

USER STUDY ON SELECTION SEARCH AND SEMANTIC TEXT MATCHING IN GERMAN TENANCY LAW

Jörg Landthaler / Ingo Glaser / Hans Lecker / Florian Matthes

Research Associates, Technical University of Munich, Department of Informatics, Software Engineering for Business Information Systems, Boltzmannstraße 3, 85748 Garching bei München, DE

joerg.landthaler@tum.de

<http://www.matthes.in.tum.de>

Research Associates, Technical University of Munich, Department of Informatics, Software Engineering for Business Information Systems, Boltzmannstraße 3, 85748 Garching bei München, DE

ingo.glaser@tum.de

<http://www.matthes.in.tum.de>

Rechtsanwalt, Produktmanager der Zielgruppe Rechtsanwälte

Haufe-Lexware GmbH Co. KG, Fraunhoferstr. 5, 82152 Planegg/München, DE

hans.lecker@haufe-lexware.com

<http://www.haufe-lexware.com>

Professor, Technical University of Munich, Department of Informatics

Software Engineering for Business Information Systems, Boltzmannstraße 3, 85748 Garching bei München, DE

matthes@tum.de

<http://www.matthes.in.tum.de>

Keywords: *User Study, Contract Drafting, German Tenancy Law, Selection Search, Keyword Search, Semantic Text Matching, Word Embeddings, TFIDF, Word AddIn*

Abstract: *Legal research is fundamental to the lawyer's daily work and benefits from better work-flow integration and high-quality search results. We present an AddIn that integrates search functionality into a text processing environment. Selection Search is a search method where users select text as input to a search query. Our user study suggests that the Selection Search can be a complementary search method to Keyword Search and that the quality of the results returned by a search technology based on Word Embeddings is perceived best by the study participants.*

1. Introduction

Legal research is an essential part of many legal professionals' daily work [LASTRES, 2015]. Legal information retrieval denotes support for legal research through technology. In contrast to general information retrieval, legal information retrieval has specific requirements. For example, legal language uses specific vocabulary with precise and vague definitions at the same time, cf. [VAN OPIJNEN AND SANTOS, 2017]. Legal professionals face several challenges. Two major challenges are the exposure to an ever-growing amount of relevant literature and a market that demands more efficient workflows. These challenges can be addressed with more efficient and effective search tools that are better integrated into legal workflows.

In [LANDTHALER ET AL. 2018a], we investigated an approach that supports lawyers in their daily work of contract analysis. Lawyers select a text fragment, for example a clause, in the contract. The selected text fragment subsequently serves as input to a search query. We call this human-computer interaction method «Selection Search». It can be seen as an alternative or complementary search method to a traditional Keyword Search that could increase the efficiency of legal research due to the faster compilation of search queries and integration into common working environments. We already evaluated different technologies for a use case in German tenancy law using two evaluation sets and framed the information retrieval task as an instance

of a semantic text matching problem. Our prior work investigated multiple text-similarity-based solution approaches, in particular, the application of the TFIDF and Word Embeddings technologies. We proposed an approach leveraging Word Embeddings characteristics: legal comment chapters are segmented into sentences, ranked and the containing chapters are returned as search results. In the following, we use the term *search methods* for the two different human-computer action methods Selection Search and Keyword Search. We call the different technologies that deliver results for the Selection Search *search technologies*. Legal information retrieval heavily depends on the search intention of the user. The present work contributes the results of a user study conducted with 15 lawyers. The study assesses the usability of Selection Search and Keyword Search integrated into a Microsoft Word AddIn, a text processing system widely used in the legal domain. Our study further investigates the perceived quality of the results of three different search technologies based on TFIDF and Word Embeddings that can be employed to deliver results for the Selection Search. In particular, we want to answer the following main research questions:

1. How do lawyers assess the usability of Selection Search and Keyword Search integrated in a Word AddIn when auditing contracts? (RQ1)
2. How do lawyers assess the quality of the results returned with SENT, CHAPTER or MLT search technologies when using the Selection Search? (RQ2)

The remainder of this paper is organized as follows: Section 2 puts our work into the context of related work. Section 3 introduces the tool that is the subject of our user study. In Section 4, we discuss the user study design and present the results in Section 5. We conclude with a summary in Section 6.

2. Related Work

[LASTRES 2015] conducted a user study with 190 associate-level participants to investigate search behavior, primarily for case law. The study reveals that associates spend a significant amount of their weekly work time on search tasks, with most of the time using paid online services, around half as much time with free online services and even less with printed resources. [PEOPLES 2005] performed a user study in case law with 56 law students mainly to compare printed and electronic resources for finding legal rules versus cases with similar fact patterns. However, the study also compared Keyword Search to Westlaw's «terms and connectors» search. It is an enhanced Keyword Search that allows users to combine multiple keywords with operators. For example, to return results where multiple keywords occur in the same sentence or paragraph. The study's outcome is that users favor an enhanced Keyword Search over a plain Keyword Search, i.e. lawyers prefer an improved control on the search functionality. A user study with 43 legal professionals was carried out by [WIGGERS ET AL. 2018] to judge the relevance of 13 document-level meta-information attributes of search results. According to this user study, the most influencing factor is the title of a document. Other relevant factors include document type, recency, level of depth, legal hierarchy, law area and author credibility. In contrast to the above-mentioned user studies, the participants of our user study mainly encompass senior lawyers. We apply an established usability measure (SUS) to evaluate the usability of different search methods and restrict our use case to the specific law domain «tenancy law» and the German legal sphere.

At the time of writing, LexisNexis released a Microsoft Word AddIn called «SmartScan»¹ in a closed beta test. Similar to our approach, SmartScan enables users to analyze a selected text fragment and displays the result of citation analysis and recommends documents from the LexisNexis database. In contrast to our approach, SmartScan does not recommend individual chapters from documents so far. Westlaw / Thomson Reuters²

¹ <https://www.lexisnexis.at/produkte/lexis-smartscan-juristische-textanalyse-automatisierte-quellenrecherche/>, Lexis Nexis SmartScan, last accessed January 2019.

² <https://legal.thomsonreuters.com/en/support/product-downloads>, Thomson Reuters Product Downloads, last accessed January 2019.

offers several Word AddIns tailored to legal professionals that focus on case law but are unrelated to Selection Search.

3. Subject of User Study

We chose Microsoft Word as a widely used text-processor and implemented the Selection Search as an AddIn. Figure 1 shows a screenshot of the Word AddIn. Users select a text fragment of interest which serves as input to a search query. We implemented different search technologies for the Selection Search in a server-side backend. The AddIn also provides a traditional Keyword Search that will serve as an alternative search method. In contrast to the keyword-based search, the Selection Search imposes stronger requirements on the search technology to deeper analyze the semantics of the search query and our hypothesis is that Word Embeddings, that capture some aspects of semantics, can be a suitable technology. Our prototype provides three different search technologies for the Selection Search and a very basic Keyword Search:

- **SENT:** The legal comment chapters are segmented into sentences. The sentences are encoded with Word Embeddings by accumulating individual Word Embeddings vectors (the «sentence» vectors can be calculated a priori and cached). Likewise, a search query is encoded with Word Embeddings. The most similar vectors are calculated with cosine similarity. In contrast to the CHAPTER search technology, more equally sized text blocks are compared by the text similarity measure.
- **CHAPTER:** The full legal comment chapters and the search query are represented with accumulated Word Embeddings vectors similar to the SENT approach. However, full legal comment chapters are represented, similar to term frequency-inverse document frequency (TFIDF), by a single vector.
- **Elasticsearch «More like this» (MLT):** This approach uses the «more like this» search functionality provided by Elasticsearch³ with default parameters. Both, the documents and the search query are encoded using the TFIDF representations.
- **Keyword Search (KWS):** We use a simple Keyword Search provided by Elasticsearch («query_string») that matches exact tokens.

³ <https://www.elastic.co>, version 6.3.1, last accessed January 2019.

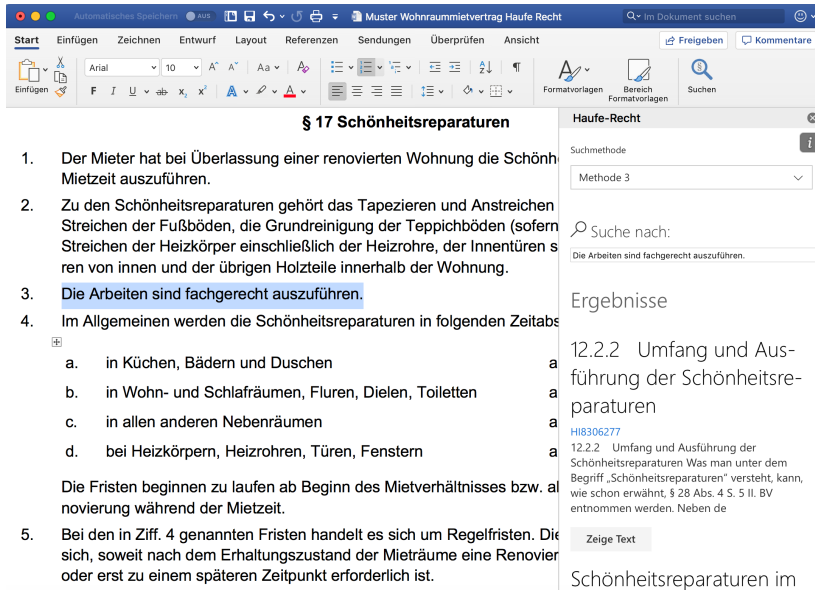


Figure 1: Screenshot of our Microsoft Word AddIn that shows an excerpt of a German tenancy contract template (left) and in the AddIn (right) results for the selected text «The operations need to be carried out technically correct».

The search methods/technology can be selected with a dropdown labeled «Suchmethode» («Search method»). However, the different technologies for the Selection Search were anonymized for the user study. Users can modify search queries in the text field labeled with «Suche nach» («Search for»). The AddIn allows to explore the full legal comment chapter of a search results by clicking on a button labeled «Zeige Text» («Show text»). For the Selection Search queries could be created not only by selecting text, but also by modifying the query in the text field. However, the user study participants were not told about this option explicitly.

The Word Embeddings have been trained using word2vec⁴ on six exemplary tenancy contract templates and three legal comments: [KINNE ET AL. 2013], [STÜRZER ET AL, 2015], [STÜRZER AND KOCH, 2017]. The sentence segmentation was calculated using spacy⁵, as described in more detail in [LANDTHALER ET AL. 2018a]. The legal comment’s plain text has not been further pre-processed for indexing in Elasticsearch.

We earlier manually created two evaluation sets: «narrow» and «broad» that reflect directly and additionally indirectly subject-specific relevant results for a given segmentation of the contract clauses regarding «cosmetic repairs», cf. [LANDTHALER ET AL. 2018a]. These evaluation sets, however, do not capture that users can select arbitrary text fragments in a contract. Nevertheless, it can serve as a baseline for the results of our user study. Figure 2 displays precision/recall curves (a curve close to top/right is best) for the different search technologies used later on in the user study. According to the precision/recall search technology evaluation, SENT performs best, MLT performs almost as good as SENT and CHAPTER performs worst with a larger distance to SENT and MLT search technologies.

⁴ <https://github.com/kzhai/word2vec>, version 0.1c (for Mac OS X), last accessed January 2019, Parameters: default parameters except iterations: 100, min-count: 1, vector-size: 300.

⁵ <https://spacy.io/>, version 2.0.5, with German core language model, last accessed January 2019, Sentence segmentation results post-processed as described in [LANDTHALER ET AL. 2018a].

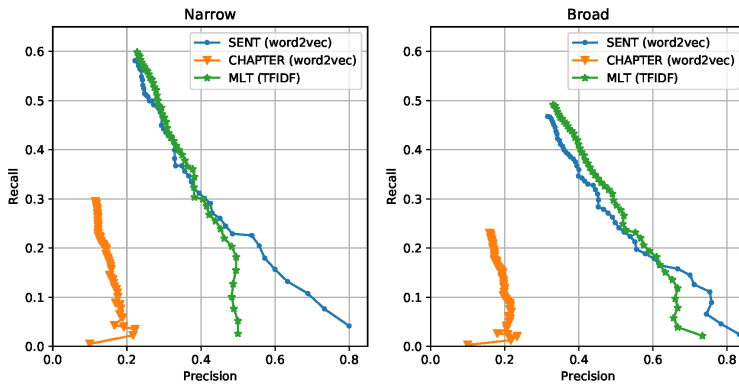


Figure 2: Precision-Recall curves calculated on two evaluation sets for our German tenancy law use case for three different search technologies: SENT and CHAPTER (Word Embeddings based), and MLT (TFIDF based). Number of search results (k) varying between 1 and 50.

4. User Study Layout

For the design of our user study, we follow in parts the guidelines of [PURCHASE 2012]. In order to answer RQ1, we asked the participants to answer a crafted legal question using the AddIn’s Selection Search and afterward applied an established usability measure: system usability scale (SUS). SUS is a collection of 10 questions designed to assess the usability of software systems as an interpretable single score, see [BANGOR ET AL. 2009]. The SUS questions were translated to German⁶. The question of the task was to decide whether a landlord can claim costs for a professional performing a cosmetic repair if the tenant did not carry out the cosmetic repair in a technically correct manner. The participants were asked to use the Selection Search method to answer the question. We controlled the presented search results by including a text similarity measure to mitigate minor deviations or faults in the search query. Therefore, we can assure that the answer to the task’s question can be given with the displayed results returned by the search in the AddIn. The users should use the anonymized search method that utilized the SENT approach. We asked the participants to use the Keyword Search to search for results for the earlier posed question and asked the SUS questions a second time. The reuse of the question could facilitate the usage and assessment of the Keyword Search, because potentially relevant legal comment chapters are already known by the participants.

In order to answer RQ2, we followed a different evaluation strategy. We instructed the participants to freely use the different available search methods in the Word AddIn on freely chosen documents and freely chosen text selections. Afterward, we directly asked the participants to judge the quality of the results of the search technologies answering Selection Search queries. We asked the participants to use the Keyword Search in an unrestricted manner and to judge it, too.

5. User Study Results

We sent an invitation e-mail to 6 600 lawyers from the customer database of our industry partner. 65 lawyers responded and received the second e-mail with detailed instructions, 22 started and 16 completed the questionnaire. We had to remove eight records due to a missing question in the SUS resulting in eight records. Due to low participation rate, we asked 25 personal contacts and employees of the industry partner resulting in seven additional completed questionnaires. In total, we could analyze 15 complete records. We provided de-

⁶ <https://experience.sap.com/skillup/system-usability-scale-jetzt-auch-auf-deutsch/>, last accessed January 2019.

tailed installation instructions for the Word AddIn, a sample tenancy contract template [HARNER, 2017] and a customized link to the questionnaire. The completion of the questionnaire took around 45 minutes on average.

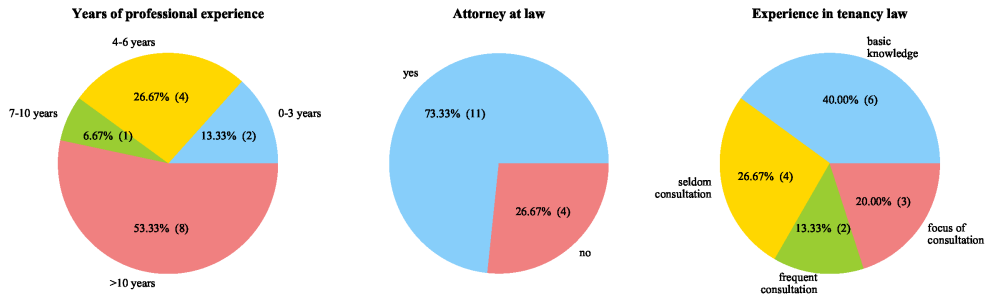


Figure 3: Participants characteristics.

We asked the participants general questions: The participants mostly are experienced lawyers, more than half of the participants have been working as a lawyer for more than ten years, cf. Figure 3. Almost three quarters of the participants are an attorney at law. The diversity of the lawyers’ areas of focus is broad. However, the largest fractions of the participating lawyers are active in labor and tenancy law (5 mentions each (33%), for up to three possible selections from a given set, result not visualized in a figure). The self-rating of experience in tenancy law ranges from basic knowledge to focus of consultation, see Figure 3. We asked for the legal research behavior, see Figure 4: The majority of participants estimated that they spend 20 to 40% of their working time for legal research. Therefore, for our participants, legal research is a substantial part of their daily work. Another interesting aspect is the high fraction of electronic resources used for legal research. 13 (87%) participants stated to start a legal research process with paid online services, while in later stages of the legal research process all resources, paid and free online services, and even printed documents are used by more than half of the participants. The SUS enables us to assess the usability of the AddIn with a single score between 0 and 100 when participants use the Selection Search or the Keyword Search. Figure 5 displays the results for both search methods with interpretations of the SUS scores proposed by [BANGOR ET AL. 2009].

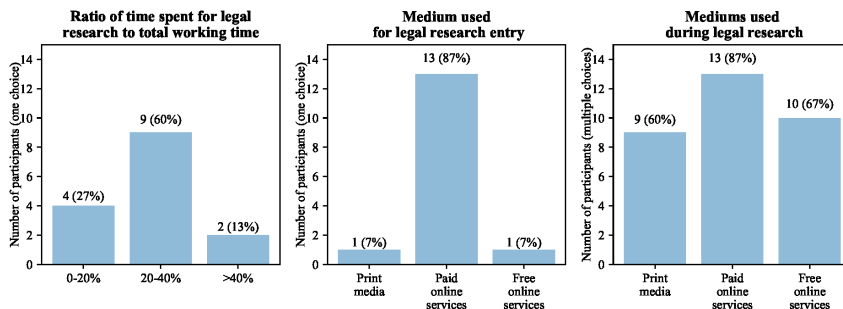


Figure 4: Legal research behavior of the participants.

The SUS score is calculated as the average of the individual SUS scores of each participant. The results are displayed as boxplots so that the distribution of the SUS scores among the participants is better visualized. The

SUS score for both, Selection Search and Keyword Search is a solid «C» grade. Some participants used the Keyword Search in the fashion of the Selection Search, i.e. by selecting individual words or text fragments as input to the Keyword Search rather than thinking of and typing keywords manually. This reduces the number of possible search queries for the participants when using the Keyword Search. The participant's opinion of the two search methods is rather ambiguous, reflected by the diversity of the individual SUS scores. In summary, the results indicate that we can answer RQ1 that the participants of the study judge the usability of both Word AddIn's search methods as good.

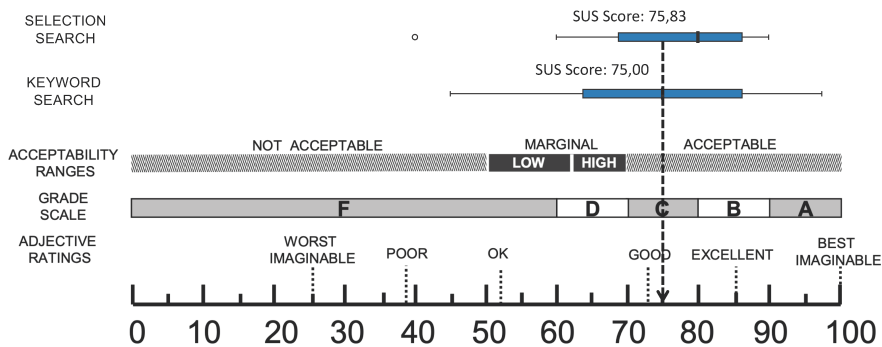


Figure 5: SUS scores for Selection Search and Keyword Search. Image adapted from [BANGOR ET AL. 2009].

When asked for the exclusive, most important reason to choose Selection Search or Keyword Search, 10 participants chose the quality of the results while 2 participants chose the usability of the tool and 3 participants voted for the control on the search functionality and one participant chose option «other reason» (results not visualized in a figure). This indicates that the quality of the search results is crucial. In parts, the participants' judgment of the different search technologies confirms the precision/recall results obtained from our manually created evaluation sets. The lawyers were asked to rate the different (anonymized) search technologies for the Selection Search with school grades. According to this evaluation, the participating lawyers seem to prefer the SENT approach, as can be derived by the boxplots in Figure 6 as well as the individual gradings of the search technologies. We additionally let the participants grade the Keyword Search with school grades.

However, the variation is very high and the preference for one of the search methods is split about half for Selection Search (SENT) and half for Keyword Search. This is also reflected by the free text comments of the participants: Two participants complained that the contract text is too general as input to a search query in general. Two participants thought that longer selected text fragments for the Selection Search lead to better results, while two participants stated the opposite. Several comments included that a quality judgment among the different search technologies is difficult and a general trend is not identifiable. The participants comments included several interesting possible improvements: inclusion of a Keyword Search functionality within the search results, the highlighting of matching text fragments and a save functionality of good result sets. Next, we forced the participants to choose their favorite search functionality (options were all Selection Search methods and Keyword Search).

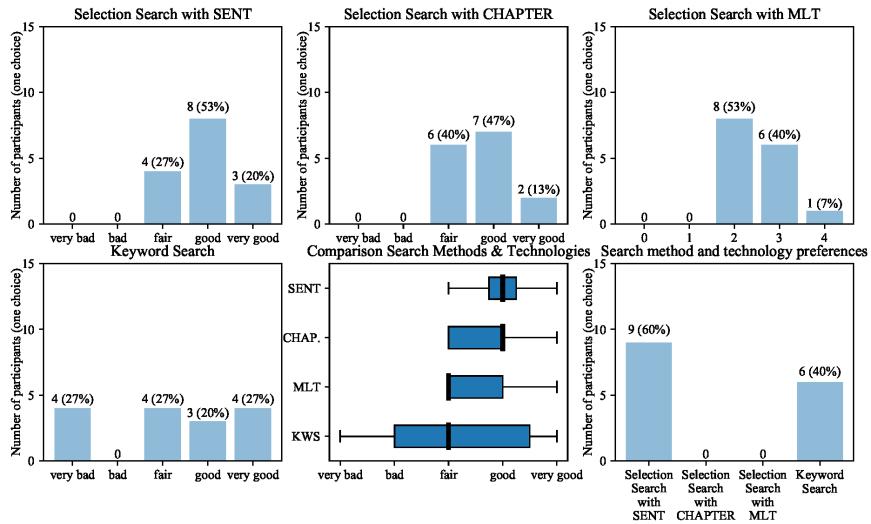


Figure 6: Participant's rating of search methods and search technologies.

As indicated by the results in Figure 6 (bottom, right), the participant's opinion is polarized and split into around half for the Selection Search with SENT search technology and Keyword Search. In summary, we can answer RQ2 that the participants slightly prefer the SENT search technology over CHAPTER and MLT.

Threats to the validity of the results include that some participants (actual number not clearly identifiable) used the Keyword Search by selecting text fragments in the contract (which was neither intended nor anticipated by us in advance). The Keyword Search implementation was very basic and did not include search query expansion, substring search or spelling correction. The participants were shown one working example of the Selection Search right before showing the SUS questions for the Selection Search. Also, we did not randomize the order of presenting the Selection Search and the Keyword Search as well as not randomizing the order of the search technologies for Selection Search. Multiple free text comments give rise to a need for better control and transparency of the search functionality. However, 14 participants (93.3%) of the participants said that they would like to use the Selection Search in the future. We did not investigate further correlations due to the small number of participants.

6. Conclusion

We conducted a user study with 15 legal professionals. Only a small number of participants could be acquired, because of the long questionnaire, a lawyer's expensive time and technical obstacles in terms of installation. Our results might not be generalizable well but in parts go in line with related scientific research. The results of our user study indicate that legal research is an essential part of many lawyers' daily work. Nowadays, lawyers mostly use online services, both paid and free services, in comparison to printed resources. Our study suggests that a text processor AddIn can be a valuable tool for lawyers when auditing contracts and most lawyers would like to use such a tool in the future. The study identifies the quality of the search results as the most important criteria for the usability of the search functionality. Our SENT approach based on Word Embeddings delivers the best results in terms of precision/recall. This is supported by the participant's perception of the quality of the results. The usability of the Selection Search is similarly well perceived as a very basic Keyword Search. However, many lawyers prefer a certain control on the search functionality as well as traceability of the mechanisms that lead to search results. A Keyword Search allows for greater control and traceability in comparison

to the Selection Search at the current stage and the input to the Selection Search is limited to the text at hand. Therefore, the Selection Search could be useful as a complementary search method to Keyword Search. We hypothesize that our Selection Search approach would benefit from deeper semantic analysis capabilities of the search queries and the knowledge base/corpus (that is, however, very challenging) that might lead to higher quality search results. It could benefit from additional control capabilities to steer the search functionality and explainable AI approaches to rationalize the search results, see for example [LANDTHALER ET AL. 2018b]. All three directions give rise to future scientific research.

7. Acknowledgment

We want to thank all participants of the user study for their precious time and valuable feedback.

8. References

- BANGOR, AARON/KORTUM, PHILIP/MILLER, JAMES, Determining what individual SUS scores mean: Adding an adjective rating scale, *Journal of usability studies*, 4. Jg., Nr. 3, Pages. 114-123, 2009.
- KINNE, HARALD/SCHACH, KLAUS/BIEBER, HANS-JÜRGEN, Miet- und Mietprozessrecht, 7. Auflage 2013, Haufe Lexware.
- LASTRES, STEVEN A., Rebooting Legal Research in a Digital Age, 2015.
- LANDTHALER, JÖRG/SCEPANKOVA, ELENA/GLASER, INGO/LECKER, HANS/MATTHES, FLORIAN, Semantic Text Matching of Contract Clauses and Legal Comments in Tenancy Law, IRIS: Internationales Rechtsinformatik Symposium, Salzburg, Austria, 2018, a.
- LANDTHALER, JÖRG/GLASER, INGO/MATTHES, FLORIAN, Towards Explainable Semantic Text Matching, Proceedings of Jurix 2018: International Conference on Legal Knowledge and Information Systems, Groningen, Netherlands, 2018, b.
- MIKOLOV, THOMAS/CHEN, KAI/CORRADO, GREG/DEAN, JEFFREY, Efficient Estimation of Word Representations in Vector Space, International Conference on Learning Representations, 2013.
- PEOPLES, LEE F., The death of the digest and the pitfalls of electronic research: what is the modern legal researcher to do. *Law Library. Journal.*, Volume 97, 2005.
- PURCHASE, HELEN C., Experimental Human-Computer Interaction: A Practical Guide with Visual Examples, Cambridge University Press, New York, NY, USA, 2012.
- HARNER, MANFRED, Mietvertrag, Eigentumswohnung, Haufe Lexware, Version from 04.01.2017, Mietvertrag, Eigentumswohnung, Haufe Lexware, <https://products2.haufe.de/#G:pi=PI17574&&;D:did=HI1949297&&>, last access January 2019.
- STÜRZER, RUDOLF/KOCH, MICHAEL, Vermieterlexikon, 15. Auflage 2017, Haufe Lexware.
- STÜRZER, RUDOLF/KOCH, MICHAEL/NOACK, BIRGIT/WESTNER, MARTINA, Vermieter-Praxishandbuch, 8. Auflage, 2015, Haufe Lexware.
- VAN OPIJNEN, MARC/SANTOS, CHRISTIANA, On the Concept of Relevance in Legal Information Retrieval, Artificial Intelligence and Law Volume 25, Issue 1, pp 65–87, 2017.
- WIGGERS, GINEKE/VERBERNE, SUZAN/ZWENNE, GERRIT-JAN, Exploration of Intrinsic Relevance Judgments by Legal Professionals in Information Retrieval Systems. In Proceedings of Dutch-Belgian Information Retrieval Workshop (DIR2018). ACM, New York, NY, USA, 2018.