MEASURING THE USE OF KNOWLEDGE IN POLICY DEVELOPMENT

Antje Witting
University of Konstanz

Abstract: Public hearings are frequently used on all levels of government to systematically collect and analyze information in the early stages of legislative policymaking. The methods currently employed measure knowledge utilization in this context by means of citation analysis of edited articles and/or reports that summarize the information shared at these meetings. By combining citation analysis and social network analysis, this article develops a methodology that can be used to capture citations in transcripts of public hearings that precede these reports. In order to demonstrate its strengths and weaknesses, the method is utilized to analyze the 2009 hearings that informed the 2010 House of Commons Transport Committee report on developing the capacity of major roads in the United Kingdom to meet the country’s strategic transport needs. The research shows a good degree of consistency between two independent coders who employed this method to distinguish citations from non-citations and classify the data. It is concluded that the method can be utilized to reliably measure knowledge utilization at public hearings, and that it can be employed in conjunction with research that focuses on measuring citations in memos, briefings, articles or reports integrating some of the evidence given at these meetings.

Keywords: knowledge utilization, content analysis, citation analysis, policy analysis

ANTJE WITTING – antje.witting@uni-konstanz.de; Dep. Politics and Public Administration at the University of Konstanz

Central European Journal of Public Policy
Vol. 9 – № 2 – December 2015 – pp 54–71
ISSN 1802-4866
© 2015 Antje Witting
Licensed under Creative Commons Attribution 3.0

INTRODUCTION

Using the knowledge that others share in memos, briefings, articles, reports or directly communicate, legislators develop policies and plans to tackle a particular problem on the policy agenda. This drawing of lessons or transferring of knowledge occurs at all stages and venues of legislative decision making. It can be defined as the process in which relevant “knowledge about policies, administrative arrangements, and institutions etc. in one time and/or place is used in the development of policies, administrative arrangements and institutions in another time and/or place” (Dolowitz & Marsh 1996: 344). This implies that the information has been read, understood and applied in policy development (Rich 1997). The available evidence suggests that in legislative processes, in particular systematic and rigorous policy analysis and advice representing the experience and values of a potential user is used and may facilitate policy learning and policy change in the long term under conditions of reciprocity and trust (Sabatier & Weisteman 1985, Mooney 1991, Caplan 1979, Weiss 1977). It has become evident that legislators primarily rely on in-house policy analysis and advice such as committee reports, but also utilize outside sources such as their constituency and the media (Howlett & Migone 2013, Hir 2005, Jackson-Elmoore 2005, Gray & Lowery 2000). The findings vary depending on the context in which knowledge is utilized (Evans & Weisteadt 2013, Weibie 2008, Shulock 1999), the attributes of the knowledge source (Howlett 2014, Desmarais & Hir 2014, Hir 2005, Adams 2004) as well as the information needs of potential users (Blatter 2015, Jackson-Elmoore et al. 2014). Hence, to understand the effect of knowledge utilization on policy learning and change, we need to better account for whose and what type of knowledge is utilized when and under which conditions (James & Jorgensen 2009).

Recent methodological innovations in research on why the use of specific information varies with changing conditions focus on analyzing the content of edited materials such as committee reports (Ledermann 2014, McAllister et al. 2014, Hir 2005, Shulock 1999). The utilization of knowledge that is orally communicated at public hearings is not captured, although a wide array of knowledge inputs enter the legislative process this way (Weiss 1999). The present article seeks to tackle this gap in the literature in three ways. First, it outlines an approach to reliably measuring citations during public hearings. Second, it evaluates the inter-coder reliability of research employed to measure citations in a 2009 hearing that informed the 2010 ‘Major Road Network’ House committee report on developing the capacity of the major road network in the United Kingdom to meet the country’s strategic transport needs. Lastly, it shows how a citation analysis of public hearings could be integrated in a study that com-
pares the use of specific information across different places and/or points during legislative policymaking.

METHOD

Knowledge use per se is difficult to measure since scholars cannot tell what and where a specific knowledge input is likely to have an effect on a particular policy output. Various methods to trace a particular policy output across various policy venues back to specific inputs are currently available (Rich 1997). Self-report surveys are a commonly employed method to capture what information users acquire and use. Only few studies that published their survey questions entail name-generating questions such as who a person would address for expertise (Rich 2004:195). Here, further lessons can be drawn from research that identifies legislators’ affiliations and coordination patterns with regards to policy learning and change (Ingold 2011, Matti & Sandström 2011, Henry 2011, Ansell et al. 2009). The available evidence has shown that name-generating questions work best to identify relationships that are salient to the respondent and can easily be recalled (Marsden 2011: 380–382). In other words, respondents tend to forget to mention people with whom they less frequently engaged (weak contacts) prior to when the survey was conducted (Marin 2004: 303–305). A few scholars have thus started to combine self-reports with observational methods such as tracing in written documents who participated in a policy venue (McAllister et al. 2014, Rich 2001). Further lessons can be drawn from citation analysis, which can specify “complex interactions among forms of information and types of use” (Rich 1997: 22).

Citation analysis is predominantly used in studies of citations in scientific publications. It has yet to be established to what extent citations in academic literature follow the same motivations as citations in the legislative process. Hence, the extent to which the motives for citing generalize across domains is still an open question. However, a handful of scholars already used citation as an indicator that specific knowledge has been read, understood and is now used to influence and possibly impact policy development. Shulock (1999), for instance, counted citations in committee reports, whereas Desmarais and Hird (2014) identified citations in Regulatory Impact Analysis, a type of document that informs the development of new regulations in the United States. Rich (2001) measured citations in news coverage between 1991 and 1995, but O'Connor and Rapchak (2012) conducted the only study that counted citations in civic discourse. The analysis focused on online political forums for thirteen months. The available studies take into account the numeric features of a cita-

Table 1  
A summary of schemes for coding citations in public comments on a semantic level

<table>
<thead>
<tr>
<th>Orientation</th>
<th>Category</th>
<th>Coding Examples</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cited</td>
<td>Type of source</td>
<td>Report, Study, Other published analysis</td>
<td>Shulock (1999)</td>
</tr>
<tr>
<td>Cited</td>
<td>Type of source</td>
<td>Think tank commentary, Think tank study</td>
<td>Rich (2001)</td>
</tr>
<tr>
<td>Cited</td>
<td>Type of source</td>
<td>Opinion pieces, Clearly biased, Not authoritative</td>
<td>O’Connor &amp; Rapchak (2012)</td>
</tr>
<tr>
<td>Cited</td>
<td>Type of source</td>
<td>Scientific studies, Laws and other legal precedents, Government reports, Newspaper/magazine articles</td>
<td>Desmarais &amp; Hird (2014)</td>
</tr>
</tbody>
</table>

Capturing citations in public comments

The problem of multiple coders recognizing and interpreting citations in public comments has not been a central subject in the literature. As an exception, an investigation of public comments presented by Arguello and his colleagues in 2008 focused on the citation of external references in individual sentences (syntactic level); here in particular on the question: “Given a mention of a specific person or organization, is it being referenced as an external source of information in the sentence?” (Arguello et al. 2008: 51). The unit of classification are references to an external source of information associated to a specific person or organization. Decisions to classify a reference to a specific entity either as a citation or a non-citation are grounded in evidence derived from the sentence that mentions the person or organization (Arguello et al. 2008: 50); and based on previously agreed classification criteria and a set of coding rules (see Table 2).

Their analysis focused on public comments submitted to the U.S. Department of the Interior’s Fish and Wildlife Service’s proposal to list the polar bear as “threatened” under the Endangered Species Act of 1973 (Arguello et al. 2008: 51). The team selected a set of sentences from this source material that contained a noun phrase marked as either a person or an organization (Arguello et al. 2008: 53). Coders were then instructed to code the sentence as ei-
ther citation or non-citation (Arguello et al. 2008: 55). The following sentence is given as an example: The IUCN is predicting a 30 percent reduction in polar bear numbers in the next 45 years. This sentence was selected for analysis because it contains the noun IUCN. In comparison, a sentence that would have instead only included an ambiguous reference to a study would not have been included. Multiple references in a sentence were counted separately (Arguello et al. 2008: 53).

### Table 2  Code Book 2.0: A scheme to recognize citations in public comments on a syntactic level

<table>
<thead>
<tr>
<th>Syntactic Level of Analysis</th>
<th>Orientation</th>
<th>Categories</th>
<th>Coding Example</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cited</td>
<td>Type of authorship</td>
<td>Person associated with the production &amp; distribution of specific information</td>
<td>Arguello (2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Organization associated with the production &amp; distribution of specific information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-citation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Agreement between two coders was evaluated twice for 1,200 sentences in the selected set of 6,000 sentences (Arguello et al. 2008: 55–56). The inter-coder reliability was considered almost perfect in terms of Cohen's Kappa (κ = 0.871), and actually increased (κ = 0.920) when the same test was repeated, since both coders had a chance to learn from their behaviour and adapt in response to the previous test (Arguello et al. 2008: 54). The results suggest that the reliability of their coding scheme to identify references in public comments is high.

The logical next step is to further develop this methodology so that it allows two independent coders to not only reliably capture the information cited in public comments but also to consistently identify the citing actors and the type of source cited.

### Capturing citations at public hearings

Following up on Arguello et al. (2008), the research that is presented in this article adopted a grounded approach to developing a coding scheme from the citation analysis discussed above. The scheme entailing classification criteria

---

2 An initial version of the coding scheme included anecdotal evidence in Category C. It was tested by three independent coders on the content of one randomly selected page. Anecdotal evidence and coding rules is outlined in Table 3. It shows that the analysis on the sentence level focuses on identifying the source of information, while the analysis on the semantic level captures the citing author and the type of source cited.

Three categories are distinguished. Category A (Type of Authorship) is adopted from Arguello et al., 2008 (see Table 1). It describes references based on evidence collected on the syntactic level: a person mentioned in a sentence can be coded "A1" (Person) when they are associated with the production and distribution of specific information, or "A2" (Organization) when, for example, an organization is cited rather than a single person. It is considered a non-citation (A3) when neither applies or when the identity of the cited source is unclear. Specific cue words to help distinguish citations from non-citations are the pronouns "I" or "we" randomly matched with visual or auditory sense verbs (e.g., “see”, “observed”, “heard”, “worked on”, “read”, “published”). Multiple citations in a sentence were counted separately.

Category B describes the citing actor. The available citation analyses have primarily focused on organizations citing specific information (B2) or have not provided any information about the citing actor (B3). However, Rich (2001) shows that the author of a citation (the citer) in a public comment can also be a person associated with the production and distribution of specific information (B1).

Category C includes two types of information. “C1” (Research) follows Shulock (1999) and includes reports, studies and other published analysis. “C2” (Commentary) entails all unpublished analysis such as orally communicated policy advice, as well as commentaries in news stories, opinion pieces and oral testimony given at hearings. The decision not to use a more specific classification system can be justified with research that has shown that public comments often lack cues that could help determine the type of source cited (O’Connor & Rapchak 2012).

---

was removed from the scheme after the test found the inter-coder agreement for this code to be insufficient. The person coding the data presented here was not part of this initial test.
Table 3  Guidelines to comprehensive coding of citations on two levels

<table>
<thead>
<tr>
<th>Orientation</th>
<th>Categories</th>
<th>Values</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syntactic Level of Analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cited</td>
<td>A. Type of authorship</td>
<td>Person associated with the production &amp; distribution of specific information</td>
<td>Arguello et al. (2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Organisation associated with the production &amp; distribution of specific information</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-citation</td>
<td></td>
</tr>
<tr>
<td><strong>Semantic Level of Analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citer</td>
<td>B. Type of citing actor</td>
<td>Person associated with the use of specific information</td>
<td>Rich (2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Organization associated with the use of specific information</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unknown</td>
<td></td>
</tr>
<tr>
<td>Cited</td>
<td>C. Type of source</td>
<td>Research</td>
<td>Shulock (1999)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commentary</td>
<td>Rich (2001)</td>
</tr>
</tbody>
</table>

The collected data in the categories A and B is numeric and can also be analyzed quantitatively, e.g., by counting the number of citations for the citer and cited. The data in category C is categorical and can be used in a descriptive analysis. Following Zhang et al. (2013), this citation data can also be integrated in a network matrix and described in network terms, as demonstrated by the hypothetical example in Figure 1.

The edgelist and matrix in Figure 1 record the total number of instances in which actors in a group of five cite one another. E.g., Actor B cites Actor D once but Actor D does not cite Actor B. Upon this basis, scholars can, for example, identify the in-degree centrality of an actor in the network, which is the relative number of times the author is cited, and compare it to the frequency with which an actor is cited and/or attributes of the cited information (Zhang et al. 2013). Further examples of social network measures that can be employed to describe the features of citation networks can be derived from the available citation analyses of academic texts (cf. Yan 2009, Leydesdorff 2007). Lessons can also be drawn from these articles about the software used for social network analysis and visualization. UCINET 6 is a network analysis software tool used to analyze and map the examples presented in this article.

![Figure 1](image_url)
an actor cites another. Frequency can be visualized by the thickness of the lines connecting the actors (edges), while the size of the symbols representing the actors in a network (nodes) can increase with their in-degree centrality (see Figure 1d for a weighted directed network that also shows the in-degree centrality of individual nodes).

**APPLICATION**

A citation analysis of the hearing that informed the 2010 ‘Major Road Network’ committee report, addressing the need of developing the capacity of major roads in the United Kingdom to meet the country’s strategic transport needs, is used as a means of testing the inter-coder reliability of research that employs the above-outlined approach. In general, the hearing included four sessions conducted between May and July 2009. The meetings were structured as follows: firstly, committee members issued a declaration of interest (e.g., trade union membership); then the committee chair asked witnesses to introduce themselves to committee members and the public as well as issue a short statement summarizing their key arguments; then the chair opened the discussion; and finally the chair closed the discussion with some concluding remarks. All the questions and answers in each session were streamed to a live audience (TV and radio) and taken down under arrangements managed by the Official Report (Hansard). A summary of the evidence given was then reported to the House of Commons, printed, and published with the meeting transcripts online.

The citation analysis of the transcripts attached to the aforementioned report was conducted in three steps. First, citation data was manually collected from the meeting transcripts attached to the report. Sentences in the source material containing cue words such as “I” and “read”, and a noun phrase marking a specific person and/or organization, were highlighted. Where no such information was available or the identity of the cited source was unclear, the sentence was coded as “A3” (non-citation). Each person included in the remaining sentences was given a unique name code and labelled as “A1”. Cited organizations were also given a unique code and labelled as “A2”. Second, the citing actor was identified, using semantic evidence, and given a unique name code. Where one person cited another at the same meeting and in relation to the same argument, only the first citation was recorded. The process was repeated to additionally link organizations to the person citing their evidence. Finally, type of data was coded using the coding scheme outlined in the previous section.

The procedure to identify and classify citations in the text included two independent coders, with each coding the evidence in full. The inter-coder agreement can be considered almost perfect (Cohen’s Kappa, $\kappa = 0.89^3$). Both coders identified a total of 95 sentences with citer-cited interactions. There was no disagreement between the coders about the citing actor, since each citation is clearly stated in the transcript of the hearing. While only 51 of these statements included a reference to a specific person, all citations could be linked to an organization that had published the evidence. One of them is included in the following quote:

[Ralph Smyth:] I would not agree that spending money on roads is the best economic way of dealing with congestion because the evidence shows that more road capacity leads to more traffic. The [DfT]’s cited figure type of source in their Command Paper of July 2008 was an 8–10% increase in traffic per year where there is new capacity (House of Commons 2010, Ev45, Q239).

In the above example, [Ralph Smyth] was identified as the citing author, whereas [DfT] was marked as the cited reference. The type of reference was identified as [figure]. [Ralph Smyth] and [DfT] were assigned their unique IDs. While [Ralph Smyth] was coded “B1” (person), [DfT] was labelled “A2” (organization). The information itself, described as [figure], was coded “C1”, i.e. research (see Table 3 for the coding instructions). The inter-coder agreement was good (Cohen’s Kappa, $\kappa = 0.67$) for the data collected in category C, but could have been better had the coders not consistently disagreed in their coding of evidence given at the hearing. While one coded it as “C1” (research), the other coded it as “C2” (commentary).

The numeric and categorical data collected was recorded in an edgelist, converted to a network matrix and then analyzed and visualized using UCINET 6. Only a small selection of research outputs is presented here to highlight potential ways to analyze and represent the data in a study of citation behaviour. The first example is shown in Figure 2 and outlines a network graph that includes all the organizations that the actor, Ralph Smyth, cited at the 2009 hearing, distinguished by the type of source that was cited. It is called a directed two-mode network, because it links a person who cites the evidence to an organization that made that specific information available. The squares denoting organizations that have published research (C1) are highlighted in grey, while those affiliated with commentaries (C2) are marked in blue. Ralph Smyth, as the person

---

3 The formula used to calculate Cohen’s Kappa: $\kappa = (Pr(a) - Pr(e)) / (1 - Pr(e))$. Non-citations ($n = 2146$) were estimated by counting the number of sentences on one page, multiplying it with the number of relevant pages (adding up to a total of 2262) and subtracting citations ($n = 116$).
citing (B1), is represented as a white node in this citation network. The strength of each edge indicates the total number of times Ralph Smith cited specific information by this organization. For example, four commentaries by members of the Cabinet Office were cited while research by the DfT was only cited once.

**Figure 2** Organizations cited by Ralph Smyth (CPRE) at the 2009 hearing by type of source

The second example (Figure 3) outlines a two-mode network graph from the perspective of the cited. It shows the other seven people at the 2009 hearing (B1) that have cited research (C1) published by DfT (A2) such as Stephen Joseph (Campaign for Better Transport) and Stephen Glaister (Royal Automobile Club). Compared with Mr Glaister who cited DfT research at four instances, Mr Joseph included only two references in his statements. Of a total of eight citations, in only two cases the legislators cited research that the DfT prepared and distributed. Like Ralph Smyth, these two actors also only referred to DfT research once.

**Figure 3** Actors citing DfT research at the 2009 hearing

The two-mode network represented in Figure 4 shows the above citation of DfT research by Ralph Smyth in the context of all citations at the 2009 hearing. Each square in the graph denotes a cited source. The colour of these squares describes the type of source cited. Its in-degree centrality, the total number of times this reference was cited by individuals (denoted as circles) at the 2009 hearing, is reflected in the size of the square.

**Figure 4** Weighted network showing type of source and in-degree for all organizations cited at the 2009 hearing

Some of the collected citations (n = 51) can also be mapped and analyzed as a one-mode network connecting pairs of individuals where at least one cites the other. Such a one-mode network is shown in Figure 5. The people cited by...
Ralph Smyth are represented and it is clearly visible that the relationship is not reciprocal. In other words, he himself has not been cited by either of the people he cites. Contrary to the example given in Figure 1, this network does not include a relationship where two individuals cite each other.

The strength of such a one-mode analysis is its focus on individuals. Where the analyst can identify the user (citer) and provider (cited) of specific information, personal data can eventually be added to the analysis with regard to their academic training, affiliations, and interests, to name a few. Such an analysis goes beyond current research that is solely focused on citing organizations. However, the case at hand also shows that transcripts of public hearings clearly specify the citing actor, but do not always include identifying information about the source cited. Compared to research that measures citations in scientific research, analysts who measure citations in the legislative process will find that the citing actors do not always clearly identify the author(s) of the cited material. In contrast to academic publications, no standardized format is used for referencing in this context (Arguello et al. 2008). This problem can hardly be tackled, which limits our ability to understand knowledge utilization on the micro level and in relation to the socio-cultural context of individual actors (Nicolaisen 2008), policy learning and change (James & Jorgensen 2009). Nevertheless, the case has also shown that the method is suitable to reliably relating specific knowledge inputs to the organizations that have published the material.

When combined with the citation analysis of edited documents, the approach outlined in this article allows tracing information flow from policy output back to knowledge input. The 2010 committee report on the major road network in the United Kingdom and the country's strategic transport needs illustrates this (House of Commons Transport Committee 2010). It was published in March 2010, and acknowledged by legislators (House of Commons 2010) soon after. Following the example of McAllistair et al. (2014), organizational participation in the hearing that preceded this report could be measured by looking at the list of persons that had given evidence at public hearings preceding the report. Like Shulock (1999) we could then analyze citations in the report. In the case of our above example of a citation by Ralph Smyth, the report clearly states with reference to paragraph Q239 in the minutes:

> Ralph Smyth of the Campaign to Protect Rural England argued that road building “would simply lead to more traffic and more congestion” (House of Commons 2010, p. 26).

The matching statement in the attached transcript of the hearing reads:

> [Q239] Ralph Smyth: I would not agree that spending money on roads is the best economic way of dealing with congestion because the evidence shows that more road capacity leads to more traffic. The DfT’s figure in their Command Paper of July 2008 was an 8–10% increase in traffic per year where there is new capacity (House of Commons 2010, Ev45, Q239).

This example highlights how the gap in current research addressing the variations in knowledge utilization could be tackled by taking citations at public hearings into consideration: While Ralph Smyth refers to research presented in a Department for Transport (DfT) 2008 Command Paper, no such reference is included in the paragraph discussing his statement. Research that solely focuses on the reports is likely to miss this case in which knowledge is utilized in one venue to support an argument that then enters another forum (committee report) in the shape of evidence. Thus, when combined with current methods to trace citations in edited materials that reference some of the evidence, scholars can trace the flow of specific information across different venues and identify changes in the citation patterns. This allows researchers to identify whose and what knowledge does or does not transfer across different circumstances.
CONCLUSION

The discussion presented here contributes to the literature on knowledge utilization in legislative processes in three ways: First, it highlights the need to develop methods that allow us to trace the flow of information across various policy-making venues. Second, it outlines available methods to measure the use of knowledge in edited documents; and thirdly it introduces an analytical framework that can be systematically used to identify citations in word-by-word transcripts of public hearings. The article explores the reliability and potential of empirical research that employs this framework to systematically observe citation behaviour in a House of Commons Transport Select Committee 2009 hearing about major roads in the United Kingdom and the country’s strategic transport needs. The inter-coder reliability of the findings was considered almost perfect. It also outlines how this methodological innovation can help address what knowledge is utilized by whom, when and under which conditions.

One possible avenue of research, not yet addressed, is the development of software that can support the analysis of citations at public hearings. Arguello et al. (2008) present evidence that manual human annotation of citations in public comments is more reliable than automatic text analysis; examples of the latter can be found in the named entity recognition (NER) literature. One practical lesson we learned when testing the reliability of the House of Commons Transport Select Committee citation data was that the current approach to recording paragraph references in order to direct others to an in-text citation is more inconvenient and time-consuming than it needs to be. For example, the Discourse Network Analyzer (DNA) is a piece of Java-based software for qualitative category-based content analysis that connects policy specialists who express the same preferences during policy discourse. It allows the coder to manually encode three variables: a) the author of a comment (person or organization), b) its content and c) a dummy variable indicating the actor’s agreement (see www.philipleifeld.de for detailed information). According to Philip Leifeld (July 2013), the software is currently being developed to also identify pairs of individuals who participated in the same event. It seems reasonable to suggest that the DNA software is revised further to help coders link an author of a citation to the reference they are citing and assign a value to this particular relationship.

In conclusion, the approach presented in the context of this paper has been useful in facilitating a reliable analysis of citations in public hearings. The methodology is clear and transparent, and researchers who are interested in reliably and effectively measuring citations in public hearings should have no problems integrating new methods as they become available, and in particular methods to capture and analyse citation network data. The better we understand knowledge utilization in the legislative process, the closer we are to grasping its effect on policy learning and policy change. Thus, further developments that enable scholars to test their causal assumptions concerning the use of specific knowledge inputs in policy development are encouraged.

REFERENCES


Zhang, G., Ding, Y., and Milojevic, S. (2013). Citation Content Analysis (CCA): a framework for syntactic and semantic analysis of citation content.