On Dimensions of Plausibility for Narrative Information Access to Digital Libraries

Hermann Kroll[1][0000−0001−9887−9276], Niklas Mainzer[1], and Wolf-Tilo Balke[1][0000−0002−5443−1215]

Institute for Information Systems, TU Braunschweig, Braunschweig, Germany
{kroll,balke}@ifis.cs.tu-bs.de and n.mainzer@tu-bs.de

Abstract. Designing keyword-based access paths is a common practice in digital libraries. They are easy to use and accepted by users and come with moderate costs for content providers. However, users usually have to break down the search into pieces if they search for stories of interest that are more complex than searching for a few keywords. After searching for every piece one by one, information must then be reassembled manually. In previous work we recommended narrative information access, i.e., users can precisely state their information needs as graph patterns called narratives. Then a system takes a narrative and searches for evidence for each of its parts. If the whole query, i.e., every part, can be bound against data, the narrative is considered plausible and, thus, the query is answered. But is it as easy as that? In this work we perform case studies to analyze the process of making a given narrative plausible. Therefore, we summarize conceptual problems and challenges to face. Moreover, we contribute a set of dimensions that must be considered when realizing narrative information access in digital libraries.

Keywords: Narrative Information Access · Plausibility · Dimensions

1 Introduction

Digital libraries are large-scale repositories containing a plethora of heterogeneous but well-curated data: texts, images, videos, research data, and many more. Today, keyword-based search (on texts or respective metadata) forms the primary access path for collections: It is easy to use and thus a commonly accepted interface, and implementing a respective search index comes with acceptable costs for content providers. In addition, keyword queries can be evaluated against different data representations; See keyword-based searches in literature and books, in knowledge bases [10, 15], and data set repositories [8]. Although the advantage of keyword-based search is apparent, moving towards alternative access paths is heavily investigated, e.g., the biomedical database SemMedDB [21] on top of Medline, the Open Knowledge Research Graph [18], or the Europeana as Linked Open Data [16] and DBpedia [2]. A good example is the current trend towards structured knowledge graphs: Here users can formulate their information needs as structured queries (e.g., via SPARQL interfaces) asking for a combination of knowledge particles (i.e., automatically joining pieces of information).

However, integrating different and heterogeneous sources into a single knowledge representation comes with new problems, such as the validity of fused information [22], identifying and canonicalizing pieces of information (entity linkage), and practical extraction problems when mining knowledge from different sources; See [33] for a good overview. A remedy might be the use of *narrative information access* to bypass the necessary integration of different sources. In [23], **narratives** are defined by directed edge-labeled graph patterns involving entities (things and concepts), events (state or state changes), literals (values), and their respective relationships (properties, temporal, and causal predicates, etc.).

Since the individual information particles within a narrative form proper subgraphs, narratives can be understood as forming a logical overlay on top of different knowledge repositories. A narrative query is then successfully answered if evidence for every single relationship of the narrative can be found in some collection's data. This evidence forms so-called **narrative bindings** connecting each relationship against a concrete piece of information stated in some knowledge repository, i.e., any form of data storage (e.g., relational databases, knowledge graphs, data sets, etc.). In previous works we formulated a conceptual model [23] for narrative information access, discussed the binding process from a technical perspective [24], showed how queries can be processed over collections of scientific publications [25], and discussed conceptual problems like context compatibility when answering queries over independent knowledge bases [26].

In contrast to fact-checking, where the task refers to deciding whether some statement is indeed a fact (and thus true) or is rejected because it has no evidence [32, 14], narrative structures may be composed out of several statements. Moreover, we do not require a universally valid decision, i.e., based on a given ground truth (a theory in physics for example), we can make a narrative plausible, whereas we cannot make the narrative plausible on a different ground truth. While question answering and its methods [19, 20, 35, 35] are related to narrative information access, we see significant differences in the evaluation: Narrative information access considers contexts [26, 22], i.e., we do not fuse information from different contexts. For a good example, consider claims from the biomedical domain: Combining claims made when treating mice with claims when treating humans can be a serious threat to the overall validity because no one guarantees that these claims should also be valid for humans then.

In summary, the key advantage of narrative information access is to make user-formulated graph patterns *plausible* in valid contexts. But beyond apparent technical problems and the problem of context compatibility in narrative information access, the notion of *plausibility* has yet to be discussed in detail. Our current model considers a narrative as plausible if evidence for every of its claims can be found. But is it as easy as that? In this paper we summarize the conceptual problems of our narrative query model. Therefore, we perform case studies in different domains to find the basic building blocks for an actionable notion of plausibility. Moreover, we open up a design space and discuss our notion of plausibility in detail by arguing on different evaluation semantics, the trustworthiness of sources, and the quality of bindings.
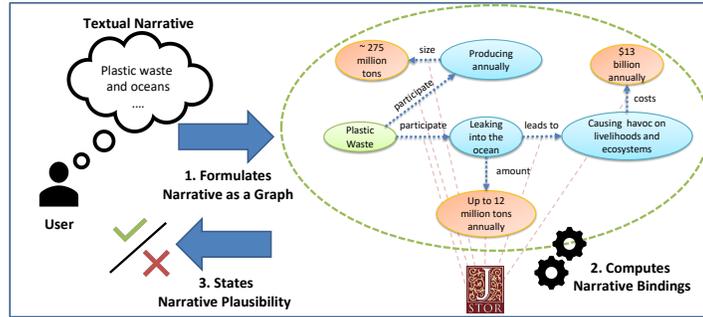
Fig. 1. Making narratives plausible by computing bindings against knowledge repositories. Here our example Narrative 1 is bound against the digital library JSTOR.

2   Case Studying Narrative Plausibility

In brief, a user models a narrative of interest as a graph structure. Then, a process searches for evidence for each relationship in a set of given knowledge repositories. If we can find evidence for the whole narrative, i.e., every narrative's relationship, we consider it plausible [23]. The process is shown in Fig. 1. Note that we may build different plausibility validation processes, which can vary in their included and excluded knowledge repositories for the binding computation, e.g., peer-reviewed scientific papers, fact-checked sources, or arbitrary websites. Due to the lack of existing implementations for such a process, we performed manual case studies, i.e., we searched for evidence through keyword queries in knowledge repositories and evaluated the results manually.

Narrative 1 *Of the approximately 275 million metric tons of plastic waste produced annually, up to 12 million tons leak into oceans, wreaking havoc on livelihoods and ecosystems (CIEL, 2020). The result is an estimated $13 billion in annual environmental damage to marine ecosystems.* [1] *(See Fig. 1).*

The central entity of our example is *plastic waste*, around which the rest of the narrative revolves. The narrative states two events directly relating to plastic waste, its *annual production*, and that *plastic waste is leaking into the ocean*. The *havoc on livelihoods and ecosystems* is a direct consequence of *ocean plastic* and represents the third event of our example. Each event that we described is further connected to a literal: There are approximately *275 million tons* of annual plastic waste production, up to *twelve million tons* leak into the ocean, and the havoc on livelihoods and ecosystems costs around *13 billion dollars per year*. Note that we may extract different graph representations of the same textual narrative but we discuss this transformation ambiguity in the next section.

We next present a plausibility validation process based on the digital library JSTOR (`https://www.jstor.org/`). One of the problems we faced in the beginning was that some relationships require more context from the narrative itself, e.g., *Producing annually an amount of 275 million metric tons* does not include

that it is about *plastic waste*. Either the model must include this detail in the event *Producing annually*, or we must verify that the context of the production must match plastic waste. Therefore, we started by evaluating that *around 275 million metric tons of plastic waste are annually produced.* We formulated different keyword queries to search for the given relationship. We retrieved a source that claimed the 275 million metric ton estimation [17]. We found two more sources, one of which reported an annual plastic production of 381 million metric tons in 2015, of which 79 percent are considered waste [34]. We performed the calculation ourselves, and the result was around 300 million tonnes of plastic waste per annum [34] or respectively 321 million metric tonnes per year [30].

We then evaluated that *up to 12 million tons of plastic waste are annually leaking into the ocean.* [30] reported that in 2010, approximately nine million tons of plastic were added to the ocean. [34] estimated that around 4.8 to 12.7 million tons of plastic entered the ocean in 2010. The plausibility validation seemed to confirm the narrative claim in this way. We moved to the next relationship: *The leaking of plastic into the ocean leads to havoc on livelihoods and ecosystems.* Even in 1987, [3] already reported an endangerment of marine environments caused by plastic in the ocean. A more recent article described dangers to the marine animal and plant life due to microplastic as well as the economic dangers [9]. For the last relation we searched evidence that the impact of plastic waste in oceans costs 13 billion dollars annually. [30] reported 75 billion Dollar annual environmental costs originating from plastics. However, the source did not state how much of the total costs are connected to plastic waste in the ocean. Apart from [30], we could not find any other sources stating the actual costs.

2.1   Summary of Other Case Studies

We performed the plausibility validation for three additional narratives: 1. Deforestation of the Amazonian Rainforest and forests in general, 2. Claims about the connection between government measures during the Covid-19 pandemic and the number of Covid-related deaths, and 3. Global warming and the effects of carbon dioxide emissions on the climate. For 3. we only considered fact-checked articles, whereas we allowed every website on the first result page of our Google Search for the 1. and 2. narrative.

An interesting finding of the deforestation narrative occurred when we evaluated the percentage of the Earth's landmass covered by forests. The narrative stated a value of 30 percent. One of our retrieved sources confirmed this narrative claim by stating a value of 30.7 percent [4]. Another source, however, stated that forests cover 38 percent of the habitable land area [31]. These values greatly differ and could be classified as counter-evidence because 30 percent could not be verified precisely. However, we noticed that [4] base their value on the total land area, while [31] makes this calculation based on the habitable land area, explaining the difference in these values.

We performed a selection of our approaches multiple times over a period of several weeks to observe how the results would change over time. We retrieved the global warming narrative from [28]. When we initially accessed this source,

the narrative reported an average surface temperature of 1.18 degrees Celsius on Earth. However, over time this value has been changed to one degree Celsius, which is less specific than the initial estimate. While performing the initial plausibility validation, we noticed various factors determining Earth's average surface temperature changes, such as the data collection methods and the point in time used as a reference. Various sources described warming of around one degree, but we could not precisely confirm the 1.18 degrees Celsius. The fact that NASA changed that value confirms our assumption that there is too much uncertainty about the data to state such a specific temperature change.

3   Concluding Discussion on Narrative Plausibility

In the following we summarize the main findings from our previous case studies.

Confirmation Bias. Confirmation bias describes "the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis at hand" [29]. As this definition suggests, bias can already occur in the search for information, and its effects highly depend on how the plausibility validation process searches for evidence. For instance, evaluating *plastic waste causes 13 billion dollars annually in environmental costs by leaking into the ocean* could lead to other results than searching without a concrete cost value.

Narrative's Context. The narrative itself spans a context that must be considered when binding its relationships individually. For instance, searching for *an amount of leakage* without *plastic waste* may result invalid results. Similarly, evaluating the *average surface temperature* without a concrete year may also yield results that do not match the user's indented context.

Finding the Narrative's Source. Imagine a narrative extracted from a study reports scientifically disputed findings. We would assume such a narrative to be evaluated as implausible in a plausibility validation process. However, the process will likely find the study's sources again – even if they are disputed.

Ambiguity of Narrative Extraction. Narratives can be interpreted differently since they make various claims, usually without explaining exactly how everything is meant. Without an objective understanding of what the narrative is exactly stating, this will affect the narrative model extracted from the actual textual narrative. Consequently, the plausibility validation process computes narrative bindings for a selected instance of the possible interpretations, making the conclusion about narrative plausibility dependent on the chosen narrative model and, therefore, on the chosen interpretation of the narrative. A plausibility validation process based on another interpretation of the same narrative might result in a different conclusion about the narrative's plausibility.

Evidence. We observed three issues when searching for evidence: 1. Indirect Evidence: The process may not find direct but rather indirect evidence (e.g., a certain percentage of the annual plastic production), which requires further computations. 2. Counter-Evidence: We may find counter-evidence, i.e., evidence that contradicts a narrative's relationship. 3. Absence of Evidence: Not finding evidence is always connected to uncertainty since we cannot be sure if evidence

does not exist or if we simply did not find it. All issues are especially relevant if we utilize error-prone NLP and retrieval methods [12, 6].

Trustworthiness of Sources. When computing bindings against different types of sources, the credibility of sources must be considered. On the one hand, we may utilize Provenance information that has already been curated [27, 13, 7]. Provenance allows knowledge curators to precisely describe how a certain piece of information was created or obtained. If we bind against that certain piece, we can then utilize its Provenance. On the other hand, we may integrate tests like the CRAAP test [5] or automated heuristics, e.g., scientific publications before websites, highly influential over low cited publications, etc. Note that we are aware of highly influential papers that have been retracted but handling such cases in digital libraries is a research field on its own [11]. But if we do not have curated information available, we must think about alternatives here.

3.1   Dimensions of Narrative Plausibility

The contribution of this work is the introduction of **dimensions** influencing narrative plausibility. When implementing a validation process in a digital library, the designers must decide how the following criteria should be handled.

Narrative Structure. We have seen that transforming a story of interest (a user's query) into a graph representation can be challenging and may yield several different representations. Moreover, each representation might end up in a different result (plausible/not plausible) which could confuse users. Do we consider the exact narrative, or do we allow a few derivations (e.g., relaxing the query a bit)? What can users then expect if a narrative is considered plausible?

Validation Approach. Which knowledge repositories are considered for the binding computation? Which methods are used to compute the actual bindings? What do these methods guarantee? And thus, what could a user then expect?

Types of Evidence. Beyond suitable evidence we may face the *indirect*, *missing*, and *counter-evidence* in practice. What should we do if we cannot find evidence for all of a narrative's relationships? Is the narrative simply considered as *not plausible*? Or should we instead show the user how much of the narrative is plausible and which parts lack evidence? Do we consider counter-evidence? And if we do, what does the existence of counter-evidence then mean, e.g., that the narrative is not plausible? As we have argued before, there is a fine line between evidence not exactly confirming a relationship and counter-evidence.

Confidence of Bindings. How trustworthy do the sources have to be? How should this trustworthiness be determined or measured? How is the quality of the actual bindings from a technical perspective, e.g., the confidence of a retrieval process or similarity measures for values? Should we use thresholds to consider trustworthy sources and high-quality bindings only? Which evaluation strategy do we select, e.g., the trustworthiness of sources over binding quality? Do we average both *values* into an overall confidence?

References

1. Plastics and the Environment. `https://www.genevaenvironmentnetwork.org/resources/updates/plastics-and-the-environment/` (last access: 2022-05)
2. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.G.: Dbpedia: A nucleus for a web of open data. In: Aberer, K., Choi, K., Noy, N.F., Allemang, D., Lee, K., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007. Lecture Notes in Computer Science, vol. 4825, pp. 722–735. Springer (2007). https://doi.org/10.1007/978-3-540-76298-0_52
3. Azzarello, M.Y., Vleet, E.S.V.: Marine birds and plastic pollution. Marine Ecology Progress Series **37**(2/3), 295–303 (1987), `http://www.jstor.org/stable/24824704`
4. Bank, W.: Forest area (% of land area). `https://data.worldbank.org/indicator/AG.LND.FRST.ZS`, online; accessed 25 May 2022
5. Blakeskee, S.: The craap test. LOEX Quarterly **31** (2004), `https://commons.emich.edu/loexquarterly/vol31/iss3/4`
6. Blatz, J., Fitzgerald, E., Foster, G., Gandrabur, S., Goutte, C., Kulesza, A., Sanchis, A., Ueffing, N.: Confidence estimation for machine translation. In: COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics. pp. 315–321. COLING, Geneva, Switzerland (aug 23–aug 27 2004), `https://aclanthology.org/C04-1046`
7. Carroll, J.J., Bizer, C., Hayes, P.J., Stickler, P.: Named graphs. J. Web Semant. **3**(4), 247–267 (2005). https://doi.org/10.1016/j.websem.2005.09.001
8. Chapman, A., Simperl, E., Koesten, L., Konstantinidis, G., Ibáñez, L., Kacprzak, E., Groth, P.: Dataset search: a survey. VLDB J. **29**(1), 251–272 (2020). https://doi.org/10.1007/s00778-019-00564-x
9. Clark, J.R., Cole, M., Lindeque, P., Fileman, E., Blackford, J., Lewis, C., Lenton, T.M., Galloway, T.S.: Marine microplastic debris: a targeted plan for understanding and quantifying interactions with marine life. Frontiers in Ecology and the Environment **14**(6), 317–324 (2016), `http://www.jstor.org/stable/44001167`
10. Elbassuoni, S., Blanco, R.: Keyword search over rdf graphs. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management. p. 237–242. CIKM '11, Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/2063576.2063615
11. Fu, Y., Schneider, J.: Towards knowledge maintenance in scientific digital libraries with the keystone framework. In: Huang, R., Wu, D., Marchionini, G., He, D., Cunningham, S.J., Hansen, P. (eds.) JCDL '20: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, Virtual Event, China, August 1-5, 2020. pp. 217–226. ACM (2020). https://doi.org/10.1145/3383583.3398514
12. Gandrabur, S., Foster, G.F., Lapalme, G.: Confidence estimation for NLP applications. ACM Trans. Speech Lang. Process. **3**(3), 1–29 (2006). https://doi.org/10.1145/1177055.1177057
13. Group, W.W.: Prov-overview. an overview of the prov family of documents. `https://www.w3.org/TR/prov-overview/` (2013)
14. Guo, Z., Schlichtkrull, M., Vlachos, A.: A Survey on Automated Fact-Checking. Transactions of the Assoc. for Comp. Linguistics (TACL) **10** (2022)

15. Han, S., Zou, L., Yu, J.X., Zhao, D.: Keyword search on rdf graphs - a query graph assembly approach. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. p. 227–236. CIKM '17, Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3132847.3132957

16. Haslhofer, B., Isaac, A.: data.europeana.eu: The europeana linked open data pilot. In: Baker, T., Hillmann, D.I., Isaac, A. (eds.) Proceedings of the 2011 International Conference on Dublin Core and Metadata Applications, DC 2011, The Hague, The Netherlands, September 21-23, 2011. pp. 94–104. Dublin Core Metadata Initiative (2011), http://dcpapers.dublincore.org/pubs/article/view/3625

17. Jambeck, J.R., Geyer, R., Wilcox, C., Siegler, T.R., Perryman, M., Andrady, A., Narayan, R., Law, K.L.: Plastic waste inputs from land into the ocean. Science **347**(6223), 768–771 (2015). https://doi.org/10.1126/science.1260352

18. Jaradeh, M.Y., Oelen, A., Farfar, K.E., Prinz, M., D'Souza, J., Kismihók, G., Stocker, M., Auer, S.: Open research knowledge graph: Next generation infrastructure for semantic scholarly knowledge. In: Kejriwal, M., Szekely, P.A., Troncy, R. (eds.) Proceedings of the 10th International Conference on Knowledge Capture, K-CAP 2019, Marina Del Rey, CA, USA, November 19-21, 2019. pp. 243–246. ACM (2019). https://doi.org/10.1145/3360901.3364435

19. Jin, Q., Yuan, Z., Xiong, G., Yu, Q., Ying, H., Tan, C., Chen, M., Huang, S., Liu, X., Yu, S.: Biomedical question answering: A survey of approaches and challenges. ACM Comput. Surv. **55**(2) (jan 2022). https://doi.org/10.1145/3490238

20. Khot, T., Sabharwal, A., Clark, P.: Answering complex questions using open information extraction. In: Barzilay, R., Kan, M. (eds.) Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers. pp. 311–316. Association for Computational Linguistics (2017). https://doi.org/10.18653/v1/P17-2049

21. Kilicoglu, H., Shin, D., Fiszman, M., Rosemblat, G., Rindflesch, T.C.: Semmeddb: a pubmed-scale repository of biomedical semantic predications. Bioinform. **28**(23), 3158–3160 (2012). https://doi.org/10.1093/bioinformatics/bts591

22. Kroll, H., Kalo, J., Nagel, D., Mennicke, S., Balke, W.: Context-compatible information fusion for scientific knowledge graphs. In: Hall, M.M., Mercun, T., Risse, T., Duchateau, F. (eds.) Digital Libraries for Open Knowledge - 24th International Conference on Theory and Practice of Digital Libraries, TPDL 2020, Lyon, France, August 25-27, 2020, Proceedings. Lecture Notes in Computer Science, vol. 12246, pp. 33–47. Springer (2020). https://doi.org/10.1007/978-3-030-54956-5_3

23. Kroll, H., Nagel, D., Balke, W.: Modeling narrative structures in logical overlays on top of knowledge repositories. In: Dobbie, G., Frank, U., Kappel, G., Liddle, S.W., Mayr, H.C. (eds.) Conceptual Modeling - 39th International Conference, ER 2020, Vienna, Austria, November 3-6, 2020, Proceedings. Lecture Notes in Computer Science, vol. 12400, pp. 250–260. Springer (2020). https://doi.org/10.1007/978-3-030-62522-1_18

24. Kroll, H., Nagel, D., Kunz, M., Balke, W.: Demonstrating narrative bindings: Linking discourses to knowledge repositories. In: Campos, R., Jorge, A.M., Jatowt, A., Bhatia, S., Finlayson, M.A. (eds.) Proceedings of Text2Story - Fourth Workshop on Narrative Extraction From Texts held in conjunction with the 43rd European Conference on Information Retrieval (ECIR 2021), Lucca, Italy, April 1, 2021 (online event due to Covid-19 outbreak). CEUR Workshop Proceedings, vol. 2860, pp. 57–63. CEUR-WS.org (2021), http://ceur-ws.org/Vol-2860/paper7.pdf

25. Kroll, H., Pirklbauer, J., Kalo, J., Kunz, M., Ruthmann, J., Balke, W.: Narrative query graphs for entity-interaction-aware document retrieval. In: Ke, H., Lee, C.S.,

Sugiyama, K. (eds.) Towards Open and Trustworthy Digital Societies - 23rd International Conference on Asia-Pacific Digital Libraries, ICADL 2021, Virtual Event, December 1-3, 2021, Proceedings. Lecture Notes in Computer Science, vol. 13133, pp. 80–95. Springer (2021). https://doi.org/10.1007/978-3-030-91669-5_7

26. Kroll, H., Plötzky, F., Pirklbauer, J., Balke, W.: What a publication tells you - benefits of narrative information access in digital libraries. CoRR (Accepted to JCDL2022) **abs/2205.00718** (2022). https://doi.org/10.48550/arXiv.2205.00718

27. Lebo, T., Sahoo, S., McGuinness, D.: PROV-O: The PROV Ontology. `https://www.w3.org/TR/prov-o/` (2013)

28. NASA: Global Temperature — Vital Signs - NASA Climate Change. `https://climate.nasa.gov/evidence/`, online; accessed 25 May 2022

29. Nickerson, R.S.: Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology **2**(2), 175–220 (1998). https://doi.org/10.1037/1089-2680.2.2.175

30. Rhodes, C.J.: Plastic pollution and potential solutions. Science Progress **101**, 207–260 (2018)

31. Ritchie, H., Roser, M.: Forests and deforestation. Our World in Data (2021), https://ourworldindata.org/forests-and-deforestation

32. Uscinski, J.E., Butler, R.W.: The Epistemology of Fact Checking. Critical Review **25**(2) (2013)

33. Weikum, G., Dong, X.L., Razniewski, S., Suchanek, F.M.: Machine knowledge: Creation and curation of comprehensive knowledge bases. Found. Trends Databases **10**(2-4), 108–490 (2021). https://doi.org/10.1561/1900000064

34. Williams, A., Rangel-Buitrago, N.: Marine Litter: Solutions for a Major Environmental Problem. Journal of Coastal Research **35**(3), 648–663 (01 2019). https://doi.org/10.2112/JCOASTRES-D-18-00096.1

35. Zhao, C., Xiong, C., Qian, X., Boyd-Graber, J.L.: Complex factoid question answering with a free-text knowledge graph. In: Huang, Y., King, I., Liu, T., van Steen, M. (eds.) WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020. pp. 1205–1216. ACM / IW3C2 (2020). https://doi.org/10.1145/3366423.3380197