–157. https://doi.org/10.2200/S00967ED1V01Y201911WBE019

RA Poldrack, JA Mumford, T Schonberg, D Kalar, B Barman, and T Yarkoni. 2012. Discovering Relations Between Mind, Brain, and Mental Disorders Using Topic Mapping. *PLoS Comput Biol* 10, 8 (2012). https://doi.org/10.1371/journal.pcbi.1002707

D R Swanson. 2008. *Literature-based discovery? The very idea* (1st ed.). Springer, Berlin. https://doi.org/10.1007/978-3-540-68690-3_1

R Zhang, M J Cairelli, M Fiszman, H Kilicoglu, T C Rindflesch, S V Pakhomov, and G B Melton. 2014. Exploiting Literature-derived knowledge and semantics to identify potential prostate cancer drugs. In *Cancer Informatics*, Vol. 13. 103–111. https://doi.org/10.4137/CIN.S13889