



Universiteit
Leiden
The Netherlands

The relevance of impact: bibliometric-enhanced legal information retrieval

Wiggers, G.

Citation

Wiggers, G. (2023, March 8). *The relevance of impact: bibliometric-enhanced legal information retrieval*. SIKS Dissertation Series. Retrieved from <https://hdl.handle.net/1887/3570499>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3570499>

Note: To cite this publication please use the final published version (if applicable).

Chapter 3

Citation Metrics

Citation Metrics for Legal Information Retrieval: scholars and practitioners intertwined?

Previously published as: Wiggers, G., Verberne, S., Zwenne, G-J. (2022). Citation Metrics for Legal Information Retrieval: scholars and practitioners intertwined? Legal Information Management 22 (2), pp. 88-103.
Changes have been made to sections 3.2.3, 3.3.2 and 3.5.1.

This paper examines citations in legal documents in the context of bibliometric-enhanced legal information retrieval. It is suggested that users of legal information retrieval systems wish to see both scholarly and non-scholarly information, and legal information retrieval systems are developed to be used by both scholarly and non-scholarly users. Since the use of citations in building arguments plays an important role in the legal domain, bibliometric information (such as citations) is an instrument to enhance legal information retrieval systems. This paper examines, through literature and data analysis, whether a bibliometric-enhanced ranking for legal information retrieval should consider both scholarly and non-scholarly publications, and whether this ranking could serve both user groups, or whether a distinction needs to be made.

Our literature analysis suggests that for legal documents, there is no strict separation between scholarly and non-scholarly documents. There is no clear mark by which the two groups can be separated, and in as far as a distinction can be made, literature shows that both scholars and practitioners (non-scholars) use both types.

We perform a data analysis to analyze this finding for legal information retrieval in practice, using citation and usage data from a legal search engine in the Netherlands. We first create a method to classify legal documents as either scholarly or non-scholarly based on criteria found in the literature. We then semi-automatically analyze a set of seed documents and register by what (type of) documents they are cited. This resulted in a set of 52 cited (seed) documents and 3086 citing documents. Based on the affiliation of users of the search engine, we analyzed the relation between user group and document type.

Our data analysis confirms the literature analysis and shows much cross-citations between scholarly and non-scholarly documents. In addition, we find that scholarly users often open non-scholarly documents and vice versa. Our results suggest that for use in legal information retrieval systems citations in legal documents measure part of a broad scope of impact, or relevance, on the entire legal field. This means that for bibliometric-enhanced ranking in legal information retrieval, both scholarly and non-scholarly documents should be considered. The disregard by both scholarly and non-

scholarly users of the distinction between scholarly and non-scholarly publications also suggests that the affiliation of the user is not likely a suitable factor to differentiate rankings on. The data in combination with literature suggests that a differentiation on user intent might be more suitable.

3.1 Introduction

Bibliometric-enhanced information retrieval (IR) aims to improve IR by using bibliometrics, for example citation metrics. Citation metrics are often associated with the notion of scientific impact; the impact of scholarly publications on other scholars. However, legal bibliometrics, and thereby legal bibliometric-enhanced IR, differs from other research domains in two manners: (1) its strong national ties [19] and (2) the often strong interconnection between research and practice, especially in civil law jurisdictions. In the Dutch legal domain this is demonstrated by the use of Dutch language in legal scholarly output, and by the lack of formal distinction between legal scholarly and practitioner (hereafter called non-scholarly) documents. This lack of distinction suggests that users expect both scholarly and non-scholarly documents to be included in legal IR systems. In turn, the developers of these systems aim to serve both scholars and practitioners as customers.

The ultimate aim of our research is to achieve bibliometric enhancement in legal IR systems; to improve the effectiveness of legal search by using citation metrics as a factor of (impact) relevance in ranking algorithms. But before we can implement such a bibliometric based relevance factor, we have to determine whether both user groups have a common understanding of impact relevance. It is important to know whether both user groups can be served using the same bibliometric-enhanced ranking function, or whether each group requires their own function. In order to determine this, we have to understand the meaning of citations in the legal domain.

In this paper we first discuss the literature addressing citations, with a focus on citations in Dutch legal documents. Next we perform a data analysis, for which we create a rule-based, semi-automatic classification

method to classify a set of 52 seed documents – 10 legal cases and 42 journal articles – into scholarly and non-scholarly publications based on document type, publisher reported intended audience, and author affiliations. We further analyze the 3086 documents that cite our seed documents: for each seed document, we register by what (type of) documents they are cited, using the same classification method as used for the cited documents. In our discussion, we link the data analysis to the literature analysis, and conclude with suggestions for using citation metrics in bibliometric-enhanced ranking for legal IR.

The following research questions are addressed in this paper:

1. Does the literature suggest the use of one bibliometric-enhanced ranking function in legal IR, or should there be separate bibliometric-enhanced ranking functions for legal scholars and legal practitioners?
2. Does a quantitative data analysis of citations in, and usage of, legal documents support the findings from the literature?¹

In answering these questions, we distinguish between the implementation of bibliometric-enhanced ranking in legal IR (should the bibliometric-enhanced ranking function consider citations from both scholarly and non-scholarly documents) and the consequences of the implementation choice (given the implementation, would this bibliometric-enhanced ranking function serve both scholarly and non-scholarly users).

The contributions of this paper are twofold: first, we examine the meaning of citation metrics in legal documents using literature and quantitative data analysis. Second, we show, using literature and data analysis, a possible approach for bibliometric-enhanced ranking for legal information retrieval.

3.2 Literature analysis

For the literature analysis, we start by reviewing the general practice of using citations as a form of impact measurement. Next, we compare Dutch

¹Research questions 3 and 4 in this thesis.

legal citation practices to this general practice. This is followed by a section on the debate on the classification of certain legal documents as ‘scholarly’, to highlight the highly intertwined legal publishing culture in the Netherlands. We conclude our literature analysis with a section on the use of citations in IR, our intended use case.

3.2.1 Citations as a Form of Impact Measurement

Cronin [37] tells us that the first written form of disseminating scholarly knowledge was the letter; two learned people would write to each other to discuss their thoughts and research. Some of these letters were copied by intermediaries for broader distribution. The networks of learned people would sometimes get together, and from this the learned societies grew. In time, these learned societies established journals as a more structured form of communication. From these journals systems like peer review and citations were developed, to ensure the quality of the content and acknowledge the work of others.

The use of citations as a proxy for impact was introduced by Eugene Garfield. Garfield stated that “Since authors refer to previous material to support, illustrate, or elaborate on a particular point, the act of citing is an expression of the importance of the material. The total number of such expressions is about the most objective measure there is of the material’s importance to current research.” [48, p. 23]. De Bellis, referring to the work of Merton, stated that: “Citing, specifically, is the same as peer-reviewing, just on a smaller scale. Hence bibliographic citations are atomic components of the cognitive and reward system of scientific communication.” [41, 86]. De Bellis also stated that “Being cited by other authors is not simply a matter of intellectual lineage. When the score gets high, it is likely that the cited document is exercising an impact on citing sources” [41, p. 32]. And that: “This forward-pushing potential, in turn, is the hallmark of scientific quality” [41, p. 32]. Another description of the meaning of citation measurements comes from Kurtz and Henneken: “The measurement of an individual’s scholarly ability is often made by observing the accumulated actions of individual peer scholars. A peer scholar may

vote to honor an individual, may choose to cite one of an individual's articles, and may choose to read one of an individual's articles." [75, p. 696]. What these authors have in common is that they consider the total number of citations a proxy for the impact of the document *on other scholarly documents and scholars*.²

Beel and Gipp state that "a citation measures impact but not quality in general." [16, p. 440]. Garfield, though a proponent of using citations as a form of impact or 'significance' [47, p. 473] measurement, does note that: "citation frequency reflects a journal's value and the use made of it, but there are undoubtedly highly useful journals that are not cited frequently" [47, p. 476]. "[T]hat does not mean that they are therefore less important or less widely used than journals that are cited more frequently. It merely means that they are written and read primarily for some purpose other than the communication of original research findings." [47, p. 476]. An example he uses is *Scientific American*, a widely read journal that he states readers read to keep up to date.

The question whether citations in the humanities behave like citations in the hard sciences (i.e. provide insight into the impact on other scholars) has been a topic of interest in the past decade. Bonaccorsi et al. have shown that distribution of citations of articles in the social sciences and humanities is similar to the distribution of citations in the hard sciences [19]. Hicks discusses different document types that play an important role in the social sciences and humanities that may not be covered by a citation index [58]: books, national literature and non-scholarly literature. The need to include books in citation indexes has been discussed by Giménez-Toledo et al. [51]. Zuccala and Cornacchia [144] have conducted research as to the methodology by which to include books and the challenges therein.

²Work by e.g. Teufel [127] narrows this down by looking at the words surrounding the citation, to see whether the author cites in a positive or negative manner, but this falls out of the scope of this paper.

3.2.2 Citations in Dutch Legal documents

The topic of legal research and the Dutch legal publishing culture has been extensively described by Stolker [120]. Stolker notes that the legal publishing culture has a strong tradition in book publishing. Even though law journal articles are becoming more important, a perspective of legal documents is not complete without considering books, confirming the statements from Hicks and Giménez-Toledo et al. [58, 51, 144]. Stolker further argues that because law is a national research topic, and documents are often bound, by topic and language, to a national audience, an analysis of such documents should be done on a national level.

Snel [117] states that there are three main reasons for citing in scholarly legal documents: to provide context for the research, to legitimize statements made in the research, and to allow others to check the quality of the research. His article is aimed at scholars, and contains advice for writing sound scholarly articles. But he mentions non-scholarly documents as possible sources of reference [117, p. 255]. To provide societal context for the legal research, he writes that authors may refer to newspaper articles. To legitimize their statements, they may refer to law articles, and legal cases [117, p. 256]. To help navigate readers to more information on the topic, they may refer to overview articles or legal handbooks [116, p. 167-168]. This demonstrates that citing non-scholarly sources is accepted practice in scholarly legal articles.

Van Opijnen [94] and Winkels and colleagues [143, 142, 140, 141] have applied citation analysis to Dutch law and case law, but did not include legal literature, such as journal articles and books. Wirt Soetenhorst presented a proof of concept of a Dutch legal literature citation index in 2017 [118], incorporating all legal articles, making no distinction between scholarly and non-scholarly legal articles. However, a literature search has not returned any information that this citation index has been completed. In his book, Stolker[120] cites several sources [130, 108] that are critical of citation metrics as a form of impact measurement for legal documents, which might explain why a legal citation index has not been created up until this point. However, he focuses exclusively on impact measurement for research

evaluation systems, not for use in IR systems. This in contrast to Garfield [47], who originally focuses on applications in library management and the creating of reading lists for scholars and students. Use for research evaluation is mentioned, but does not appear to be Garfield's original focus.

An example where this distinction – measuring for research evaluation or measuring for IR – becomes visible is document type of the cited and/or citing document. While research evaluation may take the effort and quality into consideration, regardless of the form of the document, citation indexes for IR in the hard sciences (like Garfield's original science citation index) only consider the impact on other scientific articles, as the collection the citation index is used for is limited to those scientific articles.

3.2.3 'Scholarly' Legal Documents

There is debate in the Dutch legal domain about whether a distinction can be made between scholarly and non-scholarly legal documents. Stolker describes three types of legal journals: "journals primarily focusing on the scholarly debate; journals merely focusing on dissemination (notes/annotations and short commentaries); and journals – probably the majority – doing both." [120, p. 257]. Stolker further indicates that law journals, unlike journals in the hard sciences, often do not have external peer review, but are reviewed by the editorial board. The members of this editorial board may be scholars, but may also be practitioners [67]. The Dutch legal journals are also not classified in A-, B- and C- journals, as is done in economics [119, p. 32] and other fields³. This means that, unlike in the hard sciences, there is no immediate mark which indicates which documents are scholarly and which are not that can be used for bibliometric research.

Research by Snel [116] further shows that legal scholars are not always explicit about their methodology and their choice of sources. This means that many publications do not contain a methodology section, and so this feature cannot be used to distinguish between scholarly and non-scholarly

³In Italy it appears that legal journals are distinguished between A class and other classes, see Bonaccorsi et al. citeBonaccorsi. The classification of journals into categories is mentioned to have been conducted by experts.

publications. Snel interviewed a panel of law professors, who indicated that certain approaches are so common, that they do not have to be made explicit. Examples named are using legal historical or grammatical reasoning to interpret laws, using only case law from the supreme court⁴, using the snowball approach to gather literature (rather than describing which database/IR system is used and which queries), and not explaining why non-controversial interpretations from other sources are followed. Only when deviating from one of these standard approaches, the scholar will have to make their methodology explicit.

Krans [74], in his article on the scholarly status of the annotation, indicates that for research evaluation purposes most universities classify an annotation to case law as a practitioner oriented document ('vakpublicatie'). He argues that this is not necessarily so, and that it should be judged based on the content, not the form. The president of the Dutch Supreme Court, Maarten Feteris, divides annotations in four types: (1) summarizations, (2) affirming annotations, (3) annotations that reach a different conclusion based on the same facts, and (4) annotations that shed light on arguments, points of view or consequences that the court did not consider to the full extent and that could lead to a different conclusion [40]. An article by Damen [40] shows anecdotal evidence that annotations can influence courts in later decisions. Krans uses this anecdotal evidence to argue that because of the potentially high quality and impact of annotations, the content could be scholarly [74]. Systematically this could be achieved through a reversal of the burden of evidence, where the author has to demonstrate the quality of the work in order to claim scholarly status [134].

From a research evaluation point of view the impact on judges of this fourth group of annotations could be a valid reason to classify these annotations as scholarly, as the work and quality put into it will not differ much from a journal article. However, if the determining criterion is the aim of furthering of the body of knowledge – the impact on scholars and scholarly documents – the argument that they impact judges and other cases is less persuasive. Judges write case law not for the purpose of furthering the

⁴Thereby not considering case law from lower courts.

scholarly debate, but as side product of the judiciary branch of government.

From a historical point of view it is also interesting to consider how responses to journal articles (*‘Reacties’*) should be classified. In the Dutch legal field, it is not uncommon for scholars to write a short response to a journal article of a peer⁵, in a form which is similar to the historical copied and distributed correspondence described by Cronin [37]. While such a response would constitute dissemination of knowledge and participation in the scholarly debate, the short nature of these responses, often focusing only on a single point from the original article, means it is not usually on the same level of skill and effort as a full journal article.

Snel [116] agrees that there is a lack of guidelines for what constitutes (good) academic legal doctrinal research. Because of this it is hard to make a clear distinction between scholarly legal documents and non-scholarly legal documents. Snel [116] suggests scholars to look at the content, the reputation of the author, the journal/publisher and the incoming citations when determining the reliability of a document. Citing Van Gestel and Vranken, who in turn base themselves on a VSNU⁶ report [134], Snel further indicates three factors to take into consideration: (1) originality, (2) thoroughness and (3) profundity. Originality in this context means that the document has to add something to the current body of knowledge and/or further the academic debate. Profundity is taken to mean “the extent to which the publication should provide a comprehensive answer to the research question through reliance on relevant sources”

The difficulty in separating scholarly and non-scholarly legal publications demonstrates the intertwined nature of legal scholarship and legal practice. Suggesting that impact, as measured through citations, should consider citations from both scholarly and non-scholarly documents. It also suggests that the different contexts for citing as described by Snel’s [117]: context, legitimisation of claims and reproducibility/quality control, may be more indicative of different information needs and corresponding

⁵See, for example the journals *Ars Aequi* and *Nederlands Juristenblad*

⁶VSNU refers to the Association of Cooperating Universities in the Netherlands (*Vereniging van Samenwerkende Nederlandse Universiteiten*), currently called *Universities of the Netherlands* (*Universiteiten van Nederland*).

adjusted rankings than the division between scholarly and non-scholarly legal professionals.

3.2.4 Citations in Information Retrieval

Legal IR has a number of characteristics that distinguish it from other IR domains. One of those aspects is that the same legal IR systems are used by practitioners (lawyers and legal professionals) and scholars. Legal IR systems are therefor both professional and academic search systems. Stolker states that “For the massive number of research results available via the Internet today, researchers need some guidance on both the content and the quality.” [120, p. 243]. In IR, this is referred to as the concept of relevance, which consists of multiple forms or spheres of relevance, of which topical relevance is one [109]. Another characteristic that distinguishes legal IR is a form of relevance called bibliographic relevance, where there is a legal difference between the official government sanctioned version of a document and a reprint of the same document [131, 138]. Impact, as measured through bibliometrics, can also be seen as a form of relevance.

An example of using citations as ranking criterion in academic search, including potential negative effects, is the work by Beel and Gipp [16]. They investigated the role of citations in Google Scholar and found that citations have a significant influence on the ranking, though more so for title searches than for other searches [16, p. 442,444]. It appears that since their research, Google has slightly adapted the algorithm to also include how recently the article has last been cited.⁷ This is most likely done to mitigate the Matthew effect, where highly cited documents, which are likely older to have been able to generate such a high number of citations, remain at the top at the expense of newer documents. By displaying these highly cited documents at the top, they are more likely to be cited, creating a self-reinforcing effect. Beel and Gipp [16] named this Matthew effect as one of two main points of criticism for using citations in ranking algorithms in their paper.

⁷<https://scholar.google.com/intl/en-US/scholar/about.html>

Use of citations in legal IR systems can be seen in, for example, the American legal IR system Westlaw⁸. As Jackson and Al-Kofahi [61] indicate though, the more factors like citations play a role in ranking, the harder it is for a user to understand why certain results appear in certain positions. This appears to be one of the reasons why Dutch legal IR systems have focused on thesauri and synonyms to improve their systems⁹, rather than more complex to explain methods such as Page-Rank.

Furthermore the scale of the Dutch jurisdiction, and thereby the size of Dutch legal IR companies and the datasets they have available to them, do not compare to the US and Westlaw. And Westlaw's techniques cannot simply be copied to other jurisdictions, because of the large difference between common law jurisdictions (like the US and the UK) who focus mainly on case law, and civil law jurisdictions (like the Netherlands and most continental European countries), who focus on legal codes, with case law as an interpretative tool [143].

The above mentioned literature shows that if the aim is to use impact as a factor for legal IR systems, bibliometrics from scholarly and non-scholarly publications should be taken together because (1) scholars cite non-scholarly sources, and non-scholarly sources (e.g. case law) cite scholarly sources, meaning an assessment of impact is incomplete without considering citations from all documents. (2) There is debate on whether distinguishing between scholarly and non-scholarly legal documents is even possible, and on what grounds it could/should be. When users themselves cannot reach agreement on which citations are and aren't a measure of impact for them, it is prohibitively difficult to make this distinction in legal IR systems. Since the collections of legal IR systems contain both scholarly and non-scholarly documents, bibliometric data from both types of documents is available, and can be taken together, to measure a broader form of impact than the scholar-on-scholar impact of traditional citation measures such as those proposed by Garfield [48].

⁸http://lscontent.westlaw.com/images/content/L-355700_West-Search-brochure.pdf

⁹<https://clin28.cls.ru.nl/#abstract-36>

Thus, the answer to the implementation question from the introduction is that citations from all document types should be considered, and that these citations measure not only scholar-on-scholar or practitioner-on-practitioner impact, but a broader form of impact on the legal domain as a whole. Therefore, literature does not appear to give an indication that the bibliometric-enhanced ranking for legal IR should be differentiated.

3.3 Methods

To validate the literature, we create a method to distinguish between scholarly and non-scholarly documents – based on guidance from the literature –, to analyse (1) what types of documents cite each other, and (2) what types of users use what types of documents. Our method is motivated by the discussion in Section 3.2.3, which showed that a generalized distinction is necessary to allow us to quantify the interaction between practitioners and ‘scholarly’ publications and vice-versa, to determine whether a bibliometric-enhanced ranking algorithm could serve both user groups, or whether separate algorithms need to be developed.

For this research, we used data from the Legal Intelligence system. Legal Intelligence is one of two large commercial legal IR systems in the Netherlands, covering all government publications and legal publishers. We collected 52 seed documents from the year 2014 – 10 legal cases and 42 journal articles – and the documents that cite them. For both the seed and citing documents, we extract from the logs what type of document it is (e.g. journal article, case law), the source, the title, the name(s) of the author(s) and what the usage of the document is. Along with assessing whether scholarly and non-scholarly documents cite each others, we analyse which types of documents have usage from users affiliated with a university and users affiliated with other types of organizations.

All document types in the Legal Intelligence system are included in our citation analysis, including blogs and newspaper articles, since we want to validate whether the literature is correct in that Dutch legal scholars cite non-scholarly documents and vice versa.

3.3.1 Document sampling criteria

Bornmann et al. [20, p. 214], citing Boyack [21], have remarked that the distribution of citation counts over documents is skewed.¹⁰ This means that a large portion of documents receive no citations, whereas a small number of documents receive a large number of citations [87]. For that reason, a random selection of documents would not be informative for our study, because the majority of randomly selected document has no or very few citations. We selected the documents for our analysis as follows.

We chose seed documents from the year 2014 because of the time it takes for documents in the social sciences to gather citations [108]. The citing documents were from the period 2014 to August 2019, the most up-to-date data available at the time of the research.

To be able to analyze what types of documents cite, we needed to select documents from 2014 that were likely to have been cited. Based on the assumption that documents that are sought often are also read often, and documents are often read before being cited, we used the 2015 query logs from the Legal Intelligence system. We sorted the queries by frequency of occurrence. We manually went through this list and looked at all queries that are clearly related to a case (journal identifier, ECLI number or party/case name) or journal article (journal identifier or title (more than one word)) published in 2014. Case law and journal articles from other years were skipped, to avoid a citation bias based on time since publication. The documents selected are the first documents in the query list that meet these criteria. The documents selected are shown in Table B in Appendix B.

3.3.2 Document classification

For each of the seed documents, we searched our citation index [136] for documents citing it based on the unique document identifier. These citing documents are not only journal articles, but all documents in the Legal Intelligence system. This includes books, as indicated important by Stalker

¹⁰See also Bonaccorsi et al. [19]

[120]. This resulted in 3086 citing documents.¹¹ For these citing documents we also retrieved the source, the title, the name(s) of the author(s), the document type and the usage.¹²

Our first step is to attempt to categorize these documents into scholarly and non-scholarly documents. To determine the category of documents, we consider three cumulative factors:

1. The intent criterion [37]: the document is written with the intent to further the body of knowledge and/or foster academic debate;
2. Related to this is the originality criterion [134, 130]: the document is not merely repetitive or descriptive, but adds interpretation or recommendations;
3. The thoroughness and profundity criteria [134, 130]:
 - (a) The document is based on more than one source;
 - (b) The document has proper references.

Because of the size of the data-set, it was not possible to manually assess each document. Based on our three categorization factors we looked for proxy factors in the data and settled on document type and author affiliation, further explained below. These two proxy factors are cumulative to ensure the least possible false positives in scholarly documents. To limit the manual work required, we only assess author affiliations of documents that do not have a non-scholarly document type; meaning they have either a scholarly document type or the document type alone is inconclusive as to whether the document is scholarly or non-scholarly and also has to be assessed manually.

¹¹Books are indexed per chapter. This means that if multiple chapters cite the same seed document, each chapter is treated as a separate document for the purpose of this analysis.

¹²Information about the citing documents can be found at <https://github.com/G-Wiggers/Citation-Metrics-for-Legal-Information-Retrieval-Systems>.

1. **Document type** We used the type of a document to assess the intent of the document, as well as the originality and thoroughness and profundity criteria.
2. **Author affiliation** To aid in the assessment of the intent criterion, we considered the affiliation of the author.

Documents are classified as scholarly or non-scholarly based on these two cumulative criteria.

To automate as much of the classification as possible, we developed a Python script using the proxy factors and a set of rules to determine for each of the documents whether it is classified as scholarly or non-scholarly (intended for practitioners). This process is visualized in Figure 3.1.

Classification based on document type

- If the document is a government document or case law, then it is classified as non-scholarly. Because these documents are created as a byproduct of the practice of the legislature, the executive, and the judiciary; they are not written for the advancement of scholarship and fail the intent criterion. This means, for example, that our 10 case law seed documents are all classified as non-scholarly because they are byproducts of the judiciary.
- If the document is a news article or notification of publication (short summaries with references to new books or case law), then it is classified as non-scholarly. These documents fail the intent and originality criteria.
- If the document is an annotation to case law, then it is classified as non-scholarly. Though debatable, because the theory above shows that there is a subgroup of annotations that may be considered scholarly based on quality and originality, most of these documents are not written with the intent to further scholarship but to provide interpretation of a legal decision. For this reason, they are likely to fail the intent criterion.

- If the document is a dissertation, then the intent is considered to be the advancement of scholarship and it is classified as directed towards scholars.
- If the document is a journal article or book, we add manual steps (marked in blue in Figure 3.1. Journal articles and books can have many possible intentions. Therefore, for journals and books we checked the (self-reported) publisher information to find out whether the journal or book in its entirety (on source level) was directed more towards scholars or non-scholars. Every time we encountered a new source to check, we added the outcome to a list, to allow automatic classification of other documents from the same source. We classified a document as non-scholarly if the title or description mentioned things like ‘practical information’ or ‘for practice’. If the publisher information mentioned only scholarly use, it is classified as directed towards scholars. If the publisher information mentioned both, we continued to the next step. If the publisher information mentioned nothing, we considered the source to be non-scholarly.
- If the publisher information of a journal states that it has both scholarly and other articles, we analyzed all documents from that journal in our dataset individually. If the document is an announcement or similar document, then it fails the originality and intent criteria and is considered non-scholarly. If it is an article, we check whether it analyses several cases and/or literature and uses proper references. If it meets these thoroughness and profundity criteria we consider it a scholarly article. In case of uncertainty, the documents are categorized as non-scholarly.

For documents for which the document type is inconclusive, we manually assess the last two steps in the classification schema (marked in blue in figure 3.1; whether the document covers multiple documents, and whether there are sufficient references). This manual last step is done by two legal professionals. To assess the reliability of the manual part of the classification, we calculate the inter-rater agreement in terms of Cohen’s κ [28].

Classification of authors

If a document has a scholarly document type, we analyze the affiliation(s) of the author(s) as follows:

- We check if a document had author information. Not all documents (e.g. journal articles) have author information.
- If author information is available, we retrieve the affiliation of the authors primarily from the author information in the document.
- If the author affiliation was not provided in the document, a Google search is conducted and all affiliations mentioned on the first page of the search results are considered.
- If the affiliation is to the government, the intent of the author is not considered to be the furthering of scholarship, since that is not the main goal of the government. This despite the high/scholarly level of quality of some of these documents.
- If an author has multiple affiliations and one of the affiliations is a university, we classify the document as scholarly.
- If a document has more than one author, we classify the document as scholarly if at least one of the authors is affiliated with a university.

Final document classification based on document type and authors

For both the seed documents and the citing documents, we consider a document to be scholarly when the classification based on document type is scholarly and at least one of the authors is affiliated to a university as analyzed in the author classification. These cumulative conditions were chosen to ensure the least possible false positives in scholarly documents. This is chosen since our aim is to attempt to separate between the purely scientific impact of documents, as measured by citation indexes in the hard sciences, and broader impact on the (practitioners in the) legal field.

3.3.3 Readership

To analyze the reading behavior of scholarly and non-scholarly users (reading scholarly and non-scholarly documents), we separate the document usage by scholarly users (all users affiliated with a university) and the usage by non-scholars (all users not affiliated to a university). To do this, we use the organization ID available in the Legal Intelligence data. This organization ID determines the subscription access for users affiliated to that organization. We received a list of organization IDs associated with universities. We first queried the usage by all users with an organization ID associated with a university (this data includes students as the position of the user in the organization is not included in the data), followed by users affiliated to other organization types (such as government, courts, law firms and corporations).

The usage is measured by the number of times the document is opened (clicks), where the same user can use a document on multiple occasions. The data only reflects online usage in the Legal Intelligence system.¹³

The group of users affiliated to a university is roughly 28% of total users, and is therefore smaller than the group of users not affiliated to a university. It is possible that an author affiliated with a law firm writes a scholarly article, so that a click from a user not affiliated to a university could in fact represent use in a scholarly manner. Especially if the user has multiple affiliations. However, it is not possible to determine the reader's intent from the data. For that reason, clicks from organizations other than universities are considered to be for other purposes than writing scholarly articles.

3.4 Results

The number of documents that underwent the manual last two steps of the classification (marked in blue in figure 3.1) is 311 out of the total 3138

¹³It is possible that users have alternative methods to access information, for example through paper versions of books and journals. We have no reason to assume that this would apply more to one group than to the other.

(3086+52). This means that 90% could be classified automatically and 10% needed manual classification. Of the 311 documents, 253 were assessed by both assessors. 58 documents were assessed by only 1 assessor because the second assessor experienced ‘page not found’ or ‘insufficient access rights’ errors.¹⁴ Cohen’s κ , calculated on 253 documents, is 0.58. This indicates moderate agreement in the application of the classification schema for the most difficult to classify documents. For further analyses, we used the classification of rater 1 in cases where both raters disagreed.

Table C in Appendix C shows the detailed results of the classification of the seed documents. It also quantifies the usage and citations for each seed document. The columns Scholarly citations and Non-Scholarly (N-S) citations show the classification of the citing documents according to the rules described in Section 3.3.2. For our analysis we show the usage by users affiliated to a university (shown as Usage Schol.) and the usage by users affiliated to other organization types (shown as Usage N-S).¹⁵

3.4.1 Citations between documents

To analyze the extent to which documents classified as scholarly and non-scholarly cite each-other, we counted the aggregated numbers of citations between scholarly and non-scholarly documents. Table 3.1 shows the summary of these counts. As expected based on the general theory of citation metrics, using a χ^2 test, we found that there is a significant relationship between the two variables ($\chi^2(1, N = 253) = 22.8, p = < .0001$): citations to scholarly seed papers are more likely to come from scholarly papers than

¹⁴55 errors were access rights errors, 3 were ‘page not found errors’. Of these 58 documents, 22 documents were books, 14 were articles, 13 were case law reprints in student collections, 4 were notifications/summaries, and 5 documents were other types. 48 were classified as non-scholarly, 10 were classified as scholarly.

¹⁵The difference in usage between documents could in part be explained by the access rights system of the IR system. Though the IR system works the same for every user, only results from publications the user has a subscription to are shown in the results list. All government documents are freely accessible to all users, as well as open access documents. It appears that certain journals have a higher subscription rate than others, and that digital availability of books is limited to a small share of the user group.

from non-scholarly papers. Note that this test has an expected frequency of cross-citations based on the data, and the table indicates that the categories are far from exclusive in their respective citations: citations from scholarly to non-scholarly documents make up 92% of the total number of citations from documents classified as scholarly ($138/(12+138)$).

It is also important to note here that there is a strong class imbalance on the data: out of the 52 seed documents, 13 documents were classified as scholarly articles based on the criteria in Section 3.3.2. This is why the χ^2 test is important, even though this test presupposes citations between the two groups of documents exist. The same holds for the distribution of citations over documents, which is highly skewed, as expected based on literature [20, 21, 19, 87]. 1155 of the non-scholarly to non-scholarly citations come from 1 seed document, document 14281373. 14 documents (8 documents classified as non-scholarly oriented, 6 documents classified as scholarly oriented), did not receive any citations.

	Scholarly seed	Non-scholarly seed
Scholarly citing	12	138
Non-Scholarly citing	59	2877

Table 3.1: Results: aggregated citations counts. The columns show the classification of the seed documents, the rows the classification of the citing documents.

3.4.2 Usage of documents

To analyze the usage of both classes of documents by users of the Legal Intelligence system, we counted the aggregated numbers of usage between the types of users and the types of documents. Table 3.2 shows the seed documents and the usage thereof subdivided into users affiliated to a university, and users affiliated to other organization types. Similar to the citation data, using a χ^2 test, we found that there is a significant relationship between

	Scholarly seed	Non-scholarly seed
Scholars	1062	3290
Non-Scholars	560	2577

Table 3.2: Results: usage. The columns show the classification of the seed documents, the rows the classification of the usage based on the company identifier linked to the user account.

the two variables ($\chi^2(1, N = 3086) = 46.1, p = < .0001$): a scholarly paper is more likely to be accessed by a scholar than by a non-scholar. Again, this test has an expected frequency of cross-usage based on the data, and it appears to be quite common for a non-scholar to read a scholarly paper: of all the papers accessed by non-scholars, 18% are scholarly ($560/(560+2577)$). And it is also common for scholars to read non-scholarly documents: 76% of the documents accessed by scholars are non-scholarly ($3290/(1062+3290)$).

3.5 Discussion

The above describes a method to distinguish between scholarly and non-scholarly documents and the results thereof. In this section we will briefly discuss the documents that the two manual classifiers did not agree on, followed by (1) an analysis of what types of documents cite each other, and (2) an analysis of the usage data of these documents, and compare these results with the literature. We conclude this section by discussing the implications of these results on the creation of a bibliometric-enhanced legal IR ranking algorithm.

3.5.1 Inter-rater agreement

With an inter-rater agreement of $\kappa = 0.58$, we find that there is moderate agreement between the two raters. Although we judge this as satisfactory considering that these 311 documents were the most difficult documents to classify, we analyzed the differences in classification between rater 1 and

rater 2 in more detail. We noticed three things. First, the debate about the classification of annotations to case law (see Section 3.2.3) is reflected in the results. Rater 1 strictly adhered to the classification scheme and classified all instances of annotated case law that occurred in the manual classification (e.g. because the publisher information did not identify the document as annotated case law) as non-scholarly. Rater 2 however looked at the content of the annotations, and classified 11 of them as scholarly, with a note stating that the quality of these annotations was such that they could have been published as articles. Second, rater 1 classified 7 documents that were a response to a previously published article as non-scholarly, because of the short length of these documents. Rater 2 classified these as scholarly, with as motivation that they contribute to the scholarly debate. Third, rater 1 classified 3 documents that were reports of conferences of legal experts as non-scholarly, whereas rater 2 classified these as scholarly, again referring to their contribution to the scholarly debate.

The examples mentioned above show that rater 1 focused only on the classification of the work, whereas rater 2 focused on the quality of the work. Rater 2 appeared to have used a reversal of the burden of proof similar to that mentioned by the VSNU [134].

3.5.2 Citations between documents

The analysis of the classified documents – 52 cited (seed) documents and 3086 citing documents – shows a significant relationship ($\chi^2(1, N = 253) = 22.8, p = < .0001$) of scholars citing scholars, in a setting where cross-citation is expected. This level of cross-citation the χ^2 test expects from the data shows that scholarly articles also cite non-scholarly oriented documents, as well as the other way around. When we look at Table 3.1, the largest group by far is non-scholarly documents citing other non-scholarly documents. This is partly caused by document 14281373, a legal case and therefore non-scholarly document, which has 1155 citing non-scholarly documents (See Table C in Appendix C).

However, as mentioned in Section 3.4.1, the dataset is unbalanced, meaning that the group of seed documents classified as scholarly is much

smaller than the group classified as non-scholarly. Furthermore, our chosen classification method has strict criteria before a document can be classified as scholarly to avoid false positives, which may result in false negatives, creating further imbalance.

We see that case law documents¹⁶ are widely read and cited by both non-scholars and scholars. NJB, which is a journal aimed at both legal practitioners (non-scholars) and scholars, also receives citations from both groups.¹⁷ It is interesting to see that the journal *Arbeidsrecht*, which is marketed as a journal for practitioners (non-scholars) receives no citations from either group in this dataset.¹⁸

Document 13627420 attracts a lot of response articles. The article was published in the journal for private law, notaries and registration¹⁹, which according to the website of the publisher contains both scholarly articles and non-scholarly oriented articles.²⁰ The author information in the article indicates that the author, mr. R.J. Abendroth [3], is affiliated to a law firm, with no mention of an affiliation to a university. The article is about the order of securities on a good. It received 60 citations, of which 3 are a direct chain of responses. After the original article, prof.mr. F.E.J. Beekhoven van den Boezem (scholar) writes a direct response ('Reactie') in document 15442271. Abendroth (practitioner) responds to this in document 15442265. In 16492944, mr. K.J. Krzeminski (practitioner) responds to both authors. Though this is just one example, it demonstrates an interaction between non-scholars and scholars. It also shows that the informal letter or 'Reactie', which from a research evaluation point of view may not be equal to a journal article in terms of time investment and academic rigour (as pointed out by rater 1 in section 3.5.1), from a dissemination of knowledge point of view may have just as much impact in the legal debate

¹⁶The documents in Table C below the line.

¹⁷See document ids 14151738, 12987162, 13330606, 12926733, 14177758, 13235698 and 13580788 in Table C

¹⁸See document id's 13002758, 14124128, 12987652, 14124136, 22171998, 13241348, 12882340 and 12660424 in Table C

¹⁹'Weekblad voor Privaatrecht, Notariaat en Registratie'

²⁰<https://www.sdu.nl/shop/weekblad-voor-privaatrecht-notariaat-en-registratie-abonnement.html>

(as pointed out by rater 2 in section 3.5.1).

The work of Snel [117], as discussed in section 3.2.2 shows us multiple reasons why scholarly articles may cite non-scholarly documents, and vice versa. This theoretical research, as shown in Figure 3.2, may explain the cross-citations found in the data. When looking at these reasons, we notice that Snel is not just referring to the use of non-scholarly documents in scholarly documents, or vice versa, but also to the use of one document type in another type of document. Snel [117] suggests that a highly cited case could signify that a novel problem was solved (e.g. the first case that dealt with the question whether a digital item is a good in the sense of property law), or that the court veered from a previous ruling. The high number of citations in articles could mean that the case has sparked a legal debate, and has thereby contributed to the furthering of knowledge (intent criterion). This is an example of a non-scholarly work influencing a scholarly work.

A citation from a journal article in a reference work could signify that the article has a lasting impact, for example because it has a novel contribution to legal scholarship (intent criterion) and is of high quality (thoroughness and profundity criteria). Though the reasons for citing as shown by Snel [117] differ, they are all indications of relevance for the legal domain as a whole.

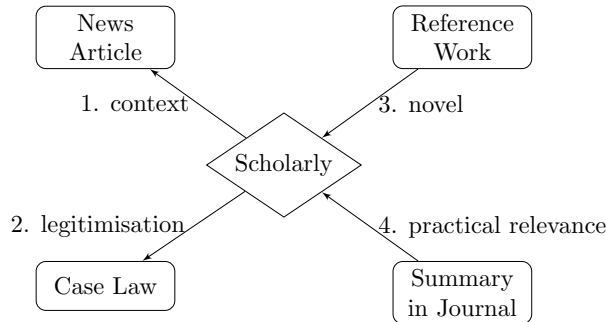


Figure 3.2: Citations in legal documents. 1. Scholarly articles may cite news to give context [117]. 2. Scholarly articles may cite case law to legitimize their claim [116]. If the case is cited often, this may indicate that the court decided a novel problem, or veered from a previous ruling. 3. If a reference work cites a scholarly article, this may indicate that the article had a novel contribution and was of high quality. 4. If the scholarly article is cited in summary in a journal, this may indicate that the article is also relevant for practitioners.

If we were to consider only the impact of scholarly documents on scholarly documents (upper left in Table 3.1), as in citation metrics in the hard sciences, we would miss part of the impact that the scholarly documents have (bottom left in Table 3.1), as well as the impact of non-scholarly documents on the scholarly documents (upper right in Table 3.1). Therefore, for bibliometric-enhanced legal information retrieval, a citation index which does not also look at non-scholarly oriented documents, both what they are cited by and what they cite has an incomplete picture of the legal field. This system of cross-citations also suggests that citations reflect not scholarly impact like in the hard sciences, but part of a broader scope of

impact, or relevance, for the entire legal field.

Given the sizeable number of documents without citation information, as discussed in Section 3.3.1 and visible in Table C, we also looked at usage data from the two user groups, to see whether that shows similar patterns, and whether it could potentially be useful to fill in the gaps in the data for use in legal IR.

3.5.3 Usage of documents

We see in Table 3.2 that even though the group of users affiliated to a university is smaller, their usage is higher than that of the group not affiliated to a university. Our results indicate that there is a relation between usage of scholarly documents and legal scholars, but that legal scholars also read documents classified as non-scholarly, and legal practitioners read documents classified as scholarly.

This finding is supported by the literature (e.g. Snel [117]). An example supporting this is document 12702866. This document is annotated case law and is therefore classified as non-scholarly. But the data shows high usage from users affiliated with a university and a relatively high number of citations by scholars when compared to the other documents in this research. Upon analyzing the document, it appears to be a seven page analysis by Prof.mr. T. Kooijmans [72] on the legal concept of recklessness. This document appears to be one of the annotations that Krans [74] argues should be classified as scholarly due to the high quality, an argument also mentioned by rater 2 (see section 3.5.1). For research evaluation this might be a strong argument to categorize these annotations as scholarly, but that has to be weighed against the intent criterion (see section 3.3.2).

It should be noted that it was not possible in this data set to distinguish between students and employees of the universities, meaning students were classified as scholars. This raises the question whether law students behave more like scholars (contributing to the upper left of Table 3.2) or like practitioners (contributing to the upper right of Table 3.2) in their legal information seeking and needs²¹. It is possible that the high number of usage

²¹The investigation of the search behaviour of students versus legal scholars falls outside

of non-scholarly publications by users affiliated to a university (upper right of Table 3.2 is partly caused by this lack of distinction. This does not, however, explain the high number of usage of scholarly documents by users not affiliated to a university (bottom left of Table 3.2). We therefor consider that while further research is required to determine whether law students have information needs and information seeking behaviour like scholars or like legal practitioners to refine these results, this does not negate the observation that there is usage of non-scholars of scholarly information and usage of non-scholarly information by users affiliated to universities.

It is interesting to note that for case law, in particular documents 14281373 and 12827114, the number of citations for the document is higher than the number of times the document has been opened. We propose two possible explanations: (1) the document has been reprinted in case law journals. Users have read the case in one of these reprint forms, but have decided to cite this version. Both documents are the official government reported versions²² and referenced by the European Case Law Identifier (ECLI). (2) Users access these cases outside of the IR system, for example by going directly to the government website publishing the cases or from a print subscription. Whatever the cause may be, this discrepancy between the usage and citations suggests that the usage data does not provide a complete picture of the readership of a document.

3.5.4 Using bibliometrics in legal IR

The data shows that legal scholars and legal professionals use the same legal IR systems and (at least to some extent) the same documents. This suggests that creating a separation between scholarly and non-scholarly documents in legal IR systems may not be useful for the users. Legal professionals open the most useful or relevant documents for their information need, regardless of the document type and/or the intended audience of the publisher. The bi-directionality of the disregard of scholarly and non-scholarly users to the distinction between scholarly and non-scholarly publications also suggests

the scope of the data set and thereby of this paper

²²As published on the government website www.rechtspraak.nl

that the affiliation of the user is not likely a suitable factor to differentiate rankings on. The data, in combination with the work of Snel [117] suggests that a differentiation on user intent might be more suitable.

The consequence of using citations from all document types in a citation index is that we move from a pure scholarly citation index and the theory behind that, so that the meaning of a citation might also differ. The citation data in this research shows that when we look solely at scholarly impact (scholarly to scholarly citations), we miss part of the picture of the total impact the document has. Similarly, we miss the impact non-scholarly documents have on the scholarly debate. When we combine data from all documents for bibliometric-enhanced legal IR, we are looking at impact on the legal field as a whole rather than solely scholarly impact. To measure this broader impact, citations alone may not provide enough information, since not all documents are cited, and for those that are cited, we only capture impact on authors (scholarly and non-scholarly). We are therefore looking at a part of the impact on the entire legal field.

To enrich the view on this impact on the legal field as a whole a combination with usage metrics appears to be an obvious combination, though it has to be kept in mind that the usage data of legal IR systems may not offer a complete view of usage of legal information, as shown in section 3.5.3. It will however, further fill in the picture of the impact of a document on the legal field as a whole. When implementing usage into a bibliometric-enhanced ranking, usage from both users affiliated to a university and users affiliated to other organization types should be considered, to reflect this impact on the legal field as a whole.

3.6 Conclusion

In this paper we addressed two research questions to try and determine how bibliometric-enhanced ranking can be introduced in legal information retrieval:

(1) Does the literature suggest the use of one bibliometric-enhanced ranking function in legal IR, or should there be sep-

arate bibliometric-enhanced ranking functions for legal scholars and legal practitioners?

The literature discussed in Section 3.2 shows that if the aim is to use impact as a factor for legal IR systems, bibliometrics from both scholarly and non-scholarly publications should be taken together because (1) legal scholarly articles use non-scholarly documents to support their claim, and in turn are mentioned in non-scholarly documents, meaning an assessment of impact is incomplete without considering citations from all documents; (2) There is debate on whether distinguishing between scholarly and non-scholarly legal documents is even possible, and on what grounds it could/should be. When users themselves cannot reach agreement on which citations are and aren't a measure of impact for them, it is prohibitively difficult to make this distinction in legal IR systems. Since both scholars and legal professionals access the same sources and use the same legal IR systems, bibliometric data from both users groups is available, and can be taken together.

Thus, when using citations from both scholarly and non-scholarly publications, we measure a broader form of impact than the scholar-on-scholar impact of traditional citation measures such as those proposed by Garfield [48], and will measure part of a broader form of impact on the legal domain as a whole.

(2) Does a quantitative data analysis of citations in legal documents support the findings from the literature? To validate the findings of our literature analysis we created a classification schema for scholarly and non-scholarly documents based on three cumulative criteria: intent, originality, and thoroughness and profundity. Most of the documents were classified based on rules, the 311 remaining documents were classified with manual steps.

The results of the citation analysis on 52 seed documents and 3086 citing documents confirm that scholarly articles cite non-scholarly documents and vice versa. The usage data shows that users affiliated to a university use both scholarly and non-scholarly documents, as well as users affiliated to other organization types. This is in line with our findings from the literature, and suggests that citations in legal documents do not measure

impact on scholarly documents and scholars in the same way as in the hard sciences, but quantify part of a broader scope of impact, or relevance, for the entire legal field.

This disregard by both scholarly and non-scholarly users of the distinction between scholarly and non-scholarly publications, and especially the fact that this occurs in both directions, also suggests that the affiliation of the user is not likely a suitable factor to differentiate rankings on. The data in combination with literature suggests that a differentiation on user intent might be more suitable. Further research focusing on differentiating queries into the user intents defined by Snel [117] will show whether a differentiated bibliometric-enhanced boost on these grounds will be possible.

Because of the modest sample size used in this research, the results of this paper cannot be extrapolated to all documents in the Dutch legal domain. Because of the national characteristics of legal domains, this example from the Netherlands can also not be extrapolated to other countries. The results do, however, provide the valuable insight that the theory and methods for impact measurement from the hard sciences cannot simply be copied to use as metric for impact in legal IR systems.

When creating a citation index that included both scholarly and non-scholarly documents, and using this for bibliometric-enhanced rankings for both scholars and practitioners alike, we encountered some missing data, since a substantial number of documents are never cited, and since citations only capture impact on authors. For documents that are never cited, the illusion could exist that they have had no impact on the field even though, like the *Scientific American* example, they may have had a different form of impact. For documents that have been cited, we have data on the impact they have had on other authors, but not on non-author users, meaning that we may be missing part of the picture of the total impact the document has had. We therefore suggest to combine citation metrics with usage metrics. Future research will focus on the correlation between citations and usage, and possibilities to combine these two metrics into an overarching view of the impact of legal documents on the legal community as a whole for use in bibliometric-enhanced legal IR systems.

3.7 Acknowledgements

The authors thank the employees of Legal Intelligence, in particular dr. T.E. de Greef and P. van Boxtel for their cooperation in this research.

The authors acknowledge prof. W.H. van Boom who, in the spirit of scholarly debate, send us his inaugural lecture on annotations after reading an earlier version of this paper.

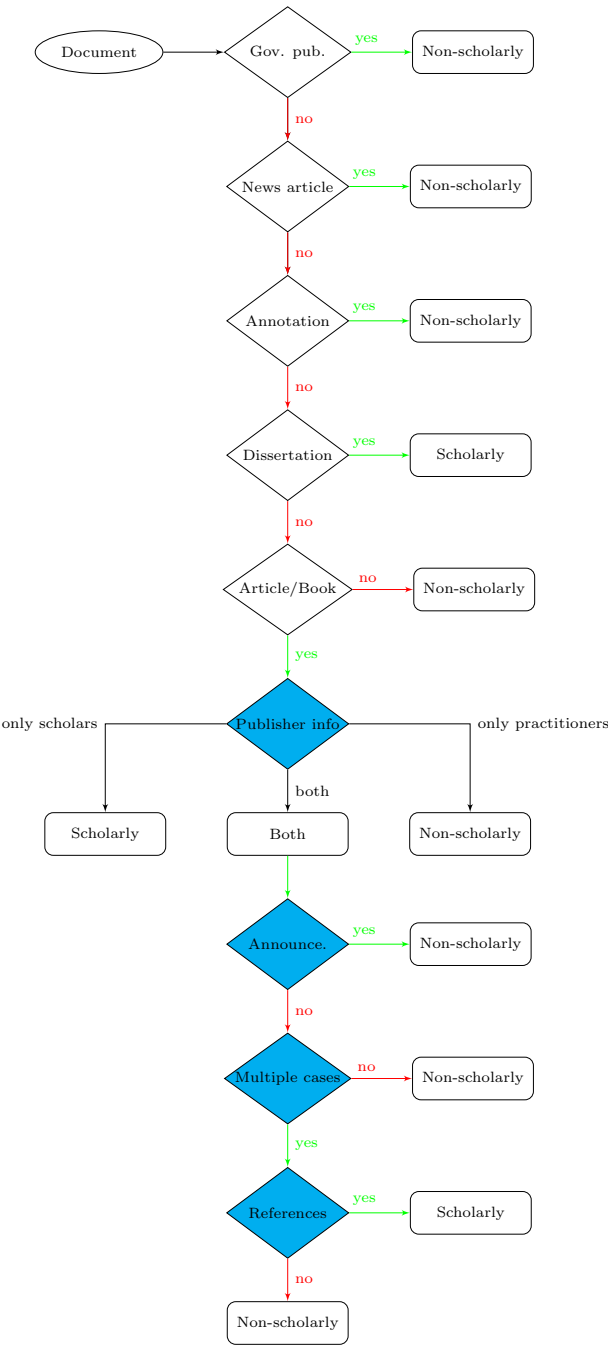


Figure 3.1: Flowchart of classification schema

