

Julian Mosen (2022): The Structure of Social Documents: Visualising Digital Traces of Collaborative Work. In: 20th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies - Doctoral Colloquium (ISSN 2510-2591), DOI: 10.48340/ecscw2022_dc04

Copyright 2022 held by Authors, DOI: 10.18420/ecscw2022_dc04

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

The Structure of Social Documents: Visualising Digital Traces of Collaborative Work

Julian Mosen

Center for Enterprise Information Research, Institute for IS Research,
University of Koblenz, Germany

julianmosen@uni-koblenz.de

Abstract. Social documents are user-generated digital artefacts, created for collaboration and communication among employees. Typical examples are wiki articles or blog posts including comments and attachments, which are stored and interacted with in enterprise collaboration systems. These documents are highly networked and contain a huge amount of meta data about who created or contributed new information. They provide a valuable data source for the reconstruction of user interactions and can be transformed into detailed descriptions of patterns and practices of digital work. The analysis of such rich stories will help to develop a better understanding how digital collaboration takes place in organisations. Unfortunately, existing research on social documents as traces of digital work is very limited. Assembling the single traces to rich stories requires knowledge about components, possible compositions, relations and emerging structures of social documents. These characteristics are less explored and limit the use of social documents for further studies on collaborative activities. The proposed dissertation aims to contribute to this problem through an in-depth investigation of the structures of social documents by developing and applying methods and tools for visualising the network and graph characteristics of real data to uncover patterns and common characteristics of digital work.

Motivation and Problem Statement

The increasing relevance of digital work and digital collaboration has been recently confirmed by the Corona pandemic in 2020 and 2021, when the German government forced companies to offer remote work to their employees. Even long before the pandemic, the research fields of *Information Systems* (IS) and *Computer Supported Cooperative Work* (CSCW) investigated how employees are technically supported to enable working together (Greif, 1988; Schmidt & Bannon, 2013) and how collaboration is digitally carried out by employees in practice.

Lately, the emergence of *enterprise social software* (ESS) and *enterprise social networks* (ESN) changed things and resulted in a redesign and transformation of groupware and collaboration systems to complex information infrastructures providing integrated collaboration platforms (Williams & Schubert, 2018) with social features originating from Web 2.0. The seamless integration of new applications, such as wikis, blogs or message boards, has led to new work practices and provided new opportunities for capturing, sharing or combining information and coordinating joint work (Nitschke & Williams, 2018; Schwade & Schubert, 2017). The use of these applications leaves abundant types of *digital traces of work* in the system. Shared files and wiki articles with version histories, blog and board posts with comments and recommendations are just a few of the heterogeneous data that remain in the systems as evidence for user interactions. Each of these digital traces contains a large amount of metadata with information about who created or modified the content at what time. Accordingly, individual user actions can be reconstructed by assembling the single traces in the right order. Their transformation into rich stories of user interactions and the analysis in context with appropriate methods, such as *trace ethnography* (Geiger & Ribes, 2011), seems to be a promising approach to gain valuable insights about the way people communicate and collaborate in organisations.

Unfortunately, there are few studies that examine the traces of digital collaboration in enterprises. Common limitations of existing studies are the focus on either homogeneous data, single types of applications, small user groups, limited timeframes and specific types of activities (Bean & Hott, 2005; Holtzblatt et al., 2010; Millen et al., 2005; Nagel & Schwade, 2020; Richter & Riemer, 2013). These limitations are mostly based on one of the following challenges: (1) Since broad user observations are very time-consuming and very expensive it makes sense to look at user-generated data. As this data is stored in company-internal systems located behind corporate firewalls with limited access to employees only, the data access is quite difficult. (2) The digital traces are often stored in a distributed manner. Each integrated application of a collaboration system can have its individual data stores, data bases and data interfaces. As a result, the data collection requires additional efforts, methods and tools. (3) Digital traces are heterogeneous data, having individual data structures depending on the type of content (Østerlund

et al., 2020). Collaboration is not limited to single documents, but can take place across different types of applications or systems (Schubert & Glitsch, 2016). Work can start in one place, for example two people editing a wiki article, and continue anywhere else, such as a discussion in a related message board. Consequently, the analysis of richer cases from practice requires a detailed understanding of the data structures and characteristics of each application and its content types.

The challenges of previous studies are addressed in this study as follows: To examine richer and more complex forms of collaboration, a large infrastructure with hundreds of users and heterogeneous types of integrated applications serves as data source. Having full technical access to the databases of the academic collaboration platform *UniConnect*¹, with more than 4000 users (researchers, employees and students) from 35 universities and research institutions, allows to look at user-generated content that is reused and combined across multiple applications. To make use of the heterogeneous traces it is essential to have a common model to analyse and compare different types of content. A fundamental type of digital trace for the planned project is conceptualised with the term *social document* (Hausmann & Williams, 2015), which describes user-generated digital artefacts, that are explicitly intended to be interacted with, can be shared and collaboratively developed, and composed of individual components (Hausmann & Williams, 2016). The concept of a social document covers a wide range of digital traces and offers a starting point for analysing the structures, that emerge when users work and interact with each other in the system. A social document consists of several components, such as a wiki article with comments and attachments. At the same time, it can be highly connected by referencing or being referenced from related documents.

Research Questions

The overarching research question addressed by the dissertation is *how and what insights can be gained about collaborative activities in enterprise collaboration platforms from analysing and interpreting social documents, their structures and their visualisation as digital traces work.*

To answer the research question and to achieve the goal of the thesis, there are four *research questions* (RQ).

- RQ1: Which terminologies are used in the context of social documents in practice and research and how can they be harmonised?
- RQ2: How are social documents structured internally and how are they integrated into their context?

¹ UniConnect is based on HCL Connections, on one of the market leading software products for enterprise collaboration (<https://uct.de/uniconnect>)

- RQ3: How can the general structure, individual instances and related user interactions of social documents be visualised?
- RQ4: How and what insights into collaborative activities can be gained from the visualisation of social documents?

Each research question is related to a *research objective* (RO) and helps to develop the *research artefacts* (RA). RO1 is to harmonise the terminology on social documents by analysing existing theories and concepts and reviewing common terminology from market leading enterprise collaboration systems. By merging and harmonising the inconsistent terminologies a taxonomy for social documents results as research artefact RA1.

Based on the findings from RO1, a model for the generic structure of social documents is developed (RO2) by reviewing existing models from previous research and analysing their applicability to social documents. Secondly, an observation of the functional and the technical implementation of social documents is carried out for market leading collaboration systems. By merging the findings, an ontology is derived that provides a generic description of structures, relations, dependencies and constraints for social documents and their components. The formalised ontology constitutes the second research artefact RA2.

RO3 is to develop a novel visualisation, that allows to represent internal structures and relations between social documents. This will be evaluated with real data from available collaboration systems. The resulting research artifact RA3 is a method for the graph-based visualisation of social documents.

The last research objective RO4 aims at gaining new insights into collaborative activities. The model from RO2 and the visualisation from RO3 are applied to a large, heterogeneous set of real data from an enterprise collaboration system (UniConnect) on a broader scale. A set of tools is developed to automatically extract social documents from the collaboration system and generate their graph-based visualisations for a larger number. The results are used to study and explore the characteristics of different types of user interactions. The research artefact RA4 is a catalogue of visualisations for real social documents including rich descriptions of the patterns and practices, hidden in those digital traces of work.

Methodological Approach

The research design is guided by the *Design Science Research Methodology* (DSRM) from Peffers et al. (2007). The research process described by DSRM and its cyclical structure provide guidance for the development and evaluation of the expected research artefacts. The cyclical alternation between development and evaluation is used to structure the work on the level of each individual research objective, but also at the higher level for the evaluation and quality assurance of each research artifact against each other. The research methods are based on desk

research and secondary research. For the research objectives RO1 and RO2, a literature analysis and several system analyses are conducted. RO3 is based on prototyping of different visualisations. Research objective RO4 makes use of database reverse engineering and quantitative data collection from existing enterprise collaboration platforms.

Important prerequisites for the success of the project are access to collaboration systems used in practice, data sources for social documents and sufficient computing and storage resources for the implementation. The access to an enterprise collaboration system is provided through the University Competence Center for Collaboration Technologies (UCT)². Necessary computational resources and storage capacities are ensured by the Socio-Physical Advanced Research Cloud Infrastructure (SPARCI)³.

Findings to Date

Regarding RQ1 and RQ2, a first literature analysis on structural descriptions of user generated Enterprise 2.0 content was conducted and a reverse engineering of the implementation of social documents in a leading enterprise collaboration system was carried out. The findings from literature and reverse engineering then were used to develop the *Social Document Ontology* (SocDOnt), an ontological description of social documents.

One important result of the literature analysis was the existence of an ontology from research on the Semantic Web. Passant et al. (2010) developed the ontology *Semantically-Interlinked Online Communities* (SIOC) and its extension SIOCT to enable interoperability and exchange of user generated content in Web 2.0 applications. A first application of SIOC for data exchange in enterprise systems was described by Lee et al. (2008), who mapped concepts from SIOC to *collaborative work environments* in BSCW, Business Collaborator and NetWeaver. The authors identified that there is a lack of important concepts for a more detailed description. This observation is consistent with those made after our reverse engineering of the technical implementation of social documents in HCL Connections (formerly IBM Connections). We identified more than ten additional concepts that are necessary to overcome the limitations of SIOC in recent collaboration systems. The new concepts have been captured in the *Social Document Ontology* (SocDOnt), which reuses existing concepts from SIOC and

² The UCT is a joint research cooperation between the University of Koblenz and the HCL Technologies Germany GmbH (<https://uct.de/>)

³ The Socio-Physical Advanced Research Cloud Infrastructure (SPARCI) is a research cloud infrastructure funded by the Deutsche Forschungsgemeinschaft (DFG) (<https://gepris.dfg.de/gepris/projekt/432399058>)

SIOCT and introduces new concepts that are necessary for a more detailed description of social documents (Williams et al., 2020).

Addressing RQ3, a graph based visualisation for social documents was developed and evaluated with multiple data sets from UniConnect (Mosen et al., 2020). Based on concepts from graph theory, a method for the graphical visualisation of social documents was implemented and allowed to compute first representations for different types of social documents and identify their characteristic structures. The knowledge about characteristic structures for each content types is a precondition for identifying the details and differences, that emerge through individual work practices.

Recently, the SocDOnt is evaluated for the description of further types of user generated content. An important content type is chat messages and related concepts, that can be found in systems for enterprise messaging and chat systems, such as Microsoft Teams, Skype, Mattermost or Slack. Like social documents, chat messages invite for interactions, are highly interconnected by referencing or embedding other documents, including tags, recommendations or having attachments.

Expected Contributions

The overall research aim is to provide a theoretical and practical contribution to the study of social documents and its interpretation as traces for digital work. The ontology-based specification of social documents addresses the limitations of existing ontologies by introducing missing concepts related to enterprise collaboration systems. For the present work, the ontology provides a precise description of the components and relations of social documents, which is the theoretical foundation for the visualisation of social documents. The implementation of a visual representation provides a new method for visualising digital traces of collaboration and can be used to retrospectively reconstruct user interactions and work practices.

References

- Bean, L. and Hott, D. D. (2005): 'Wiki: A speedy new tool to manage projects', *Journal of Corporate Accounting & Finance*, vol. 16, no. 5, pp. 3–8.
- Geiger, R. S. and Ribes, D. (2011): 'Trace ethnography: Following coordination through documentary practices', *44th Hawaii International Conference on System Sciences*, pp. 1–10.
- Greif, I. (1988): *Computer-supported cooperative work: a book of readings*, Morgan Kaufmann Publishers Inc., San Mateo, CA.
- Hausmann, V. and Williams, S. P. (2015): 'Social Business Documents', *Procedia Computer Science*, vol. 64, pp. 360–368.
- Hausmann, V. and Williams, S. P. (2016): 'Issues for the long-term management of Social Business Documents', *International Journal of Information Systems and Project Management*, vol. 4, no. 3, pp. 45–61.
- Holtzblatt, L. J., Damianos, L. E. and Weiss, D. (2010): 'Factors Impeding Wiki Use in the Enterprise: A Case Study', *28th ACM Conference on Human Factors in Computing Systems*.
- Lee, D., Peristeras, V. and Loutas, N. (2008): 'Reusing the SIOC Ontology to Facilitate Semantic CWE Interoperability', *Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part II on On the Move to Meaningful Internet Systems*, pp. 1517–1531.
- Millen, D. R., Muller, M. J., Geyer, W., Wilcox, E. and Brownholtz, B. (2005): 'Patterns of media use in an activity-centric collaborative environment', *CHI 2005*, pp. 879–888.
- Mosen, J., Williams, S. P. and Schubert, P. (2020): 'Visualizing Social Documents as Traces of Collaborative Activity in Enterprise Collaboration Platforms', *53rd Hawaii International Conference on System Sciences*, pp. 5369–5378.
- Nagel, S. and Schwade, F. (2020): 'Analysis of Task Management in Virtual Academic Teams', *53rd Hawaii International Conference on System Sciences*, pp. 461–470.
- Nitschke, C. S. and Williams, S. P. (2018): 'Traces of design activity: the design of coordination mechanisms in the shaping of enterprise collaboration systems', *CENTERIS - Conference on ENTERprise Information Systems*.
- Østerlund, C., Crowston, K. and Jackson, C. (2020): 'Building an Apparatus: Refractive, Reflective & Diffractive Readings of Trace Data', *Journal of the Association for Information Systems*, pp. 1–43.
- Passant, A., Bojars, U., Breslin, J. G., Decker, S., Boj, U., Breslin, J. G. and Decker, S. (2010): 'The SIOC Project: Semantically-Interlinked Online Communities, from Humans to Machines', in J. Padget, A. Artikis, W. Vasconcelos, K. Stathis, V. T. da Silva, E. Matson and A. Polleres (eds.): *Coordination, Organizations, Institutions and Norms in Agent Systems V*, pp. 179–194, Springer Berlin Heidelberg.
- Peppers, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S. (2007): 'A design science research methodology for information systems research', *Journal of Information Management Systems*, vol. 24, no. 3, pp. 45–78.
- Richter, A. and Riemer, K. (2013): 'Malleable end-user software', *Business and Information Systems Engineering*, vol. 5, no. 3, pp. 195–197.
- Schmidt, K. and Bannon, L. (2013): 'Constructing CSCW: The First Quarter Century', *Computer Supported Cooperative Work*, vol. 22, pp. 345–372.
- Schubert, P. and Glitsch, J. H. H. (2016): 'Use Cases and Collaboration Scenarios: How employees use socially-enabled Enterprise Collaboration Systems (ECS)', *International Journal of Information Systems and Project Management*, vol. 4, no. 2, pp. 41–62.
- Schwade, F. and Schubert, P. (2017): 'Social Collaboration Analytics for Enterprise Collaboration Systems: Providing Business Intelligence on Collaboration Activities', *50th Hawaii*

- International Conference on System Sciences*, pp. 401–410.
- Williams, S., Mosen, J. and Schubert, P. (2020): ‘The Structure of Social Documents’, *53rd Hawaii International Conference on System Sciences*, pp. 2825–2834.
- Williams, S. and Schubert, P. (2018): ‘Designs for the Digital Workplace’, *Procedia Computer Science*, pp. 478–485.