

Information Behavior in the Social Web: An Overview of the German Educational Domain

Joachim Griesbaum¹, Nadine Mahrholz¹, Jens Bertram¹, Nadine Pietras¹ and Marc Rittberger²

¹ University of Hildesheim

² German Institute for International Educational Research

Abstract

This paper explores participative Social Information Behavior in the educational domain. The goal is to capture a picture of current information practices in the Social Web. The focus is on the “places” and the scale of the Social Web in the domain, the communication dynamics and structure of communities and the specificities, quality, pragmatics and success of communication processes. The paper describes the concept and current implementation status of an online analysis approach and system that tries to answer these questions. Furthermore, first empirical results are presented. Data indicates that participative Social Information Behavior is of relevance in the domain: The volume of openly accessible user-generated content is impressive. The basic characteristics of analyzed forums suggest that such websites resemble sustainable knowledge building communities. Pre-tests regarding the analysis of communication processes denote that generated content can often be seen as a valuable information resource.

Keywords: social web, social information behavior, educational domain

Citation: Griesbaum, J., Mahrholz, N., Bertram, J., Pietras, N., & Rittberger, M. (2014). Information Behavior in the Social Web: An Overview of the German Educational Domain. In *iConference 2014 Proceedings* (p. 356–371). doi:10.9776/14113

Copyright: Copyright is held by the authors.

Acknowledgements: This investigation is part of a research project sponsored by the German Institute for International Educational Research.

Contact: griesbau@uni-hildesheim.de, mahrholz@uni-hildesheim.de, bertra@uni-hildesheim.de, nadinepietras@msn.com, rittberger@dipf.de

1 Introduction

The Social Web can be seen as a communication infrastructure that enables and allows for (potentially unrestricted) n:m-communication. Its diffusion can be assessed as a fundamental paradigm shift of communication patterns of individuals as well as of the society as a whole. Social Media offer the possibility to satisfy communication needs which were previously too costly or impossible to be satisfied. Now, everybody can be a “media outlet” (Shirky 2008). The Social Web is primarily a leisure-time-based phenomenon of predominantly young internet users. Nevertheless, there is an ongoing discussion addressing the question, if and how the potentials of the Social Web can be transferred to professional contexts (Shirky 2010, Evans 2010, Hester 2010). This paper focuses on Social Information Behavior in educational contexts. The research motivation is to get insights into the question of how students of education related study paths (future educators) actively use Social Media for their own professional educational context.

Media usage and information practices are in progress. Search engines have changed the way users access information. Furthermore, the Social Web has widely expanded the universe of available knowledge itself (Ramakrishnan & Tomkins 2007). In addition, the Social Web has created new possibilities for personal, communal and public information sharing (Shirky 2010: 173). The adaptation of Social Media is very advanced with regard to personal communication and self-expression in Social Online Networks (SONs) (Allfacebook.de 2013) and product or service related information behavior (eMarketer 2013). Contrasting, the significance and role of the Social Web for professional contexts is often unknown. Therefore, the

question remains, if the Social Web is also of significance in the educational domain, as the current state concerning places, communication structures and quality of user-generated content is widely unclear.

This is the starting point of this work. We want to get an overview of the Social Web in the domain at hand and also a picture of the communication dynamics and structure in corresponding communities as well as insights into the specificities and quality of such online communication. Apart from the scientific research interest, results of this investigation should be especially of interest for specialized information providers, as they seem to be rather slow in adapting to the current developments. Their content base is usually limited to the “old” channels of professional information. Hence, the communicative paradigm of the Social Web is barely addressed. Services are still widely restricted to a content delivery metaphor. One can argue that these restrictions are necessary and a safeguard to deliver high quality information. Nevertheless, it seems possible and plausible that there is a widening gap between users’ information practices and the services offered by specialized information providers. The Web, the Social Web and the Future Sensor Web (O’ Reilly & Batelle 2009) are very disruptive technologies. One needs to know how users behave and what they expect in order to be able to adapt to a rapidly changing environment. Here, insights into participative information behavior can be seen as a basis on which professional information providers can build upon to improve or secure the usefulness and therefore significance of their services in the long run. However, as long as aspects like the quality of user-generated content or the role of online communities are widely unclear, specialized information providers are barely able to adapt to changing user behavior and needs.

The paper is structured as follows. First, the concept of Social Information Behavior is outlined and related to the educational domain. This provides a conceptual basis for this investigation. A draft of current research concerning Social Media usage in study related contexts illustrates the relevance of the topic at hand. Following that, the research approach and methods are described and first results are presented. On this basis, the significance of participative Social Information Behavior is discussed. The paper closes with suggestions of possible adaptation approaches for specialized information providers.

2 Social Information Behavior

Social Information Behavior can be divided into two aspects: Firstly, the receptive use of user-generated content as a resource to satisfy information needs. An example is the selection of Wikipedia entries in search engine result pages to answer ad hoc information needs. Secondly, active participation in Social Media: a) to answer current information needs, e.g. by asking a specific question in a question-answering service or a forum and b) to build up knowledge by ongoing conversations with peers in communities. The latter aspect can be connected to the concept of “communities of practice” (Wenger & Snyder 2000).

As described in chapter 1, participation in Social Media is mainly restricted to leisure-based contexts (Busemann & Gescheidle 2011). Nevertheless, according to an online panel survey of Gibs (2009) there is a segment of online users who see Social Media as a “core to finding new information”. Morris et al. (2011) surveyed 624 participants using social network sites and conclude that social networks are valued for their ability to provide opinions and recommendations. Kleimann et al. (2008) investigated the use of Social Media in educational contexts. Results indicate that a substantial fraction of students use social communities to communicate with peers about study related aspects. According to an investigation of Selwyn (2009), 4% of Facebook’s “wall activity” of 909 undergraduate students could be related to studies and academic aspects. Lee et al. (2012) explored the question of resource selection in academic search tasks and state that question-answering services accounted for 5.2%. Results of an interview study of students conducted by Hrastinski & Aghaee (2012) show that Social Media are rarely used for academic needs. According to the authors, Social Media are primarily used for short answers and questions and for coordination of group work in academic contexts. Kim et al. (2011) surveyed 446 undergraduate students and found that different Social Media types are used in different information seeking contexts. For instance,

social network sites are preferred for everyday life purposes whereas question-answering services are used for leisure as well as academic needs. Geist et al. (2012) investigated relevance and quality aspects of different types of search engine based search results for education-related information needs. They conclude that user-generated content, which represents 21% of the top 20 results, has a lower relevance probability than specialized information, but is on a par with results from professional information providers which do not have an explicit educational mandate.

This short overview of investigations in the field denotes that user-generated content and communities play an important role in information behavior. In addition, the named investigations illustrate the diversity of research questions which can and probably need to be addressed in the field of Social Information Behavior. Apart from factors that influence the choice of media, characteristics of use cases, communication patterns, and quality aspects (e.g. usefulness of content and outcomes) need to be addressed as well.

3 Research Design

Interestingly, investigations often restrict their analysis to popular sites like Facebook or Twitter (cp. section 2). Beyond this, it is often unclear which other “places” “build” the Social Web. This is where this research starts. We address the following research questions:

- A) Places and scale of the Social Web in the domain: Which domains and online services form the Social Web in the educational domain? How many users are actively engaged or subscribed in communication? What is the magnitude or volume of communication? Answers to these questions give an impression of the significance of Social Information Behavior in the field with regard to audience reach and amount of user-generated content.
- B) Dynamics and structure of communities: How do communities develop with regard to user numbers and contribution quantity? How are discussions structured? What are the topics of discourse? Such data give an indication of basic characteristics of communities in the field.
- C) Quality, pragmatics and success of communication: What type and pragmatics of information needs and answers constitutes online discussions? Are discussions helpful, e.g. do they provide actionable suggestions? What is the role and what are the characteristics of socio-emotional facets of communication? Insights into such attributes of online discussions allow for a judgment of the quality of online conversations with regard to aims, results and sequence of interaction.

Part A aims for a broad overview of the whole Social Web, B for a statistical assessment of specific communities and C for a judgment of threads. Because of the multiplicity of perspectives and the wealth of data under investigation, the implementation and employment of research methods and systems that help to answer the above mentioned research questions can be seen as a challenge for themselves. Therefore, the following arguments are not only of interest concerning measured data and results but also with regard to the methodological level.

4 Approach and Methods

The subsequent description follows the structure of the research questions. In principle, the methods employed, systems implemented and categories drafted can be regarded as to be independent from each other. In the analysis, they build upon one another. The output of methods addressing research question A partly serves as input data for B etc. The whole design follows a path from broad to narrow with regard to the analyzed samples and from general to specific concerning the depth of the analysis. The following illustration gives an overview of the research approach.

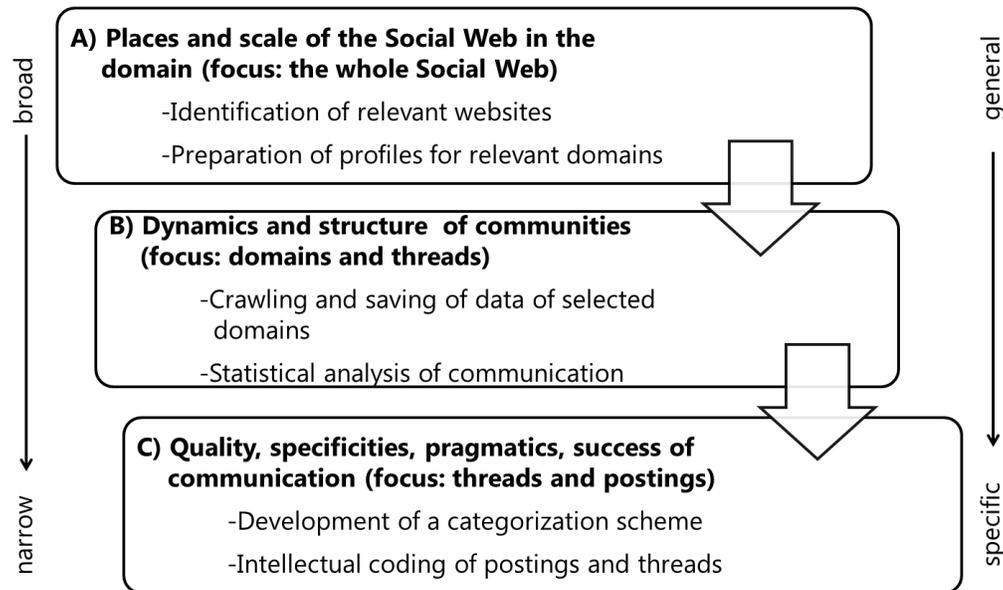


Figure 1: Research approach

A) Places and Scale of the Social Web in the Domain

In order to get an overview of the Social Web in the educational domain as a whole, the investigation starts with an identification of relevant websites. Following that, summaries of socio-demographic data and communication specific features of identified domains are prepared. To identify relevant websites, a set of 45 terms and phrases was used to query Google. The query set consisted of the most popular internal and external queries of the German Education Server (bildungsserver.de), a popular specialized information provider and search service in the field, as well as of names of education related study paths. For each query the top 100 search results were examined and pertinent forums, blogs, communities and portals were sighted in order to determine if they were suitable for further analysis. The sighted websites were selected for further analysis when the topics of these sites (or a subsection of the sites) matched topics of education related study paths and contained active discussions which related to educational topics as well as to the field of study. The results sample is what we define as the Social Web in the domain. To get an overview of the characteristics of this domain specific Social Web, profiles of each selected domain were prepared. The profiles encompass data with regard to their number of members, topics and contributions of each of these “places”. In addition, qualitative aspects were also evaluated. We specifically checked if discussions were moderated, and scanned the domain to identify its specific content focus and target group(s). If possible, the users with the largest numbers of contributions were also identified.

B) Dynamics and Structure of Communities

A subset of the identified Social Web, that means domains whose contents and volume of communication seemed to be especially interesting, was specified for further analysis with regard to communication dynamics and structure. The domains selected were Lehrerforen.de, Paedagogik-klick.de and Referendar.de. Here, we implemented a crawler based analysis tool. The tool has been developed on a Linux system and is based on the following Open Source software components: Cassandra (cassandra.apache.org), Nutch (apache.nutch.org), HTTrack (www.httrack.com), MySQL (www.mysql.com), and Nginx (nginx.com). Tools were configured with Java and Perl scripts. In the analysis we want to get an insight into the communication dynamics, seen from a community lifecycle perspective (Iriberry & Leroy 2009). Therefore, the number of new users and

contributions are analyzed over yearly time periods. This gives an insight into the “health” of the community. In addition, statistical data with regard to the number of posts per thread, post length, number of authors per thread and fraction of threads with only one posting give an indication of the basic characteristics of communication. A high percentage of rather short lived threads, or threads with a low number of replies can be seen as an indicator of probably shallow and not very successful communication which one would not connect with the notion of knowledge building communities (Wenger & Snyder 2000). Such an assessment can be underpinned with an analysis of the communication allotment of users. One can assume that the higher the mean of contributions per author, the deeper the connection to the community. Finally, discussion topics are investigated. A frequency analysis of the most frequent topic terms and their concordances gives an indication of the topical demands and needs of the communities, delivering insights in how far the discussions can be connected to specialized information.

C) Quality, Specificities, Pragmatics and Success of Communication

In addition to the statistical analysis, an in-depth content analysis is important to get insights into qualitative aspects, pragmatics and success of communication. In fact, only by triangulating both, one seems to be capable of delivering a meaningful picture. As intellectual assessment and coding is costly, a subset of 50 threads of two of the domains selected in B and another forum included in A will be analyzed for content analysis. Quality analysis of user-generated content is an active research field, in which both automatic and manual techniques are employed. For instance Agichtein et al. (2008) and Moturu & Liu (2011) aimed to quantify quality of contributions with the help of feature, usage or relationship (link structure) analysis based on automatic scoring and machine learning rating algorithms. Such methods could be aligned to our statistical approach in B. There is also active research based on intellectual analysis of user-generated content. E.g. Savolainen (2011) investigated criteria for assessing quality and credibility by a manual analysis of postings in Finish internet discussion forums. Results indicate that judgments from other messages could serve as an indicator of message content quality. Such mentions are primarily relying on author credibility and to a smaller extend focus on the judgment of the content itself. Willemsen et al. (2011) showed that there is a strong connection between content characteristics and perceived usefulness of Amazon online reviews. The focus of the analysis here is on the single contributions and, beyond that, on the development and characteristics of whole threads. Therefore, the categorization scheme needs to address the following levels: I) Initiation of the discussion, II) Course of the discussion and III) Outcome of the discussion. Concerning I and II the single contributions form the elementary units of the analysis. III is based on the judgment of the thread as a whole. As written, we are still in the process of finalizing the coding scheme and analysis system. The categorization scheme is developed collaboratively and iteratively by four researchers. The scheme encompasses the following categories:

I. Type of question: *fact oriented, personal estimation*

Intent: *Problem solving information need, uncertainty reduction, aim for suggestions, aim for emotional support*

II. Answer: *factual information, opinion, suggestion, further inquiry, gratitude, meta-discussion*

Socio-emotional aspect: *affirmation, opposition, social support, hostility*

Quality: *new topical aspect*

Topic

III. Explicit solution

As one can see, the coding scheme is complex. Judgments with regard to III (outcome) are also dependent on aggregated values of II. The system is tailored to the analysis of knowledge building processes. Especially the items *new aspect*, *further inquiry* and *opposition* can be connected to socio-cultural (exploration of knowledge, the diminishing of knowledge asymmetries), socio-cognitive (cognitive conflicts) and cognitive elaboration perspectives of learning and knowledge building (cp. Vygotsky 1979, Piaget 1979, Kop & Hill 2008, Scardamalia & Bereiter 1994). Following four cycles of development and testing, the current state of the categories is widely stable. Therefore we think, it is worthwhile to present some preliminary results from our analysis with regard to quality, pragmatics and success of communication, too. Still, one needs to consider data here is in a tentative stage.

5 Results

A) Places and Scale of the Social Web in the German Educational Domain

The query process to identify and select relevant websites for further investigation of education-related online communities revealed that most discussion sources are to be found within forums and question-answering services. Blogs and social networks barely appeared within the top 100 search results and thus were excluded from the analysis. In order to define the characteristics of the domain specific Social Web, profiles from a total of twenty-one German forums were selected (cp. table 1). These forums are exclusively or partly focused on users who aim at, are currently undertaking, or have already concluded their education related study paths. Question-answering services are for now excluded due to their broad topical scope.

Forum title	# of members in thousand	# of posts in thousand	Moderated forum	Registration for active participation	Members-only-area
paedagogik-klick.de	2	72.1	Yes	Yes	Yes
referendar.de	21	275.7	Partly	Yes	No
lehrerforen.de	15.5	304.7	Yes	Yes	No
4teachers.de	902.7	387.9	Partly	Yes	No
grundschultreff.de	1.7	59	No	Yes	Yes
fachlehrerseite.de	4	16.1	Yes	No	No
lehrerforum.de	3.9	12.8	Yes	Yes	No
schule-ratgeber.de	77.4	1	Yes	Yes	No
studis-online.de	No Data	602.8	No	No	No
uni-protokolle.de	216.5	1,908	Partly	Yes	No
uni-pur.de	No Data	4.7	No	No	No
studien-service.de	37	1,000	No	Yes	No
studieren-info.de	0.6	1.3	Partly	No	No
studiengang-verzeichnis.de	0.6	5.5	Yes	No	No

erzieherinnenausbildung.de	0.7	21.3	No	Yes	Yes
sowi-forum.com	13.8	333	Yes	Yes	Yes
krankenschwester.de	41.2	309.5	Yes	Yes	Yes
pflgeboard.de	34.6	180.6	No	Yes	No
e-hausaufgaben.de	212.6	1,422	Yes	Yes	No
vorhilfe.de	50.3	No data	Yes	Yes	Yes
dirk-bechtel.de	1.2	2.4	Yes	Yes	Yes

Table 1: Profiles of German forums in the educational domain with focus on education-related study paths (February - May 2013)

The forums are divided as follows: Fourteen forums aim at a broader target group with users of various backgrounds within the area of education, while seven forums focus only on students, pupils, and trainees.

Diverse characteristics can be seen in terms of the number of members, the total amount of topics and posts, and the communication principles. The lowest number of registered members is approximately 600 for studieren-info.de and studiengangs-verzeichnis.de. Contrary, 902,700 members are listed as members of 4teachers.de. A significant factor for a membership subscription seems to be the ability to actively participate in forum discussions, as 71% of the analyzed forums demand for log-in data in order to post or comment. A different approach can be seen for passively accessing forum content. While only 33% of the forums include certain topical areas that can be viewed by members only, the majority of the forums' content can be openly viewed or crawled by search engines at all times.

Generally, the number of members is proportional to the total amount of posts in a forum. Thus, forums with a larger scale of memberships have a higher variety of topics and posts in discussion. However, there is a discrepancy for four of the analyzed forums, where schule-ratgeber.de shows the highest variation with 77,400 members and a total of only 1,000 posts. In this forum, users need to log-in to benefit from uploaded lesson plan material in other parts of the website. Overall, there are four websites, which serve a material exchange next to their forum activity and demand. Three of these forums indicate a relative high number of members compared to the amount of total posts.

The forum with the highest member activity is 4teachers.de with about 7,000 posts composed by its most active member throughout his ten-year membership. The forum's foci are on topics related to study fields and lesson planning, include the possibility to share materials, and facilitate the exchange of experiences between students, as well as offering a student mentorship by employed teachers. Contrary, the forum studiengang-verzeichnis.de displays a member's list indicating that the highest number of comment contributions is of 12 posts. The forum contains approximately 5,500 posts and displays a total number of about 600 members. As there is no registration required for active participation, it is to infer that most comments are written by anonymous users.

An indication for the promotion of a long-term engagement of users is the employment of an actively participating moderator team. For fifteen of the twenty-one forums an at least partly intensive engagement of moderators can be seen in the discussions. These moderator-supported forums have about four times as many registered members compared to the non-moderated forums.

A more detailed profile inspection of lehrerforen.de, paedagogik-klick.de and referendar.de (cp. Fig. 2) shows that the three forums are varying in their characteristics:

- Lehrerforen.de with its 15,500 members and a total of 304,700 posts in 35,000 threads aims its discussions at students in education-related study paths as well as trainees and teachers. The forum

is supported by an active moderator team. It is openly accessible to non-members, however requires a registration prior to posting. The top ten members contributed an average of 3,600 posts and the main topical foci of the community are based on questions or discussions about teaching degrees and examinations, as well as lesson planning, didactic or educational counselling, and personal concerns.

- Paedagogik-klick.de is a comparatively small forum with about 2,000 members and 72,100 posts in 4,700 threads. The forum includes discussion areas for members only and is more intimate among its regular users, who in the top ten have posted about 4,000 posts throughout their membership. The forum is strongly moderated and influenced by a group of specialists who contribute to or lead discussions. The forum aims at various professional groups within the educational area, such as teachers, kindergarten nurses, educators, students and trainees as well as parents. The topics are based on education and learning, traineeship and career, expertise discussions, as well as various questions related to student lives.
- Referendar.de contains of 21,000 members and 275,700 posts in 27,000 threads. Contrary to Paedagogik-klick.de, Referendar.de focuses on a narrower target group of students in education-related study paths, trainees and young teachers. The forum is only partly moderated, in so far that there are six global moderators who supervise the threads but are only limitedly participating in the discussions. Forum topics encompass various questions, experience exchange and concerns around traineeships in the teaching environment. These not only include lesson planning, educational council and organizational matters, but also extend to financial questions, open job positions, etc. The forum's most active members have posted about 3,400 comments.

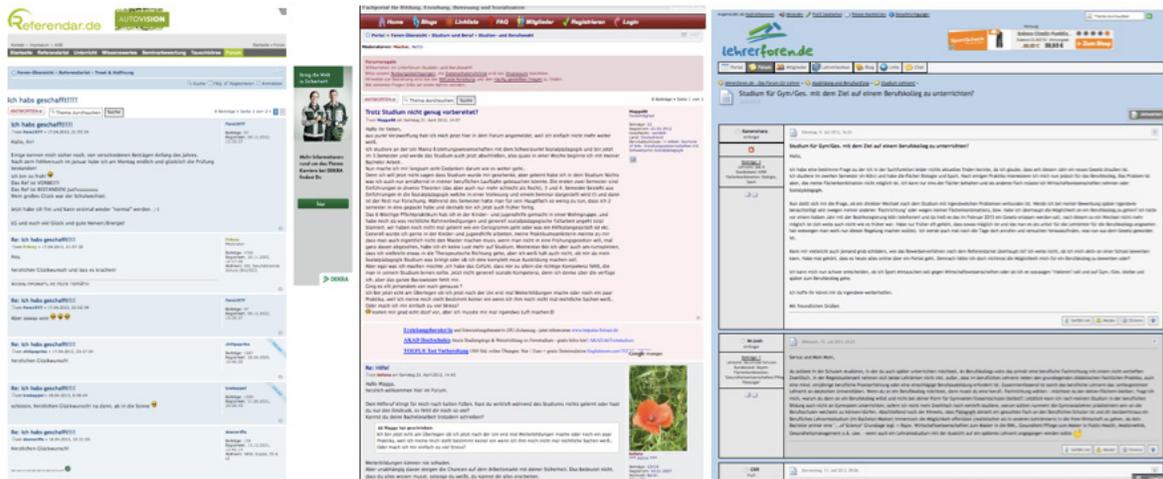


Figure 2: Thread examples for the forums Referendar.de, Paedagogik-klick.de, and Lehrerforen.de

In total, the 21 forums show a remarkable number of more than 1.6 million registered members (cp. table 2) and thus, along with an unknown number of anonymous forum users, play a significant role in the German educational domain in the Social Web.

# of forums	21		
Approximate # of members in thousand	1,637.3		
Approximate # of posts in thousand	6,920.4		
Moderated forum	Yes: 11	No: 6	Partly: 4
Registration for active participation	Yes: 16	No: 5	
Members-only-area	Yes: 7	No: 14	

Table 2: Overall figures of German forums in the educational domain with focus on education related study paths

The open access of forum topics for non-members and the possibility to crawl forum content through search engines supposes that forum content within the educational domain is not only perceived by members but widely accessible and relevant to non-members who have the opportunity to satisfy their information needs by passively consuming information. As a whole, approximately 6.9 million posts (cp. table 2) concerning topics related to educational themes are publicly available on the Internet, thus, the Social Web is highly significant for education-related information behavior.

B) Dynamics and Structures of Communities

This section presents the results of the statistical analysis of the three selected domains Lehrerforen.de, Paedagogik-klick.de and Referendar.de. After several months of implementation and fine tuning of the crawler, the final crawling was executed in June and July of 2013. To avoid duplicate content, filters were employed to exclude e.g. print-views. In addition, only selected sub-forums were indexed. Inclusion criteria were based on an explicit study or student-related focus of the sub-forums. Thus, for example, on Paedagogik-klick.de the sub-forum “*chats-The world outside of education!*” was excluded. Another example is the sub-forum “*School leadership and management*” on Lehrerforen.de. As a result, the sample of the statistical analysis is the subset of the content which is explicitly on-topic and aligned to the target group of students of education-related study paths. The following table gives an overview of the sample.

	Lehrerforen.de	Paedagogik-klick.de	Referendar.de	Sum
Topics	4,189	1,092	1,429	6,710
Posts	34,657	9,530	11,284	55,471
Active Users	2,901	470	1,693	5,064

Table 3: Sample overview

The data indicates a substantial amount of on-topic communication and also a significant number of contributors who actively express information needs or participate in problem solving and knowledge generation.

Communication dynamics

Crawling data shows that the communities already started years ago and therefore could be assessed as established domains. The oldest entry in the crawling index is from Lehrerforen.de, posted on October 25,

2002. The oldest post indexed from Referendar.de is dated on March 29, 2005. The oldest entry from Paedagogik-Klick.de was posted on January 7, 2007.

The following illustration gives an overview of new postings and participating authors per year.

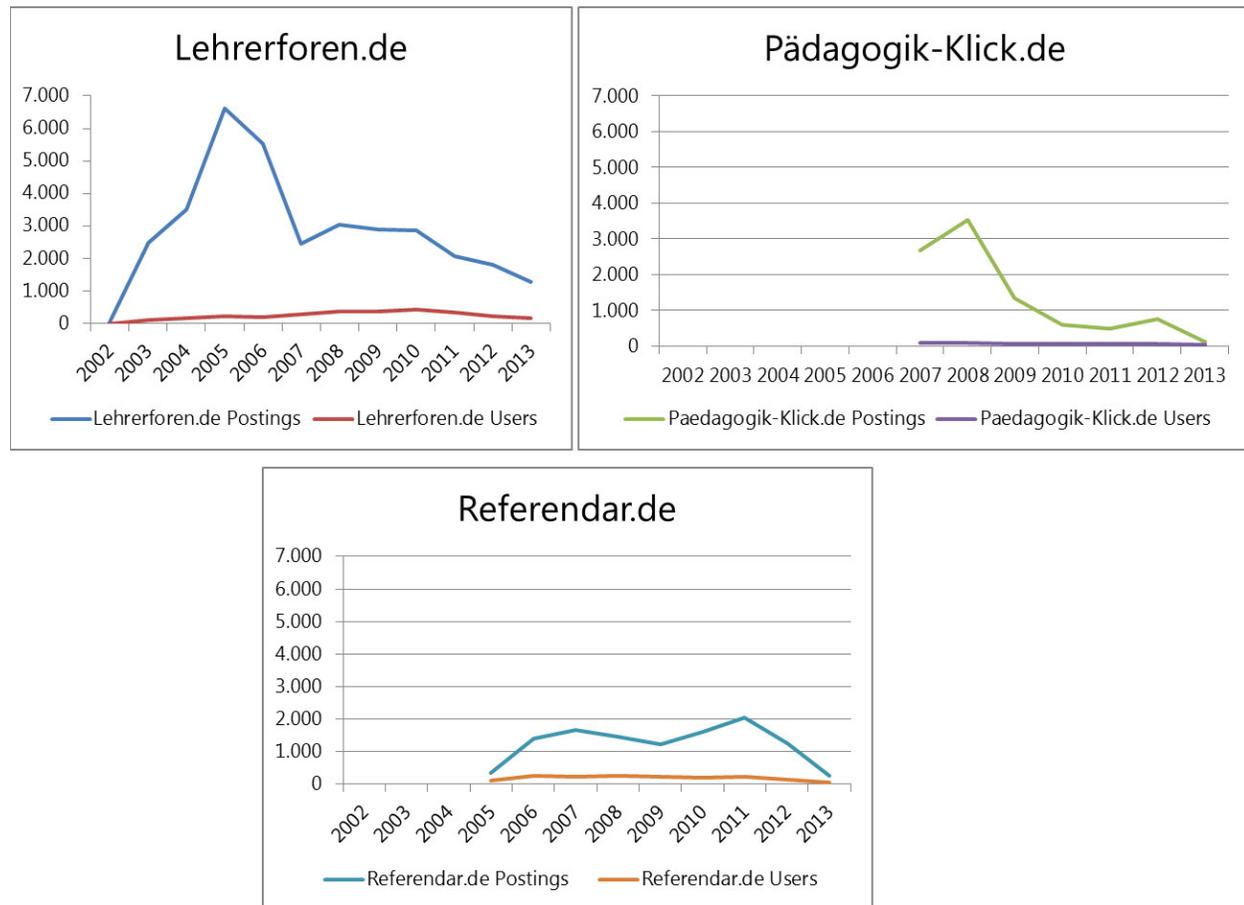


Figure 3: New postings and participating authors per year of the forums Referendar.de, Paedagogik-klick.de, and Lehrerforen.de

Data indicates that all three communities are beyond their peak of participation. In accordance with Iriberry's & Leroy's (2009) lifecycle perspective, one can argue that the communities are somewhere between a mature state and death. The reasons for the decline of participation are unclear. Probably there is a shift to other services or types of social websites, e.g. Social Networks. However, we do not want to paint a too bleak picture here. Still, there have been plenty of new contributions and new authors throughout the recent years. Nevertheless, one could recommend some suggestions to the operators of these domains, as argued by Iriberry & Leroy (2009). Operators could initiate events, provide permeated control and rewards, and establish and support new subgroups in their communities.

Communication structure

An analysis of the communication structure delivers first insights into the quality of discourse and knowledge building processes. A first summary of the data denotes substantial communication and exchange which can, at least on the surface, be connected to problem solving and knowledge building. The table below gives an overview of the communication structures within the three forums.

	Lehrerforen.de	Paedagogik-klick.de	Referendar.de	Sum	SD
Posts per topic	8.3	8.7	7.9	8.3	0.3
Posts per user	11.9	20.3	6.7	13.0	5.6
Length of postings in words	91.8	88.5	80.3	88.9	4.8
Fraction of threads with one posting (“dead threads”)	9.3%	16.6%	13.9%	11.5%	
Authors per thread	5.0	3.6	4.3	4.6	0.6

Table 4: Overview of communication structures

With a mean of roughly eight posts per thread over all forums, one can assume that communication exhibits some substance. Postings, too, are not only short statements. Instead, with a mean of 80-90 word length per posting they usually encompass several sentences. Concerning the number of authors per thread, it is to state that, in general, multiple perspectives are involved in communication. Although we get a relatively uniform picture across all three forums at large, here, and again with regard to posts per user and “dead threads”, results indicate some structural communication differences between the three forums, too. Over all forums, the number of posts per user hint to a relatively substantial participating level. In addition, the fraction of “dead threads” implies that in most cases the initiation of communication is successful. The number of authors per thread shows that there is a multiplicity of users and therefore perspectives involved in communication. The visible differences hint at a deeper community involvement of members of Paedagogik-klick.de, as the number of posts per author roughly doubles the quantity of contributions from members of Lehrerforen.de and is even three times higher in comparison to Referendar.de. In contrast, the initiation of discourse has a much higher probability of failure on Paedagogik-klick.de than in the other forums. We see this overview just as a first plunge into communication structure analysis. One can easily see the limits of the employed assessment. Firstly, there is no baseline for orientation. We think this work can be a start to build one. Secondly, more in-depth analysis and segmentation is needed. Here, more elaborate patterns of analysis are currently developed. Thirdly, the data can barely stand on its own and is not self-explaining. Triangulation with more in-depth qualitative data, as proposed in chapter 4c can lead to better insights into quality aspects and success factors of such communication.

Topics

Finally a frequency analysis of the top 100 terms of thread topics for the three communities implies a topical alignment towards professional aspects of (education for and work in) the education system (teacher, school, exam) and teaching subjects (Math, History, English, German). The titles indicate that the pragmatic of most discussions is either focused on the factual clearance of knowledge gaps with regard to certain named topics (“*child literature*”), the fostering of one’s own professional development (“*teacher or into the economy?*”) or suggestions to accomplish work-related tasks (“*lesson planning*”). The following table gives a rough draft of the discussion topics.

Lehrerforen.de	Paedagogik-klick.de	Referendar.de
Referendariat	Uni	Staatsexamen
Ref	Dipl	Bayern
NRW	Frankfurt	Examen
Examensarbeit	Austausch	StEx
Staatsexamen	Studis	Lehramt
Examen	Thread	Deutsch
Englisch	Ausbildung	Englisch
Deutsch	Erzieherin	Ergebnisse
Thema	Kinder	Geschichte
Klasse	Bildung	EWS

Table 5: Most frequent topic terms

The topics of Lehrerforen.de and Referendar.de are very similar and focus on the fostering of one's own professional development. Paedagogik-klick.de, the smallest community, has a broader topic spectrum and discussion topics often focus on didactical and pedagogical aspects of learning and teaching. As a whole, the topic analysis denotes a strong focus on topics that can be connected to specialized information and validates the results from the analysis in 5a.

C) Quality, specificities, pragmatics and success of communication

As written, the intellectual analysis with regard to the quality, specificities, pragmatics and success of communication is in a kind of pre-test stage. Nevertheless, we decided to present preliminary results to show first tendencies and also to reveal if our categorization scheme is able to capture knowledge building and the satisfaction of information needs.

The sample consists of six threads with a total of 60 posts collected from three forums. Two threads are selected from Lehrerforen.de, one thread from Referendar.de and three threads from Studis-online.de. Studis-online.de was chosen because it deviates from the other forums and also Paedagogik-Klick.de with regard to moderation and registration processes. In contrast to the other forums, Studis-online.de is not moderated and participation does not require prior registration (cp. table 1).

Initiation of discussion

The first interest is on the type of questions. Are they fact-oriented or asking for a personal estimation of the community? Data indicates both. An example for a fact-oriented question would be to ask for possibilities to use the same certificates for different kinds of academic grades. An example to ask for personal estimations would be asking for estimations if one's personal attributes would fit with being a teacher and therefore if one should study an education-related degree. Although these categories are not disjunctive, results indicate that both types of questions are usually not interwoven. In one case, a participant started a discussion with a narration of his exam nerves and a forthcoming examination date. He does not ask an explicit question but just states that he will keep the others informed of his further preparations. The second post in the thread showed social support. Then, the initiator expressed his gratitude. So, we can see this as an example for seeking emotional support without explicitly asking for it. In the whole corpus and also in the subset of the six selected threads for manual categorization, the length of the questions (measured in word counts) is far above the mean length of all posts. Thus the questions provide a richness of information need context that could be of interest to everyone who provides information services. With regard to intent, we detect all categories problem solving information need (#3), uncertainty reduction (#1), aim for suggestions (#1), aim for emotional support (#2). Interestingly, we found a combined occurrence of uncertainty reduction, aim for suggestions and aim for emotional support

and no overlap between emotional support and problem solving information need. Possibly, there is a distinction between questions that transport the personal pragmatics clearly and questions that are focused on cognitive problem solving.

Course of the discussion

With regard to answers, results indicate that all categories are present. Our data shows a mix of information (#17), opinion (#25), suggestion (#17), and further inquiry (8). Meta-discussion (#6) took place in half of the threads. With regard to gratitude (#4) we get the same picture. Emotional aspects and cognitive conflicts are also visible, but on a relatively low scale. We see no case of explicit affirmation and few oppositions (#4). This indicates a low level of cognitive discourse in the forums. Social support (#12) is more frequent than hostility (#6). New topical aspects (#25) are visible in nearly half of all posts. Therefore it seems that communication in these forums can be connected strongly to socio-cultural perspectives of knowledge building but only in few cases to socio-genetic perspectives. With regard to topic, fundamental changes throughout the discussion could not be detected. Instead, postings often broadened or narrowed the subject of the preceding post.

Outcome of the discussion

A discussion's outcome is often unclear. Only in one case, the initiator of the discourse explicitly confirmed that he is very grateful and that the discussion really helped him to reduce his uncertainty. Interestingly, the discussion still continued afterwards. Therefore, as written, the judgment of the outcome is often not directly measurable. Nevertheless, our categorization scheme allows for an estimation of emotional or cognitive values of threads. Emotional value can be approximated e.g. with the number of postings and authors expressing social support. A cognitive or knowledge building value can be calculated by the numbers and proportions of information, suggestions, opinions, further inquiries, affirmations and oppositions. Here, evaluation schemes and concrete measures still need to be worked out. Nevertheless, the data of our pre-test makes it clear that such an undertaking is principally feasible and very worthwhile in our opinion. Hence, further research will concentrate on this area.

6 Discussion

Finally, what are the results of this analysis? How can the research approach be categorized? What are the consequences for specialized information providers? To answer these questions, we will first summarize the data with regard to the research questions. Then, we will provide an estimation of the methods and an outlook for further research. After all, we will try to provide suggestions for specialized information providers.

With regard to the research questions, results can be summarized as follows. The investigation reveals the places and scale of the Social Web in the domain. Astonishingly, at least as the open Web is concerned, Social Online Networks and blogs only play a subordinate role. Visible educational-related communication is taking place in forums and question-answering services. Therefore, our analysis of forums reveals and describes the openly visible Social Web in the domain (cp. table 1). Here, the sheer amount of roughly 1.6 million registered users and approximately 6.9 million postings clearly depicts that participative Social Information Behavior is of relevance in the domain.

With regard to dynamics and communication structure of communities, data indicates that communication is substantial and that there is usually a multiplicity of perspectives involved in communication. Thus, these forums can indeed be assessed as a kind of knowledge building communities. Taking a deeper perspective, the content analysis indicates that knowledge building processes can be rather assigned to socio-cultural than socio-genetic perspectives of learning. In addition, socio-emotional aspects

play an important role in communication, as users are actively aiming for such support. At the same time, in most cases, the outcome of discussions seems to be unclear.

The research approach taken and methods employed in this investigation need to be assessed from two perspectives. On the one hand, we see a great news value. As argued, according to our knowledge it is the first attempt to measure the Social Web for a whole domain. The investigation is costly, but does not only produce new insights but also “products” that can be of value for information providers. Our overview data and crawler architecture can be used as a basis to build a Social Web search engine or Social Media monitoring system. Even our statistical analysis and the coding scheme provide new ideas to assess user-generated content. By employing a knowledge building perspective, data indicates that our coding scheme can be seen as a solid base to develop granular and elaborate measures to gauge the “knowledge building value” of threads. This is where our further research will focus on.

On the other hand, with regard to the statistical and intellectual analysis of domains, threads and communities, research instruments are still in a preliminary stage. In case of the in-depth manual content analysis, data collection is, at the time of submitting this article, not very advanced and far from being finished. Therefore results here are only preliminary.

Finally, with regard to suggestions for specialized information providers, first propositions were already presented in the paragraph above. In our crawl with roughly 55,000 indexed postings, the term “bildungsserver” (in English: education server) appears in 53 of them, that is in one of a thousand postings. If we keep in mind that the education servers play an important role in the infrastructure of specialized information providers in Germany in the field, then it is to conclude that they are disconnected from the open Social Web. So, what can be done about this? Maybe it is time to get in touch with each other. With our data published here, specialized information providers can already grasp the German Educational Social Web as a whole. There is also data with regard to influencers available. Concerning strategies of social media communication, one can distinguish three possible approaches. The one option which was already mentioned is to provide users with an access to the knowledge in the Social Web in the domain, e.g. by building up a specialized search engine. Secondly, to get a grasp of users’ needs, trends and developments in the Educational Social Web specialized information providers could employ systematic Social Media monitoring. Again, our crawler could serve as a blueprint here, but there also is a variety of other tools and services available. A third approach would be an active participation in the communication. Providers could get in contact with operators and influencers in forums or build up their own specific service-based profiles and get involved in communication.

7 References

- Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). Finding high-quality content in social media. *Proceedings of the International Conference on Web Search and Web Data Mining* (pp. 183-194), Palo Alto, California, USA, February 2008,.
- Allfacebook.de (2013). *Facebook Nutzerzahlen*. Retrieved August 06, 2013, from <http://allfacebook.de/userdata/>
- Busemann, K., & Gscheidle, C. (2011). Web 2.0: Aktive Mitwirkung verbleibt auf niedrigem Niveau. Ergebnisse der ARD/ZDF-Onlinestudie 2011. *Media Perspektiven*, 7-8/2011, 360-369.
- Croll, A., & Power, S. (2009). *Complete Web Monitoring*. Sebastopol, CA: O’Reilly.
- eMarketer (2013). *Users Seek Out the Truth in Online Reviews*. Retrieved August 06, 2013, from <http://www.emarketer.com/Article/Users-Seek-Truth-Online-Reviews/1009656>
- Evans, D. (2010). *Social Media Marketing. The Next Generation of Business Engagement*. Indianapolis: Wiley.

- Geist, K., Griesbaum, J., & Mahrholz, N. (2012). User-generated content in web search: relevance and quality aspects of search results for education-related information needs. *Proceedings of LWA 2012 – Lernen – Wissensentdeckung – Adaptivität*, Dortmund, September 2012.
- Gibs, J. (2009). *Social Media: The Next Great Gateway for Content Discovery?* nielsenwire. Retrieved November 09, 2012, from http://blog.nielsen.com/nielsenwire/online_mobile/social-media-the-next-great-gateway-for-content-discovery/
- Hester, A. J. (2010). Increasing collaborative knowledge management in your organization: characteristics of wiki technology and wiki users. *Proceedings of the 2010 Special Interest Group on Management Information System's 48th annual conference on Computer personnel research* (pp. 158-164), New York: ACM.
- Hrastinski, S., & Aghae, N. (2012). How are campus students using Social Media to support their studies? An explorative interview study. *Education and Information Technologies*, 17(4), 451-464.
- Iriberry, A., & Leroy, G. (2009). A life-cycle perspective on online community success, *ACM Comput. Surv.*, 41(2), 1-29.
- Kasper, H., Dausinger, M., Kett, H., & Renner, T. (2010). *Marktstudie: Social Media Monitoring Tools. IT-Lösungen zur Beobachtung und Analyse unternehmensstrategisch relevanter Informationen im Internet. Technical report*. Stuttgart: Fraunhofer-Institut für Arbeitswirtschaft und Organisation.
- Kim, K., Yoo-Lee, E., & Sin, S. (2011). Social Media as Information Source: Undergraduates' Use and Evaluation Behavior. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1-3.
- Kleimann, B., Özkilic, M., & Göcks, M. (2008). *Studieren im Web 2.0. Studienbezogene Web- und E-Learning-Dienste. HISBUS-Kurzinformation 21*. Hannover: HIS Hochschul-Information-System GmbH.
- Kop, R., & Hill, A. (2008). Connectivism: Learning theory of the future or vestige of the past? *The International Review of Research in Open and Distance Learning [Online]*, 9(3). Retrieved August 06, 2013, from <http://www.irrodl.org/index.php/irrodl/article/view/523/1103>
- Lee, J., Paik, W., & Joo, S. (2012). Information resource selection of undergraduate students in academic search tasks. *Information Research*, 17(1), paper 511. Retrieved November 09, 2012, from <http://InformationR.net/ir/17-1/paper511.html>
- Morris, M. R., Teevan, J., & Panovich, K. (2010). What do people ask their social networks, and why? A survey study of status message Q&A behavior. *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 1739-1748), New York: ACM.
- Moturu, S.T., & Liu, H. (2011). Quantifying the trustworthiness of social media content. *Journal of Distributed Parallel Databases*, 29(3), 239-260, June 2011.
- O'Reilly, T., & Battelle, J. (2009). *Web Squared: Web 2.0 Five Years On*. Web 2.0 Summit, San Francisco, CA.
- Piaget, J. (1979). *Judgement and reasoning in the child*. London: Routledge.
- Ramakrishnan, R., & Tomkins, A (2007). Toward a People Web. *Computer*, 40(8), 63-72.
- Savolainen, R. (2011). Judging the quality and credibility of information in Internet discussion forums. *Journal of the American Society for Information Science*, 62(7), 1243-1256, July 2011.
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences*, 3, 265-283.
- Selwyn, N. (2009). Faceworking: exploring students' education-related use of Facebook. *Learning, Media and Technology*, 34(2), 157-174.
- Shirky, C. (2010). *Cognitive Surplus*. New York: Penguin Press.
- Shirky, C. (2008). *Here comes everybody*. New York: Penguin Press.

- Statistisches Bundesamt (2012). *Studierende an Hochschulen. Wintersemester 2011/2012*. Fachserie 11, Reihe 4.1. Wiesbaden: Statistisches Bundesamt.
- Vygotsky, L. (1979). *Mind in Society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wenger, E., & Snyder, W. (2000). Communities of Practice: The Organizational Frontier. *Harvard Business Review*, 78(1), 139-145.
- Willemsen, L.M., Neijens, P.C., Bronner, F., & de Ridder, J.A. (2011). Highly Recommended! The Content Characteristics and Perceived Usefulness of Online Consumer Reviews. *Journal of Computer-Mediated Communication*, 17(1), 19-38.

8 Table of Figures

Figure 1: Research approach	359
Figure 2: Thread examples for the forums Referendar.de, Paedagogik-klick.de, and Lehrerforen.de	363
Figure 3: New postings and participating authors per year of the forums Referendar.de, Paedagogik-klick.de, and Lehrerforen.de	365

9 Table of Tables

Table 1: Profiles of German forums in the educational domain with focus on education-related study paths (February - May 2013)	362
Table 2: Overall figures of German forums in the educational domain with focus on education related study paths	364
Table 3: Sample overview	364
Table 4: Overview of communication structures	366
Table 5: Most frequent topic terms	367