Facilitating Issue Categorization & Analysis in Rulemaking

Thomas L. Bruce
Cornell e-Rulemaking Initiative
Legal Information Institute

Cornell University
Ithaca, NY 14853
tom.bruce@cornell.edu

Claire Cardie
Cornell e-Rulemaking Initiative
Computing & Information Science
Cornell University
Ithaca, NY 14853
cardie@cornell.edu

Cynthia R. Farina
Cornell e-Rulemaking Initiative
Cornell Law School

Cornell University
Ithaca NY 14853
crf7@cornell.edu

Stephen Purpura
Cornell e-Rulemaking Initiative
Information Science Program
Cornell University
Ithaca NY 14853
sp559@cornell.edu

ABSTRACT

One task common to all notice-and-comment rulemaking is identifying substantive claims and arguments made in the comments by stakeholders and other members of the public. Extracting and summarizing this material may be helpful to internal decisionmaking; to produce the legally required public explanation of the final rule, it is essential. When comments are lengthy or numerous, natural language processing and machine learning techniques can help the rulewriter work more quickly and comprehensively. Even when a smaller volume of comment material is received, the ability to annotate relevant portions and store information about them in a way that permits retrieval and generation of reports can be useful to the agency, especially over time. We describe a prototype application for these purposes. The Workspace for Issue Categorization and Analysis (WICA) allows the rulewriter to create a list of relevant substantive categories and assign them to marked portions of comment text. She can then retrieve all instances of a given issue within the comment pool. Preliminary results of experiments that apply text categorization and active learning methods to comment sets suggest that these techniques can facilitate the marking and category assignment process in lengthy or numerous comment sets. WICA will incorporate these techniques. Other possible applications of WICA within the rulemaking process are discussed.

Keywords e-rulemaking, text categorization, machine learning, comment management, annotation, notice & comment rulemaking, reply comment

1. INTRODUCTION

Although federal agencies have been exploring the use of information technology in rulemaking since the early 1990s, efforts have focused primarily on making proposed and final rules and relevant background information available on the World Wide Web, and enabling the public to submit comments online. Most agencies still have little technology beyond basic word processing to help rulewriters actually write rules and produce the supporting analyses and justifications required by statute, Executive Orders and judicial decisions.

The area in which agencies have expressed the most immediate need and desire for e-tools is comment management. The fundamental goal of "managing" comments is to allow agency rulewriters to identify pertinent claims, arguments, data, etc., that they contain. [5] Such material may add new information or in some other way change rulewriters' perception of the appropriate final rule. Whether or not this happens, the agency has a legally enforceable obligation [13] to provide a written justification with any final action on the rule. This justification, which typically appears as the Preamble to the published rule, must acknowledge and respond to significant criticisms contained in comments and explain why the agency rejected reasonable alternative approaches they propose. Knowing the relevant content in the comment pool is an essential precondition to writing the sort of Preamble that enables a rule to survive challenge in the courts.

Hence, it is not surprising that when the Cornell e-Rulemaking Initiative (CeRI) asked rulemakers in Department of Transportation agencies what e-tools would most help them, we were told: a way to automate identification of relevant substantive information in comments. Discussion fairly quickly focused on a related functionality: a way to order and quickly and easily incorporate this information into the process of drafting the Preamble.

2. CHALLENGES TO PROVIDING E-TOOLS FOR MANAGING COMMENT CONTENT

2.1 Dimensions of Comment Set Variability

The nature and number of comments received on a proposed rule varies widely from agency to agency, from rule to rule, and sometimes even within a single rulemaking. This variability poses challenges for our machine learning-based text categorization methods.

Rules that apply to a narrow band of regulated activities or entities may generate only a handful of comments. One study of comments received by agencies in the Department of Transportation during two 3-year periods found that about one-third of proposed rules elicited fewer than 10 comments. [1]. At the other extreme, occasional rules that become highly salient –

e.g., the recent Fish & Wildlife Service proposal to list the polar bear as an endangered species; the Federal Communication Commission's media ownership rule – have produced over a million comments.

Numbers don't tell the whole story, though. Some comments – typically from large corporations or trade/professional associations – are long and complex, addressing a range of substantive issues raised by the proposed rule. Even a relatively small number of such comments may contain a large quantity of relevant information. Comments from the general public tend to be short, and more likely to express sentiment and/or a conclusory position about the agency's proposal. If they do address substantive issues, coverage is often cursory. Even a very large number of such comments may contain relatively little relevant information. Both extremes, as well as comments of intermediate length and substantive complexity, will often be found in a single rulemaking.

Finally, a single rulemaking might have two, or even more, sets of comments. An agency may produce an Advanced Notice of Proposed Rulemaking to test the waters and/or get information useful in formulating a proposed rule, followed by a second comment period on the actual proposal. [9] Occasionally, comments or changed circumstances will lead an agency to alter its original proposal so much that another opportunity for public comment is required, or deemed desirable. The number and nature of successive comment sets in the same rulemaking will likely change, although the substantive issues within them may substantially overlap.

2.2 Format Complications

As new communication media have developed, agencies have expanded the ways they will accept comments: mail (conventional, then express), fax, email, and, most recently, web submission. Email and online comments may take the form of text typed directly into the body of the message, or an attached file. The proliferation of ways for submitting comments is generally considered a good thing: commenters can choose whatever each deems the easiest, most congenial method of participation. From the perspective of using information technology to facilitate rulemaking, however, this is far from idyllic. A comment set now can, and often does, contain handwritten, typed and faxed comments, text in email messages and online comment forms, and attached files in a multitude of image and text formats.[1] Some of these formats are less conducive to computer-based management and processing than others. Electronic formats that record the document as an image, rather than as a stream of characters, are particularly intractable. Such formats require conversion, sometimes in multiple, error-prone steps, before content is accessible for application of natural language processing techniques. Rulemakings with large numbers of comments requiring conversion can require considerable effort and expense in this step alone.

3. CURRENT RESEARCH

Despite these challenges, progress is being made in creating technology to help rulemakers more quickly identify relevant substantive information in comments. Researchers at CMU/Pitt have worked on the unique problems presented by very large numbers of e-mail comments that contain identical or very similar text. [14, 15] Such e-postcard campaigns, generated by interest groups, occur in a very small percentage of rulemakings, but can impose extraordinary costs on the agency when they do happen. Using information retrieval techniques, the CMU/Pitt tool isolates and tabulates the non-duplicate portions of these emails. This ensures that rulemakers will actually see the ways individual submitters have customized the form comment. This research is so advanced that the tool was used by the Fish & Wildlife Service to process e-postcard comments in the polar bear endangered listing rulemaking. The researchers are also investigating the use of natural language processing methods to identify the main claims of a comment, and determine whether they support or oppose the proposed rule or present a new idea. [6,7] This work also seeks to categorize comments according to a small set of general topics that recur in rulemaking, e.g., economic, environment, health, legal, policy.

A Stanford research group [8] has also investigated the use of information-processing techniques to organize public comments according to the rule provision(s) that each addresses. They employ a combination of manual and automatic methods to identify a set of predefined "features" in the proposed rule and associated comments. These include mentions of rule-specific concepts, definitions, and measurements, or explicit references to provisions in the rule. After representing each comment and provision in terms of such a feature vector, standard text retrieval methods are used (e.g., calculation of the cosine of the vectors) to identify comments and provisions that appear to be related (i.e., that look similar in terms of their vector representations).

In this paper, we describe a CeRI project to create an electronic workspace that helps rulewriters identify relevant substantive claims and arguments, and organize this material in a form that facilitates both analysis of the comment pool and, ultimately, preparation of the justification that must accompany final agency action. The Workspace for Issue Categorization & Analysis (WICA) will incorporate text categorization and active learning components [2,10] that can make a provisional extraction and sorting of relevant material; alternatively, the rulewriter can manually extract the relevant information and use only the organization and management components of WICA.[2]

[1] Regulations.gov, the federal government's notice-and-comment portal, tells commenters they may attach files in any of the following formats: PDF, BMP Image (Windows, OS/2), Excel Template/Work Book, GIF Image, HTML Document·, JPEG Image, PC Paint Image (Windows), Power Point 4.x (O.S), Rich Text Format (RTF), SGML Text, TIFF Image, Text Document, Word Document

[2] Commercial applications for managing comment excerpts and incorporating them into internal and external response documents are emerging. E.g. Commentworks; Tetra Tech. As a student-built prototyype using off-the-rack open-source components, WICA won't compete with software designed at substantial cost by professional developers. It has a purpose that commercial applications do not: To support research into the application of a potentially wide variety of natural language processing and

4. THE WORKSPACE FOR ISSUE CATEGORIZATION & ANALYSIS (WICA)

4.1 Basic Structure and Operation

The WICA system is composed of several software subsystems, all of which store information in the database back-end. A comment-ingestion subsystem pulls comments into the database. A category-management subsystem permits the rulewriter to define a category set that is, in effect, a taxonomy of relevant information she anticipates will appear in the comments. (These could be substantive topics, rule subsections, characteristics of commenters, etc.) Our conversations with rulewriters confirm that the anticipated category set will often need to be modified as comments are reviewed, consolidating existing categories or adding new ones. Accordingly, once constructed, the category set can be adjusted via the administrative interface in a way that sensibly preserves any work done to that point. The ability to define metadata “flags” and associate them with specific comments can help trace the path that lead to category modification, so that any necessary backtracking to apply adjustments is more efficient.

An annotation interface allows the user to associate passages of text with particular categories by highlighting the passage and then clicking on the desired category. (See Figure 1) Multiple categories can be assigned to a portion of comment text. The “swipe and click” method of annotating comment text was chosen because it is a familiar operation for most users. The text of the comment is never actually modified. Rather, the association of a segment of text with a particular category or categories is stored in the database. The user sees annotated text as highlighted, and hovering the cursor reveals the category(ies) assigned to it. (See Figure 2)

The user can also attach flags – effectively, electronic sticky notes – to comments. Flags can be pre-defined (e.g., “Workgroup must see”) or created by the user on the fly. (e.g., “Check with Tom re par 2”). Supervisory viewing modes reveal and compare annotations from multiple users on a particular comment.[3]

Through a searching subsystem, the user can retrieve comment text by category, flag, full-text search, or a combination. A variety of reports can be produced from the database, either internally or via external software.[4] Security is maintained via a readily-modified (and potentially fine-grained) system provided by the content-management layer.

From the engineer's perspective, WICA is a Web-based application built using “off-the-rack”, standards-based software and well-understood programming techniques. User interaction with the system takes place via an easily modified, AJAX-based interface built within the Drupal content-management architecture; the database back-end is the widely used mySQL open-source RDBMS.

A few implications of our design choices are worth underscoring. First, WICA can be accessed from anywhere, and users interact with it as they would with any web-based application or word-processor. Thus, training is minimal and deployment is easy – an important quality for field-testing experimental tools. Second, use of an independent content-management architecture makes it simple and fast to create a rich but controlled working environment around WICA. Features like fine-grained user management and authentication, wiki-based documentation, and discussion and user-support forums are available "for free" from the hosting content-management system. Finally, use of an off-the-shelf relational-database that can be accessed independently of the category-management subsystem has two important consequences. First, external applications – including commonly-used office applications and e-tools specifically designed for comment management – can readily make use of the comments and/or their associated categories, across one or many comment sets via direct interaction with the database.[5] Second, such applications can be built by anyone familiar with common database-programming techniques.

4.2 The Text Categorization and Active Learning Dimension

In notice-and-comment proceedings that generate a small volume of comment material, it would not be difficult for rulewriters to use the annotation interface to manually assign categories throughout the entire body of comment text. As volume of comment material increases, however, the assistance of natural language processing techniques becomes desirable to expedite identification and categorization of relevant information.

We are exploring the use of both text categorization and active learning methods. Text categorization is the process of building, by means of machine learning techniques, systems capable of automatically assigning text to one or more categories from a predefined set. In the rulemaking context, a human reader creates a “training set” by annotating a portion of comment material, from which the text categorization system “learns” what type of information to associate with each category. [11] Active learning methods aim to reduce the size of the required training set without sacrificing accuracy in automated categorization. [3, 4] Essentially, they identify the kinds of additional training examples that will be most useful to the machine learning process.

Working with comment sets annotated by law students, our research thus far has achieved overall categorization accuracy rates in the low 60-percent range. [2,10]. That is, in 60-65% of approximately 1100-1600 sentences used for these experiments, the text categorization (TC) system assigned the same

machine learning techniques to rulemaking, both in the research setting and in field testing by agencies.

[3] During our current research on text categorization, this function is used for analyzing instances of interannotator disagreement.

[4] These currently include interannotator agreement and other research reports.

[5] We presently use this capability to access the categorization data for natural language processing research.

category(ies) as the law student annotators.[6] Higher accuracies can presumably be achieved when all 11,100 sentences in this comment set are used.[7] Using active learning (AL) techniques, the number of sentences in the required training set can be reduced by about 50% and still maintain these accuracy rates. In concrete terms, human annotation of the 400-600 most useful sentences selected via active learning from the available 11,100 sentences can achieve the 60-65% accuracy rate. (This is in contrast to the 1100-1600 manually annotated sentences required to reach this level of accuracy without active learning.)

In the context of public comment analysis, some types of automatic categorization errors are more serious than others. An error analysis of the categorization results described above reveals that an average of 71.6%[8] of errors made by the system correspond to sentences that it did not categorize at all, when the human annotator assigned them to one or more issue categories. This "underinclusive" type of error is especially problematic: material that the rulemaker should consider has not been identified by the text categorization component. A second category of error occurs when both the system and the human annotator categorize the sentence, but not with the same category. These "wrong category" errors are less costly than the first type: the rulemaker will carefully focus on all text assigned to all categories in any event; the principal cost of these errors is the time the rulemaker spends adjusting the category assignment. We determined that an average of 21.4% of all errors are "wrong category" errors. A third type of error occurs when the machine categorizes a sentence that the human annotator left unmarked. In these "overinclusive" errors, the rulemaker will be focusing on text that is not valuable to her, and will then have to delete the annotation to remove it from the category folder. "Overinclusive" error account for an average of 5.0% of the categorization mistakes. Happily, it is possible to train the text categorization component to prefer one type of error over another: In an initial experiment, manipulating the misclassification cost function for the text categorization system reduces the most costly errors ('underincluive" errors) by 19%. The tradeoff is an increase in the total number of errors by 4% (i.e "wrong category" and "overinclusive" errors increase).

Further experimentation is expected to improve the accuracy rate and shift the errors away from the most costly type. Still, the achievable accuracy rate is not likely to be high enough to allow the rulewriter to ignore material left uncategorized. Hence, we do not suggest that rulewriters using the TC/AL subsystem of WICA would read only comment material assigned to one of the category folders. The system allows the rulewriter to call up all unassigned comment text, so that she can check it for significant relevant material that might have been missed. We anticipate this would be a skimming process less time-consuming than unguided comment reading, but confirmation of this assumption must await field testing.

The TC/AL subsystem has not yet been integrated into WICA, but we anticipate an improved, in-box-like workflow interface that supports prompting from the AL element in the form of a request that the user annotate a particular comment or comments. When the TC/AL subsytem has completed categorization of all the comment material, the rulewriter will be able to call up and read all comment text dealing with each category. The machine-assigned categorization can be modified during this process – the category changed, additional categories assigned, or all categories removed – if the rulewriter disagrees with the machine-categorization. Finally, all comment text not assigned to any category can be called up, skimmed, and to the extent appropriate, categorized.

The result of this process will be a database for the comment set that allows the comments to be examined from a number of perspectives. Combinations of category, flag and full-text searching will enable agency analysts to discover patterns in, and characteristics of, the comment set that would be difficult to discern from simply reading the comments and taking notes. Writing the Preamble for the next step in the rulemaking will be easier. All comment text on a particular point is readily viewable in an optimally organized way, and can be cut and pasted into whatever electronic writing tool the rulewriter favors.

4.3 Other Rulemaking Applications

Although our research focus has been on helping rulewriters understand and respond to public comments, this work has broader implications for rulemaking.

It could support a broader, more effective practice of soliciting reply comments. In theory, the comment period allows stakeholders and other members of the public not only to react to the agency's proposal, but also to address objections and suggestions made by other commenters. In fact, the sort of dialogic commenting that could really test and augment claims, arguments and ideas rarely occurs. The most detailed and extensive comments tend to be filed at the end of the comment period. Major commenters behave strategically, waiting to see what others say so that theirs is the last, most comprehensive word. Hence, robust responsive commenting requires a separate reply comment period – which many agencies resist as adding yet more time to an already lengthy process, for questionable benefits. Even when a brief reply comment period is provided, would be-commenters have the same problems as rulewriters in identifying relevant information in a large body of comment text. If, however, rulewriters could rapidly process comment text, they could provide the public with access to a database of comment material that is categorized and searchable. The result may be improved quality of reply comments. This, plus the ease with which the second set of comments can be processed, may encourage more agencies to implement a reply comment period as standard practice.

In addition, machine-facilitated content identification and categorization may help in aspects of the rulemaking process other than the public comment phase. The agency is often

[6] Within the text categorization system, NONE – i.e., text fits none of the defined categories – is treated as itself a category.

[7] This set is the FTA Grant Circulars Corpus, a combination of two successive comments sets on a Federal Transit Authority proposal. The 267 comments in the set ranged in length from 1 sentence to 1420. [2]

[8] The range is between 33% and 83%, depending on the dataset and whether the system is asked to categorize sentences according to coarse- or fine-grained issues.

legally responsible for providing a number of analyses and impact statements for a new rule. When these require review, extraction and organizing of relevant material from large amounts of scientific, technical and/or economic text, WICA can support and assist with this process.

5. RESEARCH AND POLICY ISSUES

5.1. Areas for Inquiry

We have explained elsewhere why rulemaking comments raise a number of non-standard and difficult issues for text categorization. [2] In addition to continued experimentation with TC algorithms to improve accuracy, a number of other research questions are presented.

- *Agency Annotation Behavior*: Although rulewriters now "extract" relevant information from comments – and some even use issue matrices to record information as comments are read – this will typically not be the kind of exhaustive marking being done by research annotators to produce the training set. On the other hand, rulewriters have domain knowledge (shared by at least some commenters) that research annotators inevitably lack; hence their identification of relevant material is likely to be more consistent. Either or both of these factors may impact performance. We are beginning to explore these questions with the assistance of rulewriters in the Federal Aviation Administration, the Federal Transit Authority (DOT), the Office of Civil Rights (DOT) and the Bureau of Industry & Security (Commerce), who have agreed to use WICA to annotate comment sets in parallel with research annotators.
- *Re-use of Text Categorization Models Across Rules*: As more annotated comment sets are created and categorized, it will be possible to explore the use of earlier categorization experience for training text categorization components for new comment sets. The most obvious circumstance for this type of *inductive transfer* [12] is successive comment sets in the same rulemaking. Beyond this, however, agencies often undertake several regulatory actions in a substantive area, and it will be important to learn when and how earlier experience can be applied to speed training in related contexts.
- *Exploring Tools for Discovering Unexpected Issues*: Our conversations with rulewriters suggest that comments do sometimes make claims or arguments that the rulewriter did not anticipate, but considers relevant. The current system relies upon the human reader to add categories. Particularly as ML techniques reduce the volume of comment material the rulewriter must initially read and annotate to create the training set, it will be important to investigate techniques that can identify potentially relevant material that is otherwise unannotated and, in effect, propose new categories to the rulewriter.
- *Structuring Comment Input*: Some agencies (e.g., Fisheries Service of the National Oceanic & Atmospheric Administration (Commerce)) have experimented with attempting to channel comment content by asking specific questions in the proposed rule. Structured comment input could significantly aid categorization. Experimentation in comment solicitation techniques, especially web-based formats, is needed to discover the most effective ways to encourage commenters to provide information in a manner that supports rapid and accurate automatic categorization.
- *Synergy with Other e-Tools*: WICA is designed to facilitate incorporation of other applications. The many different kinds of material reviewed and generated during rulemaking, and the potentially different types of information agencies may seek to extract, suggests the possible value of combining multiple approaches. An obvious area of interest is automatic summarization of comment material, once it is categorized.

5.2 Best Practices for Better Performance

Obtaining better performance of comment management e-tools is not just a problem for researchers. Agency behavior can affect how readily, and effectively, natural language processing and other information-processing techniques can support rulewriters. One step that would substantially facilitate research, and ultimately application, of e-tools would be to steer comment submission toward formats readily processed by machine. Paper-bound and image-based document formats should be accepted only in exceptional circumstances. To be sure, the comment process should be open to the broadest range of participants. But a large proportion of current problems in preparing comments for machine processing are a product of habit, not necessity. Comments that are hand-written or composed on a mechanical typewriter are a vanishingly small percentage of submissions in most rulemakings. And nearly all image-based electronic files are no more than a package used for the shipment of documents that originated in much more tractable word-processing formats. Substantial progress would be made simply by (1) no longer accepting faxed comments (online submission permits equally rapid transmission); (2) requiring submission of the file, as well as hard copy, of any document created with a word processor; and (3) limiting acceptable file formats to those readily converted to character-based text encodings.

Additionally, rulewriters could assist categorization by drafting practices – in the proposed rule itself and the accompanying notice and request for comment – that encourage self-categorization by commenters. Dividing the proposed rule into easily-referenced subsections, and using the same name/ numbering system to structure the accompanying explanation provides a set of organizational and structural cues that at least some commenters (especially those submitting long, multi-issue comments) will heed.

6. CONCLUSION

We have presented early results in a project to create an electronic workspace for managing content in public comments received in rulemaking. WICA is a web-based system that is handily deployed, highly interactive, flexible in its user interface, and easily extensible. It can satisfy the need of agency personnel for a working environment that supports extraction, categorization and organization of comments for purposes of analysis and construction of legally-mandated responses.

7. ACKNOWLEDGMENTS

Thanks to officials in the Department of Transportation, particularly Neil Eisner and rulemakers in the Federal Aviation Administration, Federal Transit Administration and Office of

Civil Rights; to rulemakers in the Bureau of Industry & Standards of the Department of Commerce; and to Dan Cohen.
This work was sponsored, in part, by a grant from the National Science Foundation: IIS-0535099.

8. REFERENCES

[1] S. Balla and B. Daniels, Information technology and public commenting on agency regulations, *Regulation & Governance* 1:46-67, 2007.

[2] C. Cardie, C. Farina, A. Aijaz, M. Rawding and S. Purpura, A Study in Rule-Specific Issue Categorization for e-Rulemaking. In *Proceedings of the Ninth Annual International Conference on Digital Government Research*, 2008.

[3] D. Cohn, L. Atlas, and R. Ladner. Improving generalization with active learning. *Machine Learning*, 15(2):201–221, 1994.

[4]Y. Freund, H. S. Seung, E. Shamir, and N. Tishby. Selective sampling using the query by committee algorithm. *Machine Learning*, 28:133–168, 1997.

[5] C. Kerwin, *Rulemaking: How Government Agencies Write Law and Make Policy*, page 111, 3rd edition, 2003.

[6] N. Kwon and E. Hovy, Information acquisition using multiple classifications. In *Proceedings of the Fourth International Conference on Knowledge Capture (K-CAP 2007)* 2007.

[7] N. Kwon, E.Hovy and S. Shulman. Multidimensional text analysis for erulemaking. In *Proceedings of the Seventh Annual International Conference on Digital Research, 2*006.

[8] G. Lau, K. Law and G. Wierderhold, A relatedness analysis tool for comparing drafted regulations and associated public comments, *I/S: A Journal of Law and Policy* 1:95 -110, 2005.

[9] J. Lubbers, *A Guide to Federal Agency Rulemaking*, pages 210, 284, 4th edition, 2006.

[10] S. Purpura, C. Cardie, J. Simons, Active Learning for e-Rulemaking: Public Comment Categorization. ,In *Proceedings of the Ninth Annual International Conference on Digital Government Research*, 2008.

[11] Sebastiani, F. Machine learning in automated text categorization. In *ACM Computing Surveys*, 34:1-47, 2002

[12] D. Silver, G. Bakir, K. Bennett, R. Caruana, M. Pontil, S. Russell and P. Tadepalli. Inductive Transfer : 10 Years Later. *NIPS 2005 Workshop,* 2005.

[13]P. Strauss, T. Rakoff and C. Farina, *Administrative Law*, pages 510-550, 10th edition, 2003.

[14] H. Yang and J. Callan. Near-duplicate detection for erulemaking. In *Proceedings of the Fifth National Conference on Digital Government Research*, 2005.

[15] H. Yang and J. Callan. Near-duplicate detection by instance-level constrained clustering. In *Proceedings of the Twenty-Ninth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2006

Figure 1. WICA system showing category list and annotation pane

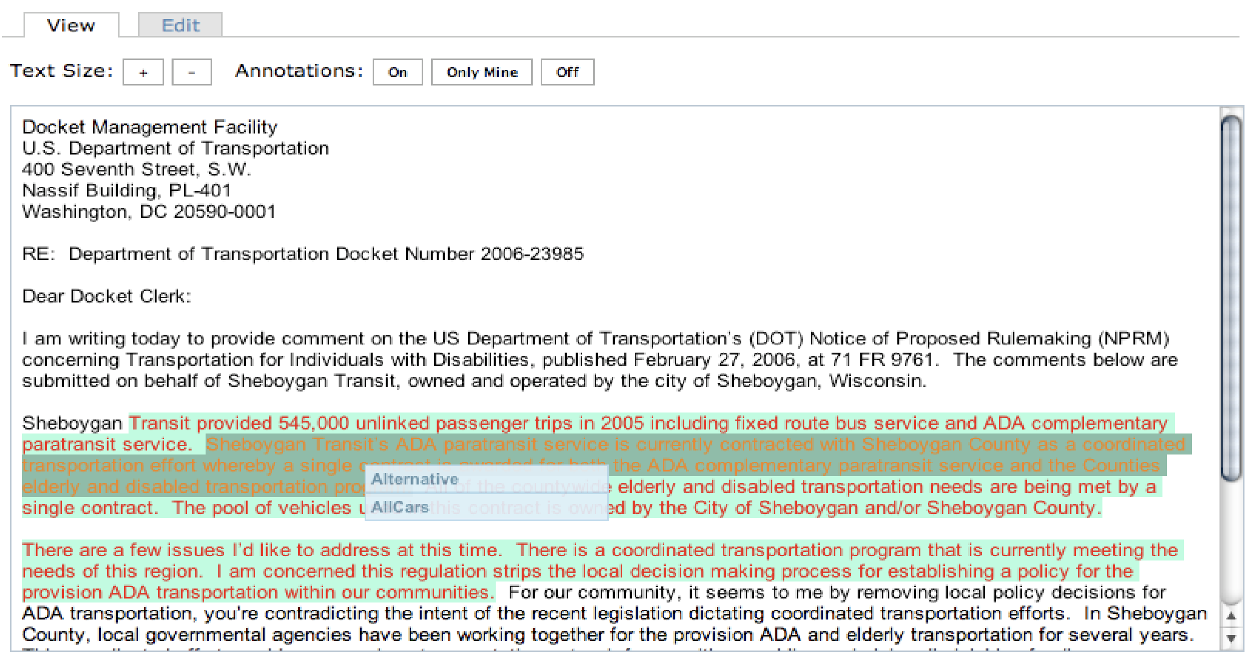

Figure2. WICA annotation screen showing overlapping passages and tooltip indicating categories