OCTVis: Ontology-Based Comparison of Topic Models

Amon Ge * Hyeju Jang * Giuseppe Carenini * Kendall Ho † Young Ji Lee ‡

*Department of Computer Science, University of British Columbia
†Department of Emergency Medicine, University of British Columbia
‡School of Nursing, University of Pittsburgh

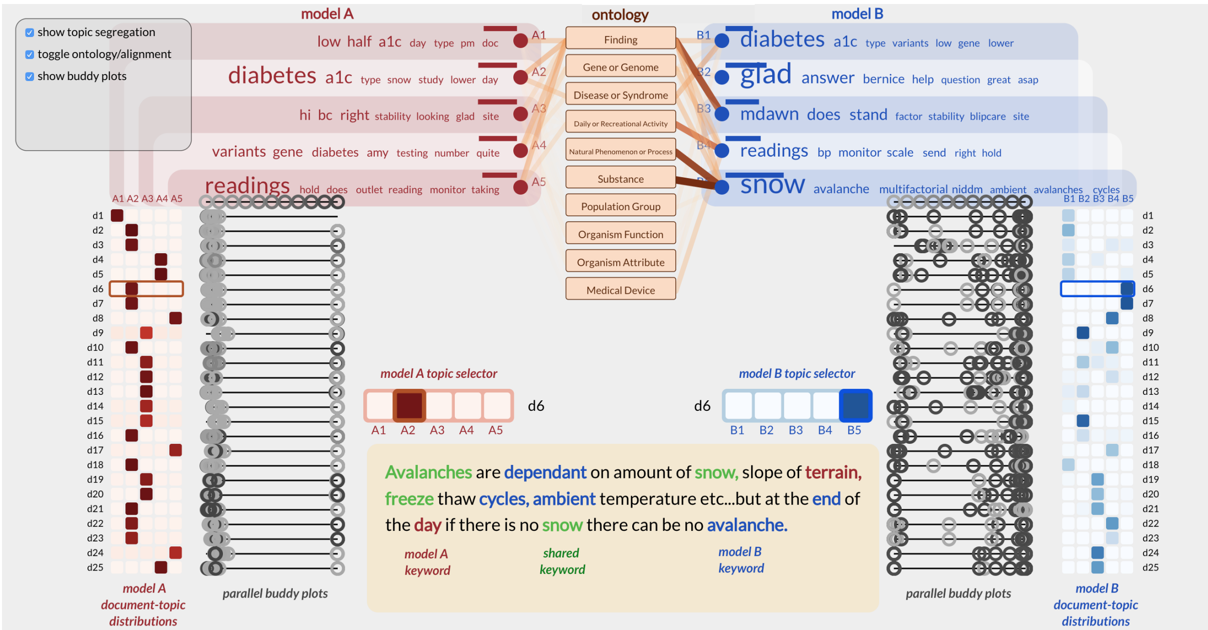

Figure 1: The *OCTVis* interface. Topic models are juxtaposed side by side, allowing comparison of topic alignment to a **domain ontology** (top) and of document-topic distributions (far left and right). The **ontology facet** (top) provides a high level overview of domain-specific ontology concepts, supporting nuanced topic model evaluation (e.g., topic cohesion, segregation) situated within the context of the text domain. A document's keywords (bottom) of a selected pair of topics are highlighted if they belong to the **left model's topic**, **right model's topic**, or are **shared**. Parallel buddy plots (grey) compare relative distances between documents across both models.

Abstract

Evaluating topic modeling results requires communication between domain and NLP experts. *OCTVis* is a visual interface to compare the quality of two topic models when mapped against a domain ontology. Its design is based on detailed data and task models, and was tested in a case study in the healthcare domain.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

Topic modeling is a natural language processing (NLP) technique that identifies topics in a set of texts, or corpus [6]. The resulting topic models can provide informative and concise summaries of the content of large corpora, which can support their exploration and analysis in diverse domains such as health care [7] and education [22]. However, evaluation of topic modeling results is challenging because automatic evaluation metrics, like topic coherence [24], at best weakly correlate with human judgments of quality [8, 9]. Thus, human-in-the-loop evaluation is needed for improving the model, as shown in prior work [1, 9, 14, 21].

This paper presents *OCTVis* (Ontology-based Comparison of Topics), a visual comparison framework for exploring results from topic models, which aims to improve the communication between domain experts and NLP researchers in formative evaluation of topic models. *OCTVis* supports inter-model comparison by incorporating topic alignment visualizations, along with a novel ontology view, which provides domain specific concepts for interpreting topics in a particular domain. For further exploration of the dataset, display of per-document topic distributions and buddy plots allow comparison of topics, texts, and shared keywords at the document level.

A use case study in the healthcare domain provides some initial evidence for *OCTVis* potential. To the best of our knowledge, this is the first user evaluation of an interface for comparing topic models.

*e-mail: {amon, hyejuj, carenini}@cs.ubc.ca
†e-mail: kendall.ho@ubc.ca
‡e-mail: leeyoung@pitt.edu

2 Data and Task Abstractions

By following the standard information visualization methodology [19], we base the design of *OCTVis* on abstracted data and task models. The data model describes text corpora, topic models, and ontologies, while the task model outlines key comparison tasks to enhance qualitative evaluation and understanding of ontology-based topic models.

2.1 Data model

OCTVis is designed to support the comparison of two topic models for the same corpus. The data model includes: the set of documents in the corpus, two topic models, and a domain specific ontology.

Text often consists of many topics; a news article about electric cars could cover battery technologies, government incentives, the environment, and so on. A topic model dataset can be represented as multiple tables. One table stores topics (attributes) per document (row), assigning a quantitative value representing the strength of that topic for that particular document [20]. For probabilistic methods like Latent Dirichlet Allocation (LDA) [4], each value in this table is the specific probability, $P(t|d)$, that the attribute topic t (column) occurs given the document d (row). For each document, we can also assign topics to subdivisions of text, most commonly per word. We have a table for each document where each row represents a word occurring in the document (or corpus) and a quantitative weight for that word associated with each topic (in LDA, this is the probability of a word w occurring given a particular topic t, $P(w|t)$).

Although most topic modeling techniques automatically generate labels for the extracted topics (e.g., most likely words from the topic in LDA), there often exist pre-defined labels for topics and subtopics within particular lexical domains, known as *ontologies*. These ontologies are hierarchical networks of categories and relationships that classify terms used in a particular domain. In our work, we use the Unified Medical Language System (UMLS) Semantic Network [17], a top-level ontology for all UMLS concepts, as a source for our domain-specific medical terminology In general, ontologies exist that cover most domains (e.g., [3, 16, 18]).

Algorithm 1: Pseudo-code for derived data: ontology mapping, topic alignment, an example of a topic metric computation (topic segregation), and document distances.

```
1: procedure MapOntologyConceptsToTopics
2:     for each word w_j in the corpus do
3:         assign w_j to a concept if a mapping exists
4:         (based on an ontology labeler, MetaMap for the UMLS
   in this paper)
5:     for each concept c_i do
6:         for each topic t_k (in both models) do
7:             weight(t_k, c_i) ← 0
8:             for word w_j mapped to c_i do
9:                 weight(t_k, c_i) ← weight(t_k, c_i) + P(w_j|t_k)
10: procedure ComputeTopicAlignment
11:     for each topic t_j in model A do
12:         for each topic t_k in model B do
13:             alignment = sim(t_j, t_k)
14: procedure ComputeTopicSegregation
15:     for each topic t_j in model X do
16:         for each topic t_k in model X do
17:             compute sim(t_j, t_k)
18:         rawSegregation ← mean sim(t_j, t_k) for all other t_k
19:     segregation ← normalized rawSegregation to sum to 1 over
   model X
20: procedure ComputeDocumentDistances
21:     for each document d_i in the corpus do
22:         for each document d_j in the corpus do
23:             distance(d_i, d_j) ← sim(p(t|d_i), p(t|d_j))
```

Algorithm 1 shows the pseudo-code to derive the data model, assuming that the two topic models to be compared have been already generated.[1] This process involves the four steps below.

[1] All the procedures here are specified for an LDA model with conditional probabilities $P(t|d)$ and $P(w|t)$. For non-LDA topic models, corresponding weights connecting topics to documents and words to topics should be used.

[Lines 1-9] Each model is mapped into the ontology. First, [lines 2-4] words in the corpus are mapped into the ontology with MetaMap [2], a toolkit that automatically codes text with the UMLS semantic types. For instance, "I need running clothes" would be coded as "I need running `[Daily or Recreational Activity]` clothes `[Manufactured Object]`". After that, [lines 5-9] we compute the strength of a mapping between a topic t_k and an ontology concept c_i. Note that this algorithm processes each concept independently, without leveraging the hierarchical structure of the ontology.

[Lines 10-13], following Alexander and Gleicher [1], we compute a mapping (i.e., alignment) between topics in the two topic models as the cosine similarity between their probability distribution of a word occurring given a particular topic, $p(w|t)$.

[Lines 14-19, we derive quantitative topic metrics per topic, including topic segregation (how well separated topics are), topic cohesion, and topic stability (see next section for details).

[Lines 20-23], we compute document distances within each topic model. For each pair of documents in the corpus, we compute their relative distance using the cosine similarity between their document-topic distributions.

2.2 Task model

The high-level goal of *OCTVis* is twofold: (i) assist communication between an NLP researcher and a domain expert during formative evaluation of topic modeling methods, and (ii) assist an NLP researcher in the development of these novel techniques. The process of refining these goals into a hierarchy of user analytic tasks to drive the interface design was based on principles from existing literature on visualizing comparisons [1, 13] and topic modeling analytics [5, 10, 12], as well as an informal iterative collection of user requirements from NLP and domain experts (including the authors), through brainstorming sessions and unstructured interviews.

In essence, the interface needs to support comparative evaluation of topic models based on task-driven metrics for topic model quality. This assessment can be decomposed into four key qualitative and quantitative metrics used to build a formal analytic task model.

2.2.1 Evaluative metrics

Topic segregation: relies on the intuition that an ideal topic model should have topics with little word or semantic overlap. While word overlap is easy to compute, semantic overlap is a more difficult problem. The interface should allow the users to gauge the semantic distance between topics qualitatively, by looking for duplicate words, synonyms and similar/related words shared across topics.

Topic coherence: The quality of a topic model depends on the quality of its topics. Essentially, words within a topic should share some overall semantic relatedness. This idea is captured by well defined quantitative metrics, like topic coherence [10], normalized pointwise mutual information [5], and topic stability [12].

Topic model comprehensiveness: measures how thoroughly the model covers the topics appearing in the corpus. This is very difficult to measure without a human-centred gold standard, because it requires understanding the entire corpus.

Topic assignment quality: A topic model should also map well both to ontology concepts and at the document level, in the sense of creating a meaningful middle level between the two. An ideal topic model is well situated both within the context of its text domain (and should map well to an ontology) and of individual documents (i.e. topics should reflect the actual text in each document).

2.2.2 Tasks

With these evaluative metrics in mind, we built a hierarchy of user analytic tasks for comparing ontology-based topic models. Here, we provide a summary of our task framework. Please refer to our supplementary material for the complete task list.

Evaluating overall topic model quality (TM) and individual topic quality (TQ): Throughout exploration of topic models, tasks for high-level comparison between the topic models (TM), as well as tasks evaluating quality of individual topics relative to one another (TQ), remain instrumental for understanding how and why topic models differ. These tasks include comparing topic model specificity, topic coverage, comprehensiveness, segregation, and coherence.

Assessing topic alignment (TA) and ontology mapping quality (O): Alignment between topics (TA) is a principle concern of topic model comparison. Tasks involving topic alignment across models include identifying the most similar and different topics across models, identifying which topic of a highly aligned pair better captures the underlying topic in the corpus, and whether one model is more coarse-grained or fine-grained than another.

Moreover, assessing the quality of mappings from topics to a domain ontology (O) is key in situating each topic model within the context of the text domain the corpus belongs to. Ontology mappings can inform topic model expectations (for instance, if a topic is strongly mapped to concepts "Pharmacologic Substance" and "Clinical Drug", we expect it to cover keywords and documents relating to drugs and medicine), suggest topic alignment (if concepts are prevalent across both models), indicate topic differences (if a concept is strongly mapped to one model's topics but not the other model's topics, it can be a red flag to investigate further differences in topic coherence, segregation, and even missed topics), and inform the quality of the ontology itself (if users identify concepts such as "Biologic Function" that might be more informative if broken down into "Physiologic" and "Pathologic Function", updating our ontology mapping in our framework will better facilitate ontology-based comparisons).

Exploring and comparing documents (DD, DT, DW): Finally, there remain a diverse array of tasks involving documents themselves. These include comparing document distances (DD) and how topic models cluster documents differently, exploring the document topic distribution (DT) to identify key topics and documents of interest for fine-grained comparison, and comparing document keywords (DW) within aligned topics across models.

3 Ontology-based Comparison of Topics (OCTVis)

Our proposed solution[2], shown in Figure 1, uses two linked views to assist in all the comparison tasks captured by our task model. To support our topic-centred tasks, a high-level topic-centred facet presents the topics from two topic models with their alignment, as well as the mappings from an existing ontology's concepts to each topic (top of Fig. 1). For more in-depth, document-centred tasks, a document-centred facet (below in Fig. 1) shows topic distributions per document, relative document similarities, as well as supporting exploration of each individual document. We now describe and justify most of the basic encodings used in *OCTVis*. For a detailed illustration of such encodings and the corresponding interactive techniques, we defer to the video associated with this paper.

Ontology Mapping: This builds on the solution presented in [1] to compare two topic models by *juxtaposition*, with a parallel-coordinates facet showing alignment between topics with color, line width, and opacity all redundantly encoding the alignment strength. Here (see top-center of Fig. 1), a vertical list of ontology concepts is inserted between the two topic models with a parallel-coordinate on each side of the list connecting each ontology concept with each topic within that side's topic model. The strength of each alignment between ontology concept and topic is computed as described in Section 2.1. Concepts in the list are sorted in descending order of combined strength from both topic models, so more frequent concepts are higher on the list than less frequent ones. Each topic is visualized as a "word cloud" to show topic keywords, with their strength encoded by font size [23]. Additional computed per-topic metrics (Section 2.2.1), such as topic coherence, are encoded with aligned horizontal bars beside each topic node.

[2]Source code: `http://github.com/humfuzz/octvis`

The order of the concept list and the concept mapping support ontology-based comparisons (O) like finding the most frequent concept per-topic and finding prevalent concepts within both or one topic model. Using the topic mappings and topic keywords, users can assess the informativeness of these concepts as well as whether they make sense within the context of the corpus.

Topic Alignment: The list of ontology concepts can be toggled off to show a direct alignment between the two topic models. Topic keywords and per-topic metrics are still presented beside each topic node, as in the ontology alignment view. Showing topic keywords beside topics supports the tasks of evaluating the quality of matched topics (TA): users can determine how well two topics match based on the computed alignment scores, compare the quality of the match using the topic keywords, as well as compare similarities and differences across model topics, between computed matches and non-matches. As well, users can evaluate individual topic quality (TQ) based on their understanding of the topic given the keywords, computed per-topic metrics, and through comparison between matched topics. For instance, users can use the topic keywords along with the computed metric bars to evaluate whether these measured metrics agree with human qualitative measures.

Document-centred comparison: Below each topic model's topics, we display a document-topic heatmap representing the distribution of topics within each document in the corpus (DT), expanding on prior work in document-centred comparison [1] (see Figure 1 bottom left and right). Each row represents a single document, and each column represents a topic from the above model. Within the heatmap, the saturation and lightness of the field encodes the strength of that topic (column) within that document (row). A checkerboard background hint relates the mapping from each column to each topic's keywords (see Figure 1 for these background hints).

To explore topic models at the document level (DW), single documents are displayed in the centre of the interface, similar to interactive topic modeling techniques presented by El-Assady et al. [11]. The user can select one topic from each of the two models and the strongest keywords for the selected topic are highlighted in the document view.

To support the comparison of document distances (DD), we adapt the parallel buddy plot encoding introduced by Alexander and Gleicher [1]. On each side of a topic model's document-topic heatmap, a buddy plot facet can be toggled to compare relative document distances in both models (see Figure 1 bottom left and right). Each line corresponds to a document, and the circles represent every other document's relative distance to the reference document. Inter-document distances in the adjacent model are encoded by horizontal distance, while distances in the opposite model are encoded by lightness; if the two models have similar document distances, each buddy plot line would follow a smooth gradient like the reference line at the top of the two buddyplots in Figure 1. Sharp changes in this gradient indicate documents that have differing distances across the two models (see [1] for more details). In *OCTVis*, we chose to use hollow circles instead of solid disks to see overlap between documents more clearly, and replaced hue encodings with lightness to prevent overloading established color encodings for the two topic models.

4 Case Study

4.1 Methodology

To test the viability of *OCTVis* on real data, we ran a case study with domain experts on two corpora of comments from online discussion forums: one for diabetes patients, the other for ovarian cancer patients. The two corpora have very different sizes (see Table 1). For each corpus, the two topic models being compared were generated with differing methods: Latent Dirichlet Allocation (LDA) [4], and Non-negative Matrix Factorization (NMF) [15]; both set to generate five topics. For the ontology mapping, we used the UMLS Semantic Network [17]. For each topic, we computed its segregation from the other topics in the same model using the average cosine similarity

Table 1: Case study discussions corpora statistics.

		diabetes	cancer
	# comments	201	56,537
# words per comment	min	1	1
	max	524	5,577
	mean	110	99
	st. dev.	93	130

of pairwise topic vectors. As for the set of documents that were actually shown in the visualization, in order to keep the case study manageable in term of time, a sample of the comments was selected and shown: 34 for the diabetes corpus and 27 for the ovarian cancer corpus. We assume, given more time, the domain and NLP experts would consider different and possibly larger samples of the data to perform a more in-depth comparison of the topic models.

Our four domain experts were all clinicians, who work with patients directly or facilitate online patient discussions. At the start of the study, we provided the experts with a definition of topic modeling and ontologies, and asked about whether they thought topic modeling can be informative in understanding a possibly large set of documents, whether they had a clear idea what a 'good' topic model should look like, and whether or not ontologies are useful in evaluating topic model quality. After this questionnaire, each domain expert was guided through the features of *OCTVis* on a fictional sample dataset and topics. During the main study, they were presented with one of two datasets and were free to explore the data using the interface. Since the goal of our interface involves facilitating communication between domain experts and NLP researchers, one of our NLP researchers discussed the results of the topic models and prompted topic model comparison with the domain expert throughout the main study. We were interested in how the interface, through topic model comparison with and without mapping to an ontology, could improve this communication and facilitate expert evaluation of the topic models. After the study, we asked the domain experts the same questions at the start to see whether their beliefs about topic models and ontologies changed, and asked for generic feedback on the interface.

4.2 Results

The case study indicated that *OCTVis* was useful in comparing the two models. From the pre-study and post-study questionnaires (see Table 2), we found that all the experts reported that their perception of topic model quality became clearer after using the interface (TM, TQ). Their belief that topic modeling can be informative in understanding a large set of documents was strong before and remained strong after the study. However, two experts lowered their assessment for the usefulness of ontologies, mainly because they found that the provided ontology was too generic.

Ontology Mapping (O): Two of our experts found the ontology mapping to be a useful high-level overview of the topics discussed. They were able to re-evaluate their understanding of topics based on their ontology mappings. For instance, in the ovarian cancer corpus, one topic with top keywords "chemo just time taxol like did good" mapped strongly to the ontology concept "Therapeutic or Preventive Procedure", and no other topics in either model mapped as strongly to this concept. With this information, the two experts were able to identify this topic as a cohesive and well-segregated topic about cancer treatment that the other model failed to capture. On the contrary, the other two experts found that the mapping contained concepts that were too general or uninformative. Because of this, enhancing *OCTVis* with techniques for interactive topic modeling (e.g., [11, 12, 14]) to allow domain experts to identify more specific and relevant sets of ontology concepts is our priority in future work.

Topic Alignment (TA): All the domain experts used the topic alignment graph to compare aligned topics and identify differences between a pair of similar topics. Generally, they found that the NMF model generated more coherent topics than the LDA model. For the diabetes corpus, one domain expert found that the LDA topics were more high level, covering the corpus with broad terms (like "diabetes" and "study"), whereas the NMF topics were more concrete (e.g. diet related terms "carbs", "sugar", "fat") . In the ovarian cancer discussions, another expert found that the LDA topics captured more medical issues (e.g., treatment, symptoms) while the NMF topics captured more psychosocial issues (e.g., emotions, people). This led the domain and NLP experts to consider the possibility of combining the two models in future work.

Table 2: Pre-study and post-study questionnaire responses. Domain experts were asked how strongly they agreed with the following statements (scale from 1-5, with 0.5 increments, where 1 is strongly disagree and 5 is strongly agree): (S1) Topic modeling can be informative in understanding a possibly large set of documents. (S2) I have a clear idea of what a 'good' topic model should look like. (S3) Ontologies are useful in evaluating topic model quality. For each expert and question, **bold** numbers indicate stronger or equal agreement *before* or *after* the study.

domain expert	S1		S2		S3	
	pre	post	pre	post	pre	post
e1	**4**	**4**	4	**4.5**	4	**4.5**
e2	4	**5**	3	**5**	**4**	3
e3	**4**	**4**	3	**4**	**4**	**4**
e4	**5**	**5**	4	**5**	**4**	2

Exploring Documents (DD, DT, DW): This was especially useful to two of the clinicians who used in-depth exploration of the documents to assess topic quality and improve their understanding of the topics. For instance, in the diabetes corpus, the document-topic distribution and highlighted keywords in individual documents helped confirm labels for topics including "exercise" and "diet", as well as determining that a "medication" topic was missed by both models. The other two experts mainly focused their analysis on the topic and ontology alignments.

5 Conclusion and Future Work

We expand on prior work in topic modeling visualization by enabling NLP and domain experts to compare topic models at multiple granularities, from high-level domain-specific mappings into ontology concepts, to in-depth document keyword exploration. To the best of our knowledge, we are the first to use a domain ontology to help interpret topics, and also the first to perform an evaluation of a topic modeling comparison interface with prospective users. Such evaluation as a proof-of-concept, through a case study with clinicians, has shown several potential benefits of our proposal. Future user studies to compare systems with and without specific features will be beneficial for more thorough evaluation. A key venue for future work is scalability to larger numbers of topics, ontology concepts, and documents. Interactive techniques including filtering and drilling down can tackle the issue of scale on many data dimensions. Topics, documents, and ontologies can be clustered hierarchically and allow users to explore the data space at multiple granularities. For instance, for large ontologies, analysis could focus on more fine-grained subtrees that are specific for relevant topics (e.g. cancer or diabetes sub-ontologies for our corpora) and allow interactive expansion and collapse of these concepts. While the alignment encoding with parallel coordinates can become cluttered with dozens of topics or concepts, heatmap matrices and other space-filling idioms can scale more effectively; our topic-ontology mappings can be redundantly encoded with these visual encodings as well.

Acknowledgments

We would like to thank our reviewers for their valuable comments during the revision of this paper, as well as Richard Lester, John-Jose Nunez, and Patrick Boutet for their detailed feedback during our development and case studies.

References

[1] E. Alexander and M. Gleicher. Task-driven comparison of topic models. *Trans. Visualization and Computer Graphics (TVCG)*, 22(1):320–329, Jan 2016.

[2] A. R. Aronson and F.-M. Lang. An overview of MetaMap: historical perspective and recent advances. *Journal American Medical Informatics Assoc.*, 17(3):229–236, 2010.

[3] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. Dbpedia-a crystallization point for the web of data. *Web Semantics: science, services and agents on the world wide web*, 7(3):154–165, 2009.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal Machine Learning Research*, 3:993–1022, Jan 2003.

[5] G. Bouma. Normalized (pointwise) mutual information in collocation extraction. In *Proc. German Society Comp. Ling. Language Tech. (GSCL)*, pp. 31–40, 2009.

[6] G. Carenini, G. Murray, and R. Ng. Methods for mining and summarizing text conversations. *Synthesis Lectures on Data Management*, 3(3):1–130, 2011.

[7] K. Chan, X. Lou, T. Karaletsos, C. Crosbie, S. Gardos, D. Artz, and G. Ratsch. An empirical analysis of topic modeling for mining cancer clinical notes. In *IEEE 13th Intl. Conf. on Data Mining Workshops (ICDMW)*, pp. 56–63. IEEE Computer Society, Los Alamitos, CA, USA, dec 2013.

[8] J. Chang, J. Boyd-Graber, C. Wang, S. Gerrish, and D. M. Blei. Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems (NIPS)*, 2009.

[9] J. Chuang, S. Gupta, C. D. Manning, and J. Heer. Topic model diagnostics: Assessing domain relevance via topical alignment. In *Proc. Intl. Conf. Machine Learning (ICML)*, 2013.

[10] E. T. M. L. D. Mimno, H. M. Wallach and A. McCallum. Optimizing semantic coherence in topic models. In *Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 262–272. Association for Computational Linguistics, 2011.

[11] M. El-Assady, R. Sevastjanova, F. Sperrle, D. Keim, and C. Collins. Progressive learning of topic modeling parameters: A visual analytics framework. *Trans. Visualization and Computer Graphics (TVCG)*, 24(1):382–391, Jan 2018.

[12] M. El-Assady, F. Sperrle, O. Deussen, D. Keim, and C. Collins. Visual analytics for topic model optimization based on user-steerable speculative execution. *Trans. Visualization and Computer Graphics (TVCG)*, 25(1):374–384, Jan 2019.

[13] M. Gleicher. Considerations for visualizing comparison. *Trans. Visualization and Computer Graphics (TVCG)*, 24(1):413–423, Jan 2018.

[14] E. Hoque and G. Carenini. Interactive topic hierarchy revision for exploring a collection of online conversations. *Information Visualization*, 2018.

[15] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

[16] F. Mahdisoltani, J. Biega, and F. M. Suchanek. YAGO3: A Knowledge Base from Multilingual Wikipedias. In *CIDR*. Asilomar, United States, Jan. 2013.

[17] A. T. McCray. An upper-level ontology for the biomedical domain. *International Journal of Genomics*, 4(1):80–84, 2003.

[18] Microsoft. Microsoft concept graph and concept tagging release, 2016.

[19] T. Munzner. A nested model for visualization design and validation. *Trans. Visualization and Computer Graphics (TVCG)*, 15(6):921–928, Nov 2009.

[20] T. Munzner. *Visualization analysis and design*. AK Peters Visualization Series, CRC Press, 2014.

[21] D. Oelke, H. Strobelt, C. Rohrdantz, I. Gurevych, and O. Deussen. Comparative exploration of document collections: a visual analytics approach. *Computer Graphics Forum*, 33(3):201–210, Jul 2014.

[22] A. Ramesh, D. Goldwasser, B. Huang, H. Daume, and L. Getoor. Understanding MOOC discussion forums using seeded LDA. In *Proc. 9th Workshop Innovative Use of NLP for Building Educational Applications*, pp. 28–33. Association for Computational Linguistics (ACL), Jun 2014.

[23] C. Seifert, B. Kump, W. Kienreich, G. Granitzer, and M. Granitzer. On the beauty and usability of tag clouds. In *12th Intl. Conf. Information Visualisation*, pp. 17–25, Jul 2008.

[24] K. Stevens, P. Kegelmeyer, D. Andrzejewski, and D. Buttler. Exploring topic coherence over many models and many topics. In *Proc. Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 952–961. Association for Computational Linguistics, Stroudsburg, PA, USA, 2012.