

Annotation Consensus: Implications for Passage Recommendation in Scientific Literature

Shannon Bradshaw
Department of Math and Computer Science
Drew University
36 Madison Ave
Madison, NJ 07940
sbradsha@drew.edu

Marc Light
Research and Development Department
The Thomson Corporation
640 Opperman Dr.
Eagan, MN 55123
marc.light@thomson.com

ABSTRACT

We present a study of the degree to which annotations overlap when several researchers read the same set of scientific articles. Our objective is to determine whether there is sufficient evidence to suggest that information about which passages initial readers tend to annotate might be used to recommend important passages to later readers of the same material. We found that readers exhibit a high degree of overlap in the passages they annotate, that these passages account for a small but significant fraction of the total document, and that such passages are distributed throughout a document rather than concentrated in the same few sections in each paper (e.g., the results section). These findings indicate that work on developing a passage recommendation model based on annotation is warranted.

Categories and Subject Descriptors

H.1 [Information Systems]: Models and Principles; H.3.7 [Information Systems]: Information Storage and Retrieval—*Digital Libraries*

General Terms

Experimentation

Keywords

Annotation, annotation consensus, knowledge management, passage recommendation

1. INTRODUCTION

Maintaining currency is increasingly difficult: Scientists and other knowledge workers find it increasingly difficult to maintain currency in their areas of expertise. This is largely due to the increasing amounts of information with which they must be familiar. For example, 40,000 new citations are

added to the MEDLINE repository of life sciences literature each month [12]. A number of studies of information gathering behavior among scientists in many fields indicate that researchers spend a significant amount of their time (as much as 35%), searching, gathering, and reading information necessary to their work [14, 24]. If we can dramatically reduce the time required to keep pace without degrading a researcher's understanding of related work, more time can be spent in pursuing new findings, developing innovative technologies, etc.

Reducing the reading burden: A portion of this problem lies in reducing the amount of time required in order to read and understand individual articles. A large body of automated summarization work targets this problem [8]. We suggest an alternate approach and here present an initial step in determining whether such a strategy might prove fruitful. The technique we envision is one of passage recommendation based on the consensus of many prior readers of a document who have annotated as they read. A means of presenting the type of recommendations we envision would be to highlight within an article all passages that 2/3 or greater of all previous readers have marked in some way (e.g., underlining, margin note, etc.). Though motivated by the hypothesis that such an approach would be successful, in this paper we present only a preliminary step in this direction. We do not outline a recommendation model or system, rather we simply present a paper-based study of annotation consensus. Our objective in performing this study was to determine whether there is sufficient overlap among the annotations of our test subjects to warrant work developing an annotation-based passage recommendation system.

Research questions: In a study of annotation behavior among a group of scientists reading the same set of papers, we asked the following research questions:

- **Q1:** To what degree did readers overlap in the passages they highlighted, underlined, or in some other way annotated?
- **Q2:** On what fraction of the reading material did annotators focus the majority of their attention?
- **Q3:** What is the distribution of annotations throughout documents?

Q1 rationale: We are interested in whether there is significant consensus as to the important reading material in a document. If so, then the passages jointly annotated by a small set of initial readers are likely to be of particular

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HT'07, September 10–12, 2007, Manchester, United Kingdom.
Copyright 2007 ACM 978-1-59593-820-6/07/0009 ...\$5.00.

interest to later readers of the same document. Aggregated annotation information may be used to recommend a reading path other scientists might follow through a document so as to quickly learn its important contributions.

Q2 rationale: A passage recommendation strategy is useful in reducing the reading burden only if the number of passages recommended is small relative to the size of an entire document. Therefore annotations must be concentrated among a small subset of the passages of a document in order to use annotation consensus to recommend passages to future readers.

Q3 rationale: If annotations do cluster in a relatively small fraction of the passages within documents, then we would like to know if there is some factor driving annotation consensus that derives from the structure of a paper. For example, if most annotations are found in the discussion section of a paper then there is no need for a passage recommendation strategy of the kind we envision. If such a phenomenon exists, then readers are likely to already know to skip to the discussion section and begin reading there first.

2. CONTRIBUTIONS

Our findings indicate that scientists with roughly the same degree of subject familiarity who read the same article exhibit significant overlap in the passages they annotate when reading an article. Furthermore, our results suggests that these passages account for only a small fraction of the article on average. Finally, annotations are distributed widely throughout an article and do not appear to be isolated to any one particular section or section type. Our conclusion is that the consensus of annotators of a document may be useful in directing the attention of future readers with related interests so as to reduce the amount of time required to find and absorb the important information that document contains.

3. RELATED WORK

Look and feel of annotations: Our research is informed by work targeted at understanding annotation behavior. Going back as far as 80 years, many scholars have studied the ways in which readers most commonly mark documents [3, 7, 21]. Recent work in this area focuses on understanding reading and annotation behavior on paper in order to better design electronic annotation systems, particularly with regard to their user interface [22, 16, 17, 25, 10, 28, 23].

Passage-level annotation consensus: In the work presented here, we move beyond the study of what types of marks readers make. We measure the degree to which the various marks of a several readers cluster together as we “overlay” several copies of the same document each marked by a different annotator. We looked for agreement of a sort as to the passages of importance within a collection of documents. This facet of annotation behavior is largely unexplored; however, one piece of work in this area is that of Cathy Marshall [17]. As part of a larger study, Marshall looked at the degree of overlap among annotations for a small number of copies of a textbook used in an undergraduate university course. Building on this work, we considered annotation consensus across many documents in the domain of scientific literature. In contrast to Marshall’s work, we control for factors such as group discussion and lectures that might bias the behavior of readers toward greater overlap in annotation.

Social information navigation: Also related is research on social information gathering, filtering, and search. A variety of systems exist that aggregate the tags individuals apply to the documents they read in order to enable others to find documents on topics of interest to them. CiteULike¹, Del.icio.us², and Technorati³ provide examples. Bao et al. [1] explore the use of such tag aggregators to improve web search. They present one algorithm that ranks documents by matching web queries to tags applied to web pages using the Del.icio.us service. They outline another that determines the importance of a page based on the number times it has been tagged. Moving from tags to other forms of social navigation, Farzan and Brusilovsky [11] describe a Footprints-like [27] approach based on the ways in which students interact with a set of on-line resources. In the Knowledge Sea II system in which this approach is implemented, students are presented with a color-coded map indicating the most popular documents within the collection (based on the activity of other students) and the purposes for which other students found the information useful. Later work by the same authors [13] extends these ideas in work targeted at the ACM digital library. Though this work does make use of highlighting and other forms of passage-level annotation, such annotations are employed in an information navigation context only at the document level to help users choose among many search results. In general, social navigation approaches to date either replace or augment traditional information retrieval techniques with the collective opinion of a body of readers. In this respect, this type of technology is related to what we envision as several of the next steps we will take in our own research. The primary difference is that in future work we will explore the recommendation of passages within documents a person has already chosen to read rather than recommendation of the documents themselves.

Annotation for collaboration: Less related, though still important to our work, is research on the sharing of annotations for the purpose of collaboration or discussion. This research informs our own largely in that it points to the value in reusing the annotations of others and lends additional credence to our conclusion that aggregated annotation may serve as an effective basis for a passage recommendation model. Several systems support synchronous online discussion in an educational setting by providing students with the ability to annotate and communicate around various forms of information included in homework assignments [4, 9, 2]. Other work explores the use of annotation as a reading and learning tool within the classroom [26, 20]. Finally, many researchers have explored shared annotation as a tool for synchronous and asynchronous collaboration around document creation and other forms of knowledge work [29, 5, 18, 15, 19].

4. STUDY OVERVIEW

Guiding principles: In performing the study described below, our objective was to gain an initial understanding of annotation consensus among researchers reading the same article. We identified an environment in which the backgrounds of the readers was related, but with significant differences so as to ensure varying perspectives that occur in real-world

¹<http://www.citeulike.com>

²<http://del.icio.us>

³<http://technorati.com>

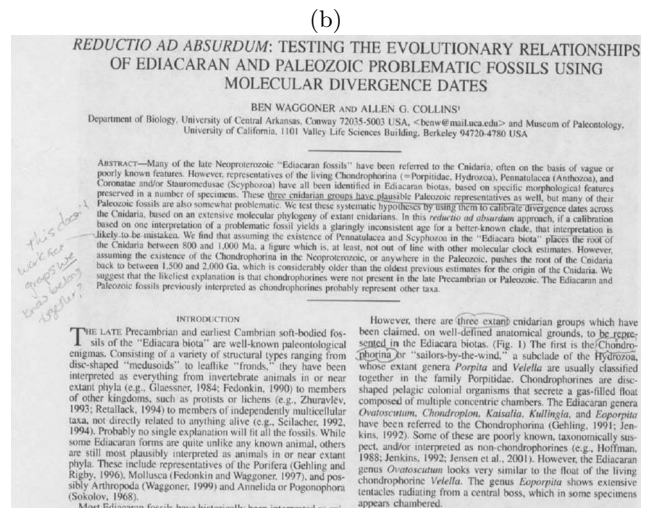
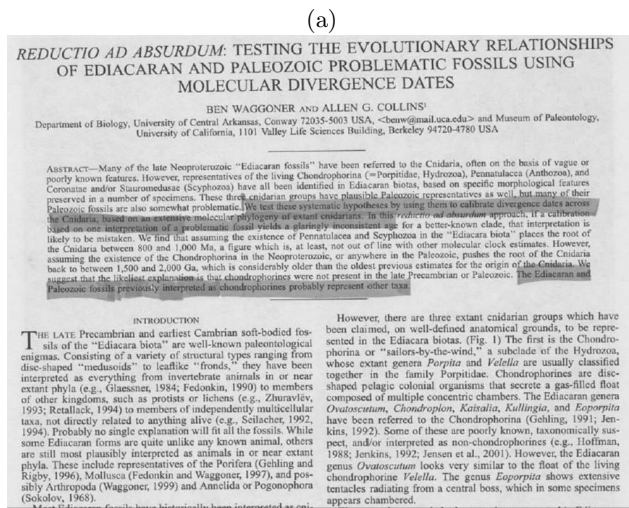


Figure 1: Examples of marks made by two different annotators of the same document. Each annotator worked with his/her own copy.

situations. In addition, we structured the study so as to eliminate bias due to lecture, discussion, or other interactions that might influence annotation behavior. Finally, we sought out an environment in which several researchers were motivated to read the same set of scholarly articles for their own purposes.

Environment: We performed our study in the context of a graduate seminar on molecular phylogenomics in the Department of Biology at the University of Iowa. The class was composed of eight graduate students from four distinct laboratories in three different graduate programs (molecular biology, genomics, and geoscience). The first half of the course (eight weeks) consisted of lecture on the topic of molecular phylogenomics. In the second half, each student was required to select two papers that he or she would present to the rest of the class. Each presenter provided the rest of the graduate students with two papers to read at least three days in advance of his or her presentation. To provide a focus for reading, the presenter also emailed to the rest of the class a brief (few sentences) description of the topic he or she would be presenting. One student per week presented both papers he or she had selected. To avoid interrupting the natural use of the readings as reference material during a presentation and the following discussion, we collected the annotated papers at the end of each class. However we attended every class meeting and observed in-class discussion to ensure that the papers we collected had been annotated prior to the presentation and that no additional marks were made during the presentation or subsequent discussion. We did not, however, participate in discussion.

Instructions to test subjects: The instructions we gave our test subjects as to reading and annotation procedure were exactly as follows:

1. Read the topic summary provided by the presenter so that you understand the subject of the talk and thus what information within the papers will be relevant to the discussion.
2. Annotate the papers with an eye toward that which is relevant to this topic. The function of the annota-

tions should be to enhance your understanding of the material and ability to quickly review it later. Typical annotations include highlighting, underlining, asterisks, etc. However, annotations are not limited to these marks, please use those you commonly use in your research.

3. When highlighting passages, figures, etc. of relevance please use the pink highlighter provided, a pen or a pencil. Yellow and other colors of highlighters do not photocopy.
4. In determining how much or what to highlight please use your own discretion; however, please bear in mind that either extreme (the entire paper or no highlighting at all) are of no use to us in this study.
5. Please bring your annotated papers to class on the day of the corresponding presentation. We will collect the papers, photocopy them, and return them to you at the next class.

5. DATA COLLECTION

Collecting hardcopy: We collected paper copies of 16 articles that had each been annotated by four to seven readers, depending on how many students chose to participate in our study in a given week. For each paper, therefore, we gathered between four and seven distinct copies, each of which had been read and marked by a different reader. The types of annotations readers applied to the articles they read included:

- Highlighting using a marker
- Underlining
- Circling or boxing
- Vertical lines drawn in the margin (margin bars).
- Exclamation points, asterisks, and other emphasis marks.
- Margin notes

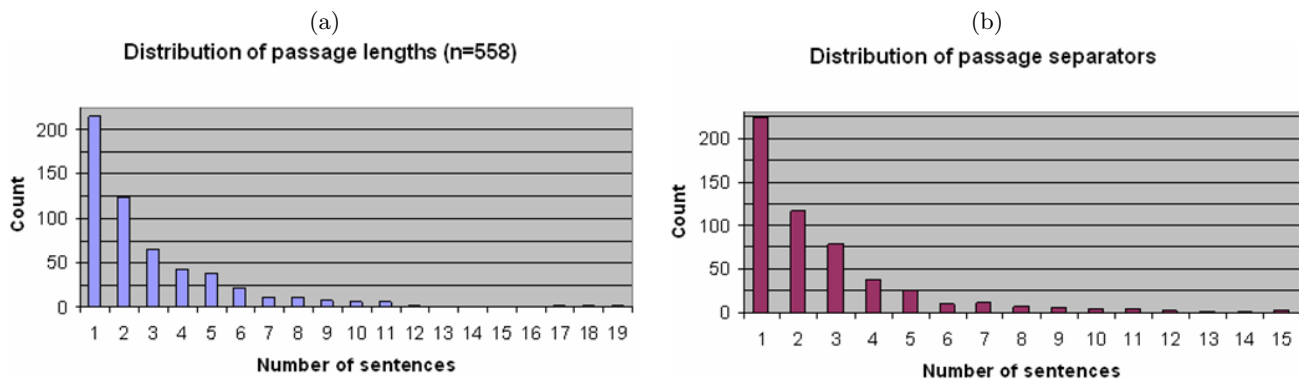


Figure 2: (a) Distribution of the number of sentences in passages of interest. (b) Distribution of the number of un-annotated sentences found between passages of interest.

Such marks are in keeping with the annotation behavior reported by Marshall [17]. Figure 1 provides examples of the marks made by two different readers of the same paper in our study.

Transcribing hardcopy: To transform the information in the papers we collected into a form we could use for evaluation, an undergraduate transcribed the paper annotations into xml tagged versions of each document using Mitre’s Calisto tool [6]. We broke each paper down to sections, paragraphs, and sentences. For each sentence, we tracked which readers had marked it in any way. For our purposes, the relationship between a textual unit and a reader is binary; if a reader marked a sentence in whole or part in any way, we considered it marked. For example, reader GK underlined sentence 180 in paper 13; reader SL circled two separate words in the same sentence; and reader TM wrote a margin note next to this sentence. All three readers were marked as having annotated sentence 180 in paper 13. No matter how many separate marks a reader made in a sentence, for our purposes, a reader either marked a sentence or she did not. As discussed in Section 3, the types and frequency of annotations made by individual readers of a document has been studied by many researchers. Agreement among readers as to passages of interest as indicated by annotations is a largely unexplored area of annotation behavior. In this paper, we are concerned solely with annotation consensus.

6. GENERAL STATISTICS

Document statistics: Our data set contains 16 papers, containing a total of 3516 sentences. We excluded bibliographies from consideration; only two annotations in all the data we collected were found in a references section. The average number of sentences in an individual document is 220 with the smallest being 46 and the largest, 364. Of the total of 3516 sentences, 1639 (47%) were annotated by at least one reader. The smallest fraction annotated in any one document was 31% and the largest was 72% (in the document containing only 46 sentences).

Annotator statistics: Readers annotated an average of 13% of the sentences they read. The smallest fraction annotated by any reader was 8.5%, the largest was 23%. The reader we identified as GK annotated 23% of the sentences. The frequency with which this subject annotated was the only one of the eight that was significantly greater than the rest.

7. PASSAGES OF INTEREST

Finding passages: As is probably common to most research articles, the authors of the documents in our study introduce an idea and then proceed to work through that idea for several sentences. As a result, several readers marked different sentences in close proximity that identify the same or very similar take away information from an article. In some instances this information spans multiple paragraphs; in others it is contained within a single paragraph. As a result, we sought a means of determining passages of interest to readers that reflects what we found in our data, but does not introduce bias into the study. We settled on an automated technique that defines a passage of interest to be a continuous sequence of sentences in a single document for which at least one reader annotated every sentence. Passages under this definition are bounded by at least one sentence that no reader annotated or by the beginning or end of the document.

Passage statistics: Among the annotated sentences in our data we found 558 distinct passages. The mean passage length is three sentences (median of two), with a minimum of one and a maximum of 19. Figure 2a depicts the distribution of passage lengths. The mean number of sentences between passages is three (median of two) with a minimum of one and a maximum of 47. The distribution of the number of sentences separating passages is charted in Figure 2b. There are five additional separators of various lengths beyond the largest length of 15 shown in this chart. We trimmed the tail to make the bulk of the data more easily viewed.

8. ADDRESSING Q1

In this section, we present findings that address the question: **To what degree did readers overlap in the passages they highlighted, underlined, or in some other way annotated?**

Normalizing consensus: Our study included a total of eight graduate students serving as test subjects. For each of 16 papers, we collected annotations from between four and seven subjects. Seven subjects annotated copies of six different papers, six annotated copies of five papers, five annotated copies of two papers, and four annotated copies of three papers. To normalize the number of annotators across papers, we broke down the consensus levels into four percentiles:

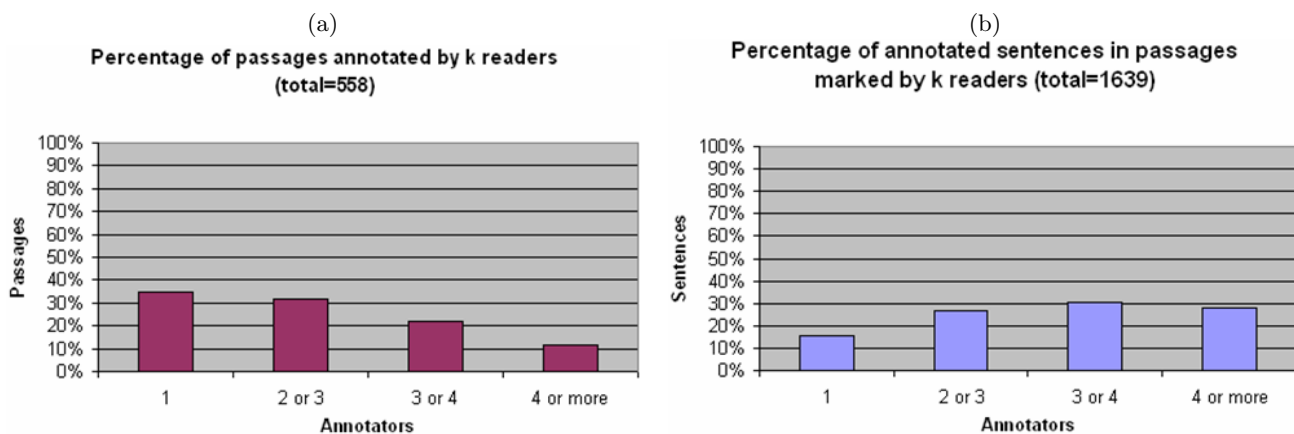


Figure 3: (a) Fraction of passages annotated by k or more readers. (b) Fraction of all annotated sentences found in passages annotated by k readers.

1. $0\% < \text{consensus} \leq 25\%$ (1 reader)
2. $25\% < \text{consensus} \leq 50\%$ (2 or 3 readers)
3. $50\% < \text{consensus} \leq 75\%$ (3 or 4 readers)
4. $75\% < \text{consensus} \leq 100\%$ (4 or more readers)

The percentages here represent the ratio of readers who annotated a sentence to the total number of readers for a paper. For example, a sentence annotated by three of the four people from whom we collected annotations for paper 14 would fall into the third percentile above, because it was annotated by 75% of the readers for that paper. With between four and seven annotators per paper these percentiles indicate one annotator, two or three annotators, three or four annotators, and four or more annotators, respectively.

High levels of consensus: In evaluating the level at which readers agree in their annotation of passages, we considered passages as a whole. We identified any reader who marked at least one sentence in a passage in any way as an annotator of that passage. We then calculated what fraction of the annotators for a given document marked each passage. The results are shown in Figure 3. Figure 3a shows that passages marked by one or two annotators account for 371 or 67% of all annotated passages, with consensus of three or four annotators on 121 (22%) passages, and consensus of four or more on 66 (12%) passages. However, the overwhelming majority of passages marked by one or two readers consist of a single sentence. When we look at the number of sentences in annotated passages, we see that the majority of annotated content is marked by 50% or more of the readers. Figure 3b shows the fraction of annotated sentences contained in passages at each level of annotation consensus. There were a total of 1639 passages annotated by at least one reader in our study. Sentences in passages with consensus of 3/4, 3/5, 4/6, or 4/7 annotators account for 496 (30%) of all annotated sentences. Sentences in passages with consensus of 75% or more of annotators account for 455 (28%) of all annotated sentences. The majority of annotators agree on almost 60% of the annotated sentences found throughout the documents in our study. These sentences account for just over 27% of the total number of sentences in our dataset.

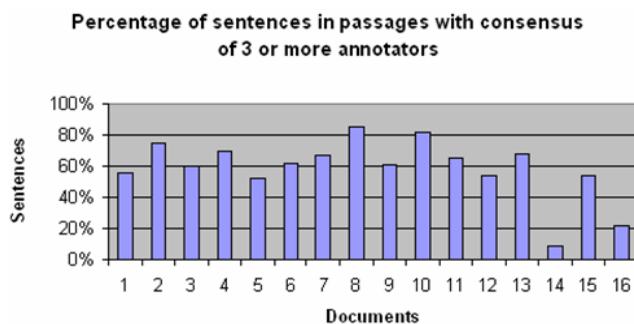


Figure 4: Fraction of annotated sentences found in passages with a consensus of three or more annotators. The results are presented for each document individually.

Consistent with individual documents: At the individual document level, Figure 4 charts the fraction of annotated sentences contained in passages on which there was annotator consensus of 3 or more. With the exception of documents 14 and 16, the results meet or exceed what we found when considering the data set as a whole. This indicates a reliably high level of consensus among readers for approximately 60% of annotated sentences. Document 14 was the last document discussed on the next-to-last class of the seminar course we studied. Document 16 was the last document presented on the last day. Respectively, they are also the longest and third longest documents included in our study, containing 100 sentences more than most of the other documents. The marks made by annotators of these documents represent a departure from the annotation frequency of these same readers in other documents in our study. We believe these subjects were simply fatigued or overburdened by their end of semester workload and that they gave documents 14 and 16 only a cursory look.

Statistical Significance: We determined the statistical significance of annotation consensus at the sentence level. Our objective was to demonstrate that the degree to which the annotations of individuals overlap could not have happened by chance. We performed a Monte Carlo simulation to determine the distribution of overlap if sentences were annotated

at random in proportion to what we found for each of our eight annotators. We found that the overlap in annotations between readers in our study is significant ($p=.01$) for as few as two or more annotators of the same paper.

9. ADDRESSING Q2

In this section, we present findings addressing the question: **On what quantity of the reading material did annotators focus the majority of their attention?**

Calculating the concentration curve: We have defined an annotation as a reader/sentence pair. If a reader marked a sentence in any way, we treat the entire sentence as having been highlighted by the reader. If a single sentence was marked by five different readers then that sentence is involved in five annotations. Using this definition of an annotation, we calculated a curve that shows how the concentration of annotations relate to the quantity of material read. The result is depicted in Figure 5. To generate this curve, we determined the number of individual annotations found in each passage of interest. For each passage, every sentence was marked by at least one reader. Some were marked by more. We simply counted the number of reader/sentence pairs in each passage and then sorted passages by this number. The Y axis values are determined by the sum of sentences in passages as we move from those containing the largest number of annotations to those containing the least. We could have achieved even higher concentration numbers had we worked with individual sentences and sorted sentences by the number of annotators. However, we performed this calculation at the passage level because our data indicates that it is passages around which reader attention is concentrated, not individual sentences.

Annotations concentrated in small fraction of material: At least one reader annotated 46% of all sentences read by our test subjects. However, the overwhelming majority of these annotations are concentrated in a small but significant fraction of the material. Figure 5 shows that half of all annotations are concentrated in just 1/5 of the sentences. Nearly 80% of annotations are concentrated in passages accounting for only 1/3 of the total data set. This result combined with our results for Q1 indicate that a passage recommendation system based on the annotations of prior readers of a document may significantly reduce the time required for readers to learn the key contributions of a research article.

Consistent with individual documents: Individual documents exhibit a similar annotation concentration curve. For example, in every document, 80% of the annotations are found in passages accounting for between 1/4 and 1/2 of the total document. For most documents, this concentration is found in 1/3 to 2/5 of the document.

10. ADDRESSING Q3

In this section we address the question: **What is the distribution of annotations throughout a document?**

Rationale: A passage recommendation system is necessary only if the key passages in all articles are not always found in the same portions of a document or in some other way known from the outset. If, for example, the majority of take away information is concentrated in the discussion section then new readers can simply focus their attention on this section. If annotations are concentrated in the same few sections of most documents, then it is likely that the

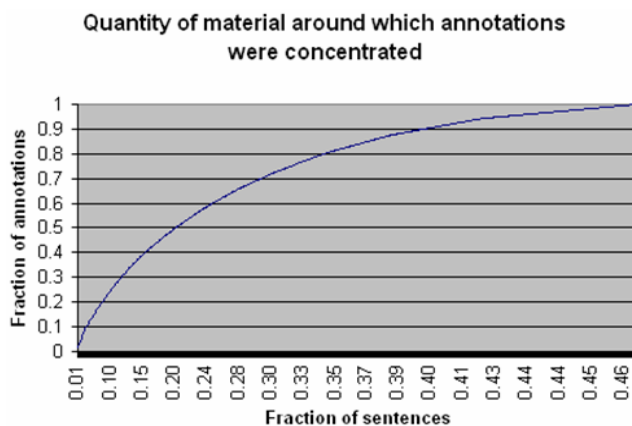


Figure 5: Curve depicting the concentration of annotation events (one annotator marking one sentence) in varying quantities of reading material.

attention of readers will be focused in these sections as well. There would, in such a case be little use for a passage recommendation system of the kind we envision.

Analysis procedure: To evaluate the distribution of annotations throughout documents in our study, we broke up each document into 10 consecutive segments (percentiles) containing an equal number of sentences. For each document, each segment contained 10% of the sentences in the entire document. We excluded bibliography sections. For example, for a document containing 200 sentences, the 10th percentile contained the first 20 sentences in that document. The 20th percentile contained the next 20 sentences. The 30th percentile contained the next 20 and so on until the end of the document. Looking across documents, then, we evaluated what fraction of all annotations in the data set fell into each percentile. Once again, an annotation is the event of one reader annotating a single sentence.

Annotations distributed throughout documents: Our findings indicate that readers do not focus their attention to any extreme in one part of a document and that a passage recommendation system for scientific literature may be useful in helping readers find the key information in an article more quickly. The readers in our study annotated in such a way that their marks were distributed evenly through all documents, for the most part. There was a small drop off in annotation concentration in the last 20% of most documents, but even these portions of the documents received significant attention from our readers. Figure 6 shows the distribution of annotations throughout documents read during our study. Close to 10% of all annotations occurred in the first nine segments containing 10% of the sentences for each document. There was a slightly greater concentration of annotations (12.6%) near the middle of the documents and a noticeable drop off in annotations near the end of documents.

Annotations distributed throughout section types: Articles in the field of biology exhibit a fairly consistent structure across disciplines and subspecialties with regard to a handful of section types. In particular, these sections include: *Introduction, Methods, Results, Discussion, Conclusion*. We performed an analysis similar to that described above for specific section types across all documents. As with the

analysis above, we found that annotations are distributed fairly evenly throughout specific section types.

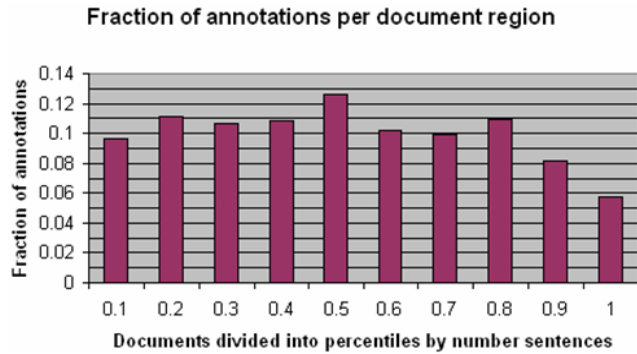


Figure 6: Documents were divided into regions by number of sentences. The 10th percentile represents the first 10% of sentences in the documents. The 20th percentile is the second 10% of sentences and so on. The Y axis shows the fraction of annotations that fell into each percentile.

11. DISCUSSION

Our objective in performing the study described here was to gain an understanding of the relationships between the annotations of multiple readers of the same research articles. We hoped to gain some insight as to whether aggregating the annotations of individuals has potential to benefit later readers of the same documents with similar interests and objectives. Our findings indicate that the annotations of a small set of initial readers of a document may make an effective basis for directing later readers to passages they would find useful.

60% or more of our test subjects annotated the same passages from documents in our data set. This behavior was largely consistent across all 16 documents. The sentences in these passages account for just over 27% of all the sentences in our data set. Therefore, we found a high degree of annotation consensus on a relatively small, but significant fraction of the material our subjects read. It is likely that future readers of the same documents having related interests would find these passages useful. Furthermore, since the volume of this material is only 1/4 of the size of the entire document collection, highlighting these passages may provide later readers with a means of learning the essential take away information in these articles in a fraction of the time it would take ordinarily.

Finally, annotations do not appear to be concentrated in one or more specific section types (e.g., the results section). Rather, it appears that the location of information of interest in a document is unique to that document and based on the content rather than the structure. The annotations of readers in our study were densely concentrated into a relatively small fraction of each document, yet these passages were distributed almost uniformly throughout an entire document.

12. FUTURE WORK

The readers in our study had a fair amount of common ground in approaching the documents they read. However,

they did not have a specific purpose in reading the documents. They were not looking for an answer of some kind. In many situations, researchers read articles with a particular agenda in mind. We hypothesize that two readers with largely the same objective in reading a paper would exhibit an even greater degree of overlap in annotation than what we saw in this study. We also hypothesize that two researchers reading the same paper for different reasons would overlap less. Any passage recommendation system based on annotation would need to make these distinctions. In addition, the volume and quality of annotation differs from one individual to another as we saw with our test subject, GK. These variations would need to be taken into account in any passage recommendation model.

Our next step in this work will be to evaluate the differences in annotation consensus when readers have the same, different, or no clear objective in mind. We are currently in the initial stages of analyzing data we have collected for such a study.

Beyond this we will expand our experiments to studies larger in scale so that we can gain an understanding of variation in individual annotation behavior.

13. ACKNOWLEDGMENTS

We would like to thank Dylan Scott, an undergraduate student at the University of Iowa, for meticulous attention to detail in transcribing the paper-based annotations used in this study. We also thank Kurt Anstreicher, Professor and Chair, Department of Management Sciences at the University of Iowa for developing Monte Carlo simulations to test the statistical significance of our annotation consensus results. Finally, our thanks to Sarah Abramowitz, Associate Professor of Mathematics at Drew University for consulting on several aspects of evaluating and presenting the results of this study.

14. REFERENCES

- [1] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, and Z. Su. Optimizing web search using social annotations. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 501–510, New York, NY, USA, 2007. ACM Press.
- [2] D. Barger, A. Gupta, J. Grudin, and E. Sanocki. Annotations for streaming video on the web: System design and usage studies. In *Proceedings of the Eighth International World Wide Web Conference (WWW8)*, 1999.
- [3] J. E. Barnett, F. J. D. Vesta, and J. T. Rogozinski. What is learned in note taking? *Journal of Educational Psychology*, 23(2):181–192, 1981.
- [4] A. J. B. Brush. *Annotating Digital Documents for Asynchronous Collaboration*. PhD thesis, University of Washington, 2002.
- [5] J. Budzik, S. Bradshaw, X. Fu, and K. J. Hammond. Clustering for opportunistic communication. In *Proceedings of the Eleventh International World Wide Web Conference*. ACM Press, 2002.
- [6] The callisto annotation tool (mitre corporation). <http://callisto.mitre.org/>.
- [7] C. C. Crawford. Some experimental studies of the results of college note-taking. *Journal of Educational Research*, 12:379–386, 1925.

- [8] H. T. Dang. Overview of duc 2006. National Institute of Standards and Technology (NIST), 2006.
- [9] J. Davis and D. Huttenlocher. Shared annotation for cooperative learning. In *Proceedings of the Conference on Computer Supported Cooperative Learning*, 1995.
- [10] C. Decurtins, M. C. Norrie, and B. Signer. Digital annotation of printed documents. In *Proceedings of CIKM 2003*, 2003.
- [11] R. Farzan and P. Brusilovsky. Social navigation support through annotation-based group modeling. In *In proceedings of 10th International Conference on User Modeling*, 2005.
- [12] P. Fontelo, M. Ackerman, and G. Kim. Development of evidence-based medicine resources: Bridging clinical research to medical practice. In *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004.
- [13] J. Freyne, R. Farzan, P. Brusilovsky, B. Smyth, and M. Coyle. Collecting community wisdom: integrating social search & social navigation. In *IUI '07: Proceedings of the 12th International Conference on Intelligent User Interfaces*, pages 52–61, New York, NY, USA, 2007. ACM Press.
- [14] W. Hayes. Text mining: Next steps for drug discovery (invited talk). NAACL/HLT 2004 BioLink Workshop, 2004.
- [15] M. Leland, R. Fish, and R. R. Kraut. Collaborative document production using quilt. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1998.
- [16] C. C. Marshall. The future of annotation in a digital (paper) world. Presented at the 35th Annual GSLIS Clinic: Successes and Failures of Digital Libraries, 1998.
- [17] C. C. Marshall. Toward an ecology of hypertext annotation. In *Proceedings of HyperText 98*, 1998.
- [18] C. C. Marshall, M. N. Price, G. Golovchinsky, and B. N. Schilit. Introducing a digital library reading appliance into a reading group. In *Proceedings of Digital Libraries 1999*, 1999.
- [19] C. Neuwirth, D. Kaufer, R. Chandhok, and J. Morris. Issues in the design of computer support for co-authoring and commenting. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1990.
- [20] P. Nokelainen, J. Kurhila, M. Miettinen, P. Floreen, and H. Tirri. Evaluating the role of a shared document-based annotation tool in learner-centered collaborative learning. In *Proceedings of IEEE ICALT*, 2003.
- [21] K. O'Hara and A. Sellen. A comparison of reading paper and on-line documents. In *Proceedings of CHI 1997*, 1997.
- [22] K. O'Hara, F. Smith, W. Newman, and A. Sellen. Student readers' use of library documents: Implications for library technologies. In *Proceedings of CHI 1998*, 1998.
- [23] T. Ostler. Textlight: Towards an intelligent text browser. Master's thesis, The University of Westminster, 1998.
- [24] Content user profile: Update on scientists. *InfoAboutInfo Briefing*, Outsell Inc., 7, April 30 2004.
- [25] F. Shipman, M. Price, C. C. Marshall, and G. Golovchinsky. Identifying useful passages in documents based on annotation patterns. In *Proceedings of ECDL 2003*, 2003.
- [26] C.-Y. Wang and G.-D. Chen. Extending e-books with annotation, online support and assessment mechanisms to increase efficiency of learning. In *Proceedings of ITiCSE '04*, 2004.
- [27] A. Wexelblat and P. Maes. Footprints: History-rich tools for information foraging. In *CHI*, pages 270–277, 1999.
- [28] J. L. Wolfe. Effects of annotations on student readers and writers. In *Proceedings of Digital Libraries 2000*, 2000.
- [29] Zotero: The next-generation research tool. <http://www.zotero.org/>.