# Facebook as a Source for Human-Centered Engineering: Web Mining-Based Reconstruction of Stakeholder Perspectives on Energy Systems

*Bianka Trevisan[1], Denise Eraßme[1], Tim Hemig[1], Sylvia Kowalewski[1], Johanna Kluge[1], Simon Himmel[1], Anna Borg[2], Eva-Maria Jakobs[1] and Martina Ziefle[1]*

*[1]Institute of Linguistics and Communication Studies*
*HCI-Center, RWTH Aachen University*
*Campus-Boulevard 57, 52062 Aachen, Germany*

*[2]CBM Gesellschaft für Consulting, Business und Management mbH*
*Niederbexbacher Straße 67, 66450 Bexbach, Germany*

## ABSTRACT

In this paper, a new approach in acceptance research is presented analyzing Facebook-user profiles and text data of stakeholders for purposes of human-centered engineering. Thereby, data from topic-specific German Facebook-pages are collected by means of a self-developed Facebook-parser and analyzed semi-automatically with Web Mining-methods applying a multi-level annotation scheme for sentiment analysis. The aim of the analysis is to get insights about (1) who is discussing energy systems in the Web, (2) how users evaluate energy systems (positive, negative, neutral), and (3) which gender-related differences in user discussions (acceptance factors) can be observed. As an application example, the renewable energy form deep geothermal energy is used, which has increasingly become subject of public discussion in the context of the German energy turnaround. The results of the Facebook-study show that in human-centered engineering of energy systems gender-sensitivity is of major importance: Men are primarily focusing on the overall economic efficiency and the environment protection, women pay more attention to costs and benefits that affect themselves. Thus, engineers of complex energy systems such as geothermal energy should therefore take into consideration the aspect of divergent gender perspectives.

**Keywords**: Web Mining, Facebook, acceptance research, human-centered engineering, energy system, stakeholder perspective, gender diversity, deep geothermal energy

## INTRODUCTION

The paper reports new ways of acceptance research by using and adapting Web Mining methods. The basic idea is to exploit user-generated content as access to stakeholder perspectives on complex energy systems. Therefore, Facebook-data is analyzed with the aim of identifying gender-related differences in technology evaluation. The study is part of the project TIGER that investigates the public awareness and acceptance for geothermal systems by combining methods such as questionnaire, focus group, and Web Mining.

The paper aims at answering the following research questions: Who expresses him- or herself judgmentally about renewable energies such as deep geothermal energy on the Internet (stakeholder)? By what features are these stakeholders characterized (profiles)? What aspects of energy systems are perceived and claimed acceptance-relevant in Internet discourses? How important is geothermal energy in comparison to other forms of energy production?

Web Mining has its origin in computational linguistics. Up to now, this approach hasn't been applied as a method in acceptance research. The study reported here is based on a Text Mining approach described in Trevisan and Jakobs (2010). The approach allows identifying how persons and social groups perceive and evaluate complex technology systems that mean to identify evaluation objects, aspects, and value criteria. In Trevisan et al. (2012), the approach is refined by a linguistic-based multi-level annotation model (MLA). The model allows for a holistic analysis of evaluative statements, where different evaluative means are considered and relevant linguistic information related to each other. A shortcoming of the presented method is that the stakeholders behind the statements can't be described and identified (age, gender, education, etc.) because they usually publish content anonymously under a nickname. However, stakeholders' demographic data is needed as a precondition for the triangulation of focus group, questionnaire, and Web Mining-results (Arning et al., 2013).

The paper presents an approach to solve the problem by using Facebook-messages, since stakeholders provide many demographical data about themselves on social networking sites. Thereby, stakeholder profiles and comments from different German Facebook-page types are elicited and analyzed, e.g., company pages, group pages or news.

The paper is structured as follows. First, the principal approach of human-centered engineering in acceptance research is introduced, which includes a specification of the objectives of the associated project TIGER as well as an outline of the methodological challenges and opportunities of previous related research. Second, the methodological design for the stakeholder analysis with Web Mining methods is described. Third, the results are presented. They concern user profiles of stakeholders on topic-specific Facebook-pages as well as stakeholders' evaluation on renewable energies. Fourth, the results are discussed. Lastly, a conclusion and an outlook are given.

# HUMAN-CENTERED ENGINEERING IN ACCEPTANCE RESEARCH

## TIGER: Project's objectives

2007 the EU Heads of State and Government agreed on the climate formula "20-20-20". According to this formula, $CO_2$ emissions should be reduced by 20%, the share of renewable energy increased to 20% and energy efficiency raised by 20% until the year 2020. Germany is a pioneer in achieving the climate protection goals and invests increasingly in renewable energy forms (solar, biomass, etc.). Anyway, the achievement of these goals and the associated energy turnaround *(Energiewende)* lack of acceptance for renewable energies such as *deep geothermal energy*. This causes socio-economic costs that have to be avoided, contemporary.

However, up to now, approaches are still missing which conceptualize the perception and evaluation of complex energy systems and, thereby, take different interacting levels into account (social, cultural, individual- and group-based, but also technical and economic variables) as well as different stakeholder-perspectives (user vs. non-users, target groups vs. decision makers) and associated conflicts. In the interdisciplinary project *TIGER - Deep Geothermal Energy - acceptance and communication of an innovative technology,* a multi-methodological approach is designed, combining methods from social sciences (in-depth interview, questionnaire) and computational linguistics (Text and Web Mining). In applying this approach, group- and location-specific advantages and disadvantages of deep geothermal energy are investigated; the relevance and relative weighting of the identified acceptance factors is determined by the triangulation of method-related results (see Figure 1):

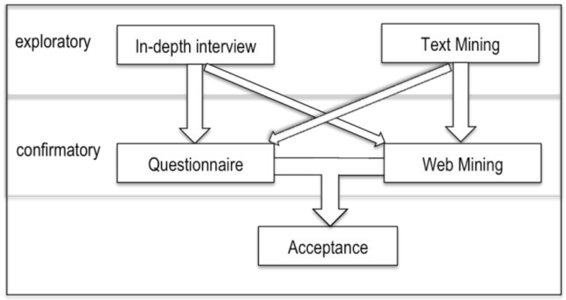


Figure 1. Multi-method approach

The results can be compared on the content level, i.e., the response frequency per acceptance factor is compared per method and database. However, the occurrence of an acceptance factor cannot be related to a target group, as, up to now, target groups can only be determined with traditional methods (in-depth interview or questionnaire) (Arning et al., 2013). Thus, the methodological gap, respectively, the identification of related stakeholders or target group-members remains a challenging task in acceptance research and in the respective project.

## Methodological challenges and opportunities

**Challenges**: Up to now, stakeholders in the Web remain hidden as most of them use a nickname for their activities on the Web (e.g. in blog comments) instead of their real name. Rarely, more information can be found beyond the nickname such as age, gender, education, etc. One way to close this methodological gap is to guess by some indicators stakeholder profiles, e.g., by *metadata* (nickname, posting periods) (Trevisan 2014/in press), *posting frequency* (allows for the identification of the activity type, active vs. passive user) or *linguistic profile* (refers to users formulation style, colloquial vs. standard style) (Jakobs et al., 2014/in preparation). For instance, the nickname can provide hints on the user's gender (e.g. *Bill09*); linguistic characteristics can give information about the educational background in the sense of colloquial (e.g. *Cool!*) or standard German (e.g. *Very good!*). Neunerdt et al. (2013) sketched the usefulness of *linguistic characteristics* for the identification of user profiles, already. However, this "presumed" data can neither be considered as a prerequisite for the modeling of stakeholder profiles nor is it commonly considered valid for purposes of data triangulation.

**Opportunities**: In this paper, a new methodology for stakeholder profile elicitation and modeling is presented. To date, Text Mining-studies in acceptance research are based on blog or news comments (text data) that do not deliver stakeholder-related information (Trevisan and Jakobs, 2010, Trevisan et al., 2013). The present study uses Facebook-data (text data: hosts and user comments; metadata) as a source for acceptance research. An advantage of the presented approach is that the stakeholder's profile behind the statements can be described (gender, age, marital status, education, university graduation, profession, etc.). Thus, shortcomings of recent Text Mining-studies in acceptance research can be solved.

The presented methodology is based on data elicited from topic-specific German Facebook-pages dealing with *(deep) geothermal energy*. *Topic-specific pages* serve, in general, on the one hand as an information platform regarding a specific topic, on the other hand as a topic-specific discussion forum. The discussion is held between the host(s) of the topic-specific page (host 1-n) and interested users (user 2-n). Thereby, a host posts a statement in the manner of an article or cites an article from an extern source, which he comments. In return, users comment on the posted article. In doing so, they provide statements and evaluations related to the initial post or article, which might be of neutral, positive, or negative polarity. Similarly, referencing in online discourses is described for blogs by Hoffmann (2010) and modified for discourses in topic-specific blogs by Trevisan (2014/in press). Moreover, in Facebook both – the host and the user (as reader) – are able to positively express and/or reinforce their statement by "liking" a specific part of a discussion, e.g., previous posted comment or the initial post. Figure 2 depicts in chronological order the structure of postings (hosts' comment on the posted article, article, stakeholder or user comments) of a random Facebook-page.

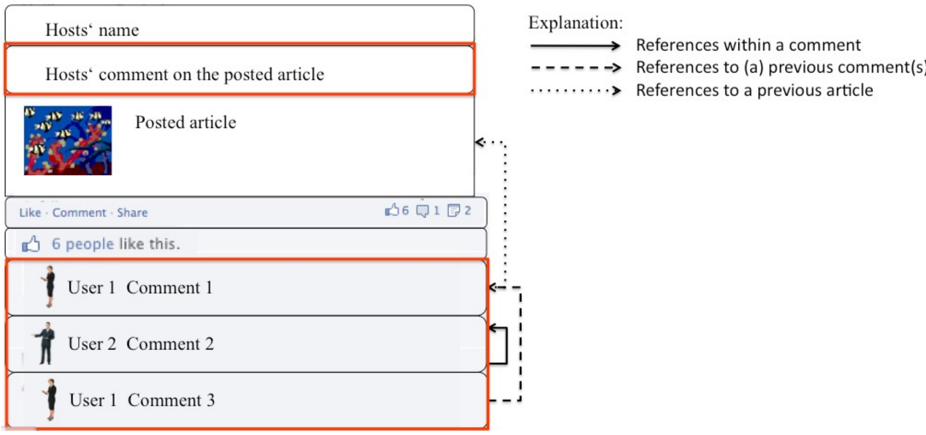


Figure 2. Postings (article, comments and related stakeholders) in chronological order representing a random interest group (Adapted from Trevisan, 2014/in press)

In this paper, the focus lies on hosts' and user comments, as framed in Figure 2.

# METHODS: STAKEHOLDER ANALYSIS WITH WEB MINING

## Data collection

To build up a data corpus, it was searched for topic-specific German pages on Facebook dealing with deep geothermal energy. The search was led by two criteria: (i) the topic of the page should relate to geothermal energy and/or renewable energies; (ii) more than 100 persons should be members of the group in order to gain a sufficient amount of data. The data – text data and metadata – is collected from the source code, semi-automatically (time period: 2009 – 2013).

**Text data:** First, in a pre-processing step each post or hidden comment has to be opened manually by click on *more*. Second, a self-developed Facebook-parser collects the manually prepared data. The *Facebook-parser* is implemented in Java; using the parser, text can be collected automatically from the source code of Facebook-pages. The text data is stored relatively: for each article found, the article text and all related comments are collected and converted to plain text (.xml). Internet lingo-specific elements such as emoticons had to be recognized and converted separately due to their representation in the source code (as an image vs. composed of punctuation marks). Methodological challenges concern the reliable extraction of the data from Facebook-pages. The challenge refers to the fact that text data in Facebook exists in different formats, i.e., articles are formatted differently in the source code as comments.

**Metadata:** Using the *Facebook-parser*, metadata can be collected automatically from the source code of Facebook-pages. Thereby, the number of collected metadata is limited (1) by the number of metadata made available by the website (*here:* Facebook) and (2) by the amount of metadata information that was provided by the user himself. Some users have an extensive user profile including information about the marital status, education, university graduates, or profession; others tend to give only little information about themselves, in many cases, the minimum information required: the name.

Finally, our database consists of 634 user profiles from 13 different Facebook-pages. Table 1 shows the distribution of users per gender and Facebook-page type.

Table 1. Distribution of users per gender and Facebook-page type (FB=Facebook, m=male, f=female)

| FB-page | | FB-user | | |
|---|---|---|---|---|
| type | Σ | m | f | Σ |
| company page | 2 | 52 | 32 | 84 |
| group page | 5 | 80 | 62 | 142 |
| interest page | 1 | 32 | 5 | 37 |

| | | | | |
|---|---|---|---|---|
| news | 2 | 30 | 6 | 36 |
| organization | 3 | 245 | 90 | 335 |
| Σ | 13 | 439 | 19<br>5 | 634 |

Five types of Facebook-pages are distinguished: company page, group page, interest page, news and organization. In this study, the page types are defined as follows:

*Company page:* Manually generated page by a company, e.g., an energy supplier. Internal events and news in the business context of the company are published and put up for discussion such as latest scientific findings on *geothermal energy*.

*Group page:* Manually generated page by a stakeholder. Generally, stakeholders with a common interest (e.g. geothermal or renewable energy) use the virtual space for discussion.

*Interest page:* Automatically generated page based on the interests of Facebook-users. The page is not connected with or endorsed by anyone related to this topic.

*News:* Manually generated page, e.g., by the publisher of a magazine. The aim is to announce current and independent information, for instance, on the use of *geothermal energy*.

*Organization:* Manually generated page by an association, e.g., in the area of *geothermal energy*. Main tasks of the association are to inform the public, to encourage the dialogue between public and policy makers, and to improve the framework conditions continuously.

## Data processing

**Text data**: The text data is processed in five successive methodological steps. First, the text data is formatted for further analysis into .txt. Second, the txt-formatted data is tokenized by means of an adapted TreeTagger-tokenizer (Neunerdt et al., 2013). Third, the tokenized data is PoS-tagged using WebTagger (Neunerdt et al., 2013). *WebTagger* is a POS-tagger developed for the morpho-syntactic annotation of Web comments. Fourth, the automatically tokenized and tagged data is manually corrected. The corrected data build a basis for the upcoming processing steps, the multi-level annotation in EXMARaLDA. Fifth, the pre-processed text is annotated with a multi-level annotation scheme applying the annotation model depicted in Trevisan et al. (2012) and Trevisan, (2014/in press).

The model is originally used for the enhancement of automatic sentiment analysis and opinion detection in German blog comments. The annotation model consists of different annotation levels with various purposes of annotation, such as the identification of evaluative utterances. In the present Facebook-study, the data is annotated semi-automatically on four selected annotation levels: *graphemic level, lexico-semantic level, polarity level, and pragmatic level*. At the *graphemic level*, expressions at the text surface as well as grapho-stilistic features that show special notational styles are annotated such as emoticons or quotation marks (e.g. *"[a successful drilling]"*). At the *lexico-semantic level*, topic-specific terms are recorded, e.g., geothermal energy-related nouns such as *drilling hole*. At the *polarity level*, three types of annotation levels or scopes are distinguished: a level for the analysis of polarity indicating *tokens* (a word such as a noun or an adjective, e.g. *displeasure*), *multi-tokens* (a grammatical phrase such as a noun phrase, e.g. *the earthquake-risk area*), and *sentences* (e.g. *Geothermal energy is a pioneering technology*.). Here, the polarity indicating tokens are annotated. Finally at the *pragmatic level*, information is given about the evaluative substance of the evaluative utterance, e.g., someone has the intention to *BLAME*, *PRAISE*, *CRITICIZE* something. Thus, each evaluation-indicating token is, in addition to the automatic annotation by the WebTagger, enriched with topic-specific and evaluation-indicating functions. Figure 3 shows an extract of the multi-annotated text data (in EXMARaLDA).

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **X [txt]** | Absolut | passt | das | hier | rein | . | Das | gehört | in | die | Kategorie | " | Energieeffizienz | " | ;-) |
| **X [pos]** | ADJD | VVFIN | PDS | ADV | ADJD | $. | PDS | VVFIN | APPR | ART | NN | $( | NN | $( | $. |
| **X [lemma]** | absolut | passen | d | hier | rein | . | d | gehören | in | d | Kategorie | " | Energieeffizienz | " | ;-) |
| **[GraphemicLevel]** | | | | | | | | | | | | MARK | | MARK | EMO |
| **[Lexico-semanticLevel]** | | | | | | | | | | | | | EL | | |
| **[PolarityToken]** | + | + | | | + | | | | | | | | + | | + |
| **[PragmaticLevel]** | IRONIZE | | | | | | IRONIZE | | | | | | | | |

Figure 3. Extract of the multi-annotated text data (in EXMARaLDA)

In the example two ironic utterences are given (*Absolut passt das hier rein. Das gehört in die Kategorie "Energieeffizienz"./ Absolutely fits the purely here. This belongs in the category of "energy efficiency".*). Thereby, irony is indicated by graphemic means such as the quotation marks (Tag: MARK) or the twinkling emoticon (Tag: EMO, *;-)* ) at the end of the second utterance. The quotation marks enclose the topic-specific positive connotated (Tag: +) word *Energieeffizienz* (Tag: EL, *energy efficiency*), whereby the meaning of the word *and* the utterance is reversed into its opposite.

Lastly, the identified topic-specific terms are categorized. For this purpose, semantically related words are organised in lexical fields (e.g. RADIATION). *Lexical fields* summarize partially synonymous words and words who are “associated in experience” (Croft and Cruse, 2004; see also Fillmore, 1985), e.g., COSTS{energy costs, euro, expenses}.

**Metadata**: For each collected text data a metadata file is created in plain text (.xml). The metadata (file) provide context-, user- and content-related information, such as information about the user’s age provided by the birthdate (e.g. 4.1.1983), the publication date of the text data (e.g. 12. November 2013), or the text type (e.g. comment). The sequence of the metadata in the file is set by the analyst, manually. Figure 4 shows an example for a metadata file.


```
<file>FB_Linksammlung_Erneuerbare_Energien_2013-05-23_04-07-23_350_article</file>
<project>FuEne</project>
<textsorte>Article</textsorte>
<media>Website</media>
<method>FBParser</method>
<source>
    <name>Facebook</name>
    <url>../FuEne/Facebook/Geothermie/Linksammlung_Erneuerbare_Energien.html</url>
</source>    <date>2013-05-23 04-07-23</date>
<timestamp>1369318043</timestamp>
<created>2013-11-12 04-58-28</created>
<topic>Energy</topic>
<title>(2) Linksammlung Erneuerbare Energien</title>
<author>
    <name>[redacted]</name>
    <url>https://www.facebook.com/gerhard.koudela?hc_location=stream</url>
</author>
<tags>
    <tag>Erneuerbare Energie</tag>
</tags>
<feedback>
    <shared>0</shared>
    <likes>2</likes>
</feedback>
```


Figure 4. Metadata file. Note: the name of the author was blackened from data protection legal reasons

For the following metadata analysis, each .xml-file is imported in a .csv-file.

## Data analysis

**Text data:** For data analysis, text data from one Facebook-group page is analyzed. The database consists of 828 articles and comments (approx. 27.600 token). Table 2 shows the distribution of comments and tokens per gender.

Table 2. Distribution of comments and tokens per gender (m=male, f=female)

| | f | m | Σ |
|---|---|---|---|
| Comments | 98 | 730 | 828 |
| Token | 1.829 | 24.895 | 26.724 |

The preprocessed data is analyzed by frequency and sentiment. In the *frequency analysis*, it is counted (i) how often which topic-specific term (e.g. drilling hole) is mentioned in the entire corpus *(total ranking)* and (ii) which lexical field respectively category is mentioned most often *(relative ranking)* (see for a detailed description of the ranking procedure Trevisan and Jakobs, 2010). In the *sentiment analysis*, it is determined which topic-specific terms occur with which positive, negative, or neutral connoted words within an evaluative utterance, e.g., fault-prone geothermal. The aim is to find out how components and features of energy systems such as deep geothermal energy are evaluated and which reference objects (gas, oil, etc.) are mentioned.

**Metadata:** The metadata are analyzed quantitatively. Aim of the quantitative analysis was to determine the distribution of gender, age, marital status, education, university graduates, and profession among the elicited user profiles. The quantitative results are transferred in percent.

# RESULTS

## User profiles of stakeholders on topic-specific Facebook-pages

The overarching aim of the Facebook-study is to identify gender-related differences in technology evaluation. However, results of the metadata analysis show that not all users provide their complete demographic data on Facebook. Women seem to have a more sophisticated data protection awareness compared to men. While 27% of male Facebook-users have an incomplete profile, 41% of female Facebook-users share fragmentary data with the Facebook-community. Table 3 gives an overview about gender-related differences in data protection awareness respectively the amount of users' profiles with missing demographic data.

Table 3. Gender differences in data protection awareness and users' profiles with missing data (m=male, f=female, n=occurrence)

| | m | f |
|---|---|---|
| n | 120 | 80 |
| Percent | 27% | 41% |
| ∑ | 439 | 195 |

Nevertheless, the analyzed user data provides insights about users or stakeholders' gender, age, marital status, education, university graduates, and profession; missing demographic data (e.g. age, gender) are classified as "unknown". Thus, results are presented in the following two diagram types: (1) the first diagram shows the ratio of the queried class in proportion to the unknown and (2) the second diagram only the known variables.

First results indicate that most users specify their gender ($n_f$=30%, $n_m$=69%): only 1% of the Facebook-members do not provide their gender on their user profile (see Figure 5). Generally, the non-disclosure of gender represents the lowest proportion of unknown data throughout the whole Facebook-study results (see Figure 6, 7, 8, 10). In addition, the gender distribution among the Facebook-members demonstrates that mainly men are members of energy systems-focusing Facebook-pages ($n_f$=31%, $n_m$=69%).

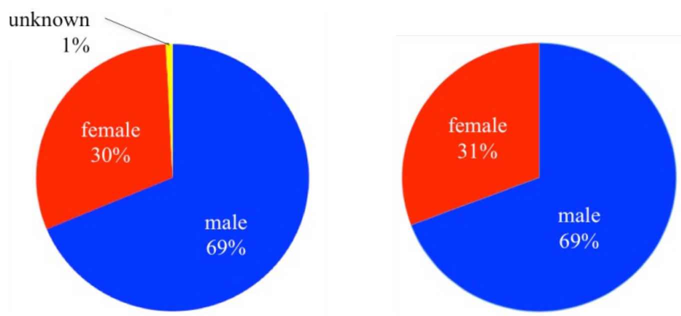


Figure 5. Gender distribution. Left: with unknown gender. Right: without unknown gender

Regarding the age distribution it is evident that the highest proportion accounts the group of Facebook-members in the range between 20 and 40 years (46% // 56%, see Figure 6). Second most frequent, the Facebook-members are in the age range between 40 and 60 years (29% // 35%). The proportion of over 60-years-old is relatively low (6% //

7%) as well as the proportion of under 20-years-old is vanishingly small (1% // 2%). In 18% of the elicited profiles, the age is unknown, which forms the second lowest proportion of unknown data, totally (see Figure 5, 7, 8, 10).

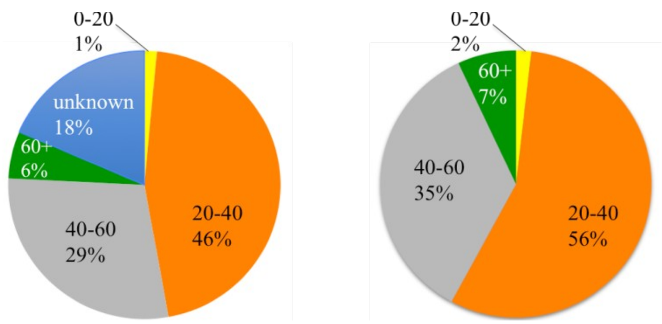


Figure 6. Age distribution. Left: with unknown age. Right: without unknown age

Regarding the marital status, only 16% of the elicited Facebook-members shared this information on their profile (unknown=84%) (see Figure 7). Among the users with known marital status, the majority is married (52%), 26% are in a relationship, and 20% are singles. 1% of the users are divorced or live apart.

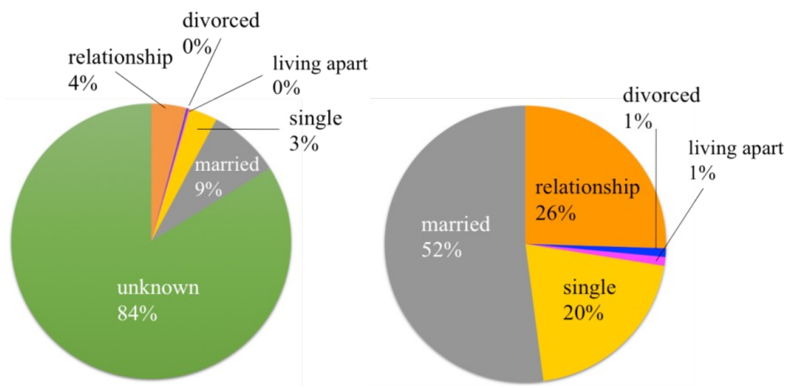


Figure 7. Distribution of marital status. Left: with unknown martial status. Right: without unknown martial status

The distribution of graduates shows that the majority (61%) of the Facebook-members disclosed this information for the community (see Figure 8). However, among the given information 56% of the users are college graduates and 44% high school graduates. Thus, the distribution is almost equal.

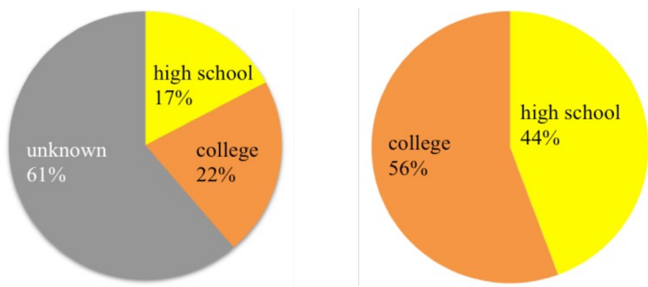


Figure 8. Distribution of graduates. Left: with unknown graduates. Right: without unknown graduates

Thereby, among the graduates a predominate percentage are male Facebook-members 78% ($n_f$=22%) (see Figure 9).

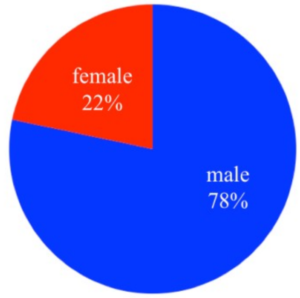


Figure 9. Gender distribution among graduates

In 68%, the job position of the Facebook-members is unknown (see Figure 10). Among the knowns, the distribution is almost equal: 46% are employees, 56% are self-employed.

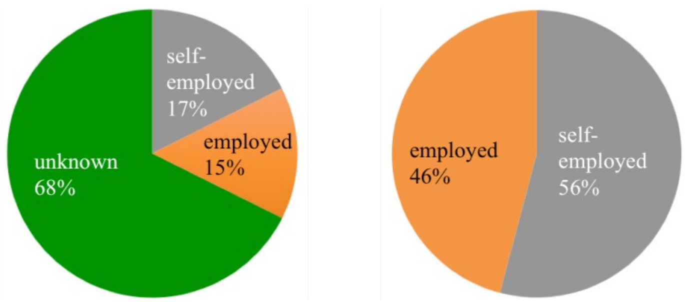


Figure 10. Distribution of users by job position. Left: with unknown job position. Right: without unknown job position

Regarding the distribution of jobs per sector the profiles show that only 32% share information about their job (unknown=68%) (see Figure 11). Nevertheless, the data reveal that the majority of the Facebook-members on energy systems-focusing pages are employed or self-employed in other sectors (66%) than the energy sector (34%).

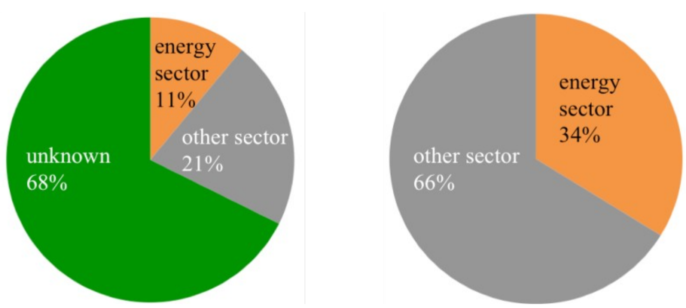


Figure 11. Distribution of jobs per sector. Left: with unknowns. Right: without unknowns

## Stakeholders' evaluation on renewable energies

Regarding the results of the frequency analysis it has to be considered, generally, that the corpus proportion is substantially higher among men than among women (see Table 2). Therefore, the results of the frequency analysis always have to be rated in terms of their overall weight. In the following, the results of the total and relative ranking, the comparison objects and the sentiment analysis results are presented.

With regard to the results of the *total ranking* it is evident that in the male text data *energy* (n=66), *electricity* (n=55) and *energy turnaround* (n=53) are by far mentioned most frequently (see Table 4). In the female text data, *euro* (n=6) is mentioned most frequently, even though the distribution over all terms is balanced. Comparing the term occurrence between men and women it is shown that only two terms emerge in the comments of both genders. These are *energy* ($n_m$=66, $n_f$=3) and *euro* ($n_m$=24, $n_f$=6).

Table 4. Total ranking of the most frequently mentioned terms (top 10) (m=male, f=female, n=occurrence)

| m | | | f | |
|---|---|---|---|---|
| **Term** | **n** | | **Term** | **n** |
| energy | 66 | | euro | 6 |
| electricity | 55 | | costs | 4 |
| energy turnaround | 53 | | heating system | 4 |
| profit | 31 | | heating costs | 3 |
| subvention | 30 | | house | 3 |
| billion | 25 | | energy | 3 |
| euro | 24 | | savings | 3 |
| money | 22 | | payback period | 3 |
| wood | 18 | | residential ventilation | 2 |
| heat | 18 | | renovation | 2 |
| ∑ | 342 | | ∑ | 33 |

With regard to the results of the *relative ranking* it is clearly visible that almost all identified lexical fields of the female users occurred on the male side except BENEFIT (n=6), IMAGE (n=3), and DISADVANTAGE (n=3) (see Table 5). In total, the most discussed topic or lexical field for both gender is COSTS ($n_m$=131, $n_f$=26). The lexical field COSTS subsumes, for instance, the *financing* of energy form expansion, increased *duties for taxpayers*, or increased *prices for electricity*. Highly ranked lexical fields, which only occurred in males' discussions are SUPPLY (n=46), SUBSIDY (n=42), or LOBBYISM (n=21). The second most mentioned lexical field between male users is PROFIT (n=106), which was also mentioned in females' discussions, but with a much lower weighting to the overall discussions (n=2). PROFIT subsumes concepts such as *compensation for electricity fed into the grid*, *economic efficiency*, or *market share*. Furthermore, a high priority at the males the ENVIRONMENT (PROTECTION) (n=51) and at the females the lexical fields PROPERTY (n=13) and ENERGY EFFICIENCY (n=9).

Table 5. Relative ranking of identified lexical fields (m=male, f=female, n=occurrence)

| m | | | f | |
|---|---|---|---|---|
| **Lexical field** | **n** | | **Lexical field** | **n** |
| COSTS | 13 | | COSTS | 26 |
| PROFIT | 10 | | PROPERTY | 13 |
| ENVIRONMENT (PROTECTION) | 51 | | ENERGY EFFICIENCY | 9 |
| SUPPLY | 46 | | BENEFIT | 6 |
| SUBSIDY | 42 | | DAMAGE | 6 |
| EMISSION | 34 | | PLANT | 5 |
| PLANT | 28 | | EMISSION | 4 |
| LOBBYISM | 21 | | ENVIRONMENT | 4 |
| CONSUMPTION | 20 | | IMAGE | 3 |
| FUEL | 19 | | PERFORMANCE | 3 |
| WATER | 18 | | DISADVANTAGE | 3 |
| WASTE | 18 | | PROFIT | 2 |
| ENERGY EFFICIENCY | 17 | | CONSUMPTION | 2 |
| PRODUCTION | 16 | | FUTURE | 1 |
| DAMAGE | 15 | | | |
| EMOBILITY | 11 | | | |
| PROPERTY | 10 | | | |
| AGRIBUSINESS | 10 | | | |
| NEED | 10 | | | |
| RISK | 8 | | | |
| FUTURE | 8 | | | |
| PERFORMANCE | 6 | | | |
| SECURITY | 6 | | | |
| ∑ | 65 | | ∑ | 87 |

Regarding *comparison objects* of geothermal energy, the users mentioned nine different energy forms. Table 6 shows an overview of the mentioned energy forms.

Table 6. Reference objects of (deep) geothermal energy (m=male, f=female, n=occurrence)

| **Gender**<br>**Energy form** | **m** | **f** | **∑** |
|---|---|---|---|
| solar energy | 38 | 4 | 42 |
| nuclear energy | 37 | 4 | 41 |
| wind energy | 29 | 6 | 35 |
| biogas | 23 | 2 | 25 |
| biofuel | 18 | | 18 |
| coal | 14 | 1 | 15 |
| natural gas | 5 | 2 | 7 |
| oil | 7 | | 7 |
| pumped-storage plant | | 1 | 1 |
| ∑ | 172 | 20 | 191 |

Thereby, the most mentioned are *wind energy* (n=42), *nuclear energy* (n=41), and *solar energy* (n=35). Since much more text data from male users is elicited (see Table 2), the number of entries for each energy form turns out much higher among male users ($n_m$=168 vs. $n_f$=20). Overall, the most frequently energy form named by men is *wind energy* (n=38), among the female users *solar energy* (n=6). Moreover, some energy forms are only mentioned from one gender such as *biofuel* and *oil* (from male users) and *pumped storage plant* (from female users).

The results of the *sentiment analysis* show that, based on the comments of the *male users*, renewable energies compared to fossil energies are evaluated more positive. One exception is biomass. Biomass is controversial discussed, in general. Proponents emphasize the energy efficiency; opponents criticize the use of food for energy production, especially in the context of food run shorts in third world countries. Also negatively evaluated is lobbyism, the greed of energy suppliers, and the lack of environmental awareness of the energy companies. In this context, it is proposed that the energy companies participate in the costs due to environmental pollution, e.g., the elimination of nuclear waste. In addition, positively evaluated are the progress of new technologies and innovations in energy supply, the renewable energy forms solar and wind energy, and the decentralized marketing of energy in favor of lower electricity prices, especially for low-income earners.

In the women comments it is evident that, particularly, the own benefit of renewable energy is evaluated, i.e., the energy efficiency of solar energy or the installation of a pellet heating in their homes. Also positively evaluated is the energy turnaround, the decentralization, and wind energy. Negatively evaluated are fossil energy resources such as coal and nuclear power.

# DISCUSSION

The results show, that in human-centered engineering of energy systems gender-sensitivity is of major importance. As evident from the results, men and women consider different evaluation aspects relevant. While men are more focusing on the overall economic efficiency (costs, profit) of renewable energies and the environment protection, women pay more attention to their costs and benefits (property, energy efficiency) that affect themselves, e.g., in the utilization of geothermal energy for power and heat generation.

Comparing the results to previous studies, it is apparent that only few discussion aspects identified in Facebook-articles and -comments emerge in other studies such as cost, wind energy, or efficiency. For instance, Kowalewski et al. (2014, in this proceedings) mentioned beside environmental aspects unknown costs as one of the most important barriers in the context of deep geothermal energy; in addition, Trevisan et al. (2013) named the *heat flow* or the *dealing with citizens by the municipality* as relevant.

Comparing the results with the German national average (www.destatis.de) it is recognizable that the overall distribution precipitates differently in the case of Facebook-data. In the national average, the gender distribution is balanced (2012: m=48,9%, f=51,1%), within the Facebook-data the number of male clearly outweighs the number of female users (m=69%, f=31%). The same applies for the age groups: in the national average the percentage per age group is balanced (2011: <25 y.=23.9%, 25-45 y.=25.8%, 45-65 y.=29.3%, 65+ y.=11.8%); in contrast, Facebook-users are mainly in the age of 20 to 40 years. In contrast, the proportions for the marital status agree most closely with the Facebook-data assuming that the Facebook-category "in a relationship" (26%) and "single" (20%) correspond to the category "single" (42.3%) of the national average. 52% of Facebook-users spent themselves as married in comparison to 42.4% of the national average (state: 2011). Lastly, 56% of the analyzed Facebook-users are college graduates, whereas only 14% of the national average reached in 2012 a college degree (applied university or university) or doctorate. This suggests that a very special group of people is interested in the discussion on renewable energies such as geothermal energy and that a discussion over the national average would certainly bring forth a different focus and new respectively unknown acceptance factors.

Regarding the methodological approach, through the analysis of the Facebook-data valuable insights on gender-specific evaluation of energy systems such as the renewable energy source (deep) geothermal energy are gained. Nevertheless, the data allow only partially statements about the reasons for claiming acceptance inhibiting and promoting factors, as user profiles are in most cases incomplete. A more accurate picture and scientifically sound statements could be achieved through the collection and analysis of further Facebook-profiles and -comments (representative data corpus).

# CONCLUSION AND OUTLOOK

The previously unsolved problem of missing demographic data in Internet discourses could be achieved by the use of Facebook-data. Thereby, the data allow statements about stakeholder profiles of users who express themselves judgmental about renewable energy sources such as deep geothermal energy. Moreover, the related text data (articles and comments) is used for the identification of acceptance inhibiting and promoting factors. Therefore, Text and Web Mining-methods such as the multi-level annotation scheme for sentiment analysis are applied, successfully. Future research will focus on the analysis of further Facebook-data and the triangulation of method-specific interdisciplinary results, e.g., from questionnaires, interviews, or focus groups with Web Mining-results. It is expected that by triangulating, a knowledge gap can be closed, which concerns previously unidentified acceptance factors.

# ACKNOWLEDGMENTS

This research was funded by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety.

# REFERENCES

Arning, K., Trevisan, B., Ziefle, M., Jakobs, E.-M. (2013), *"Eliciting user requirements and acceptance for customizing mobile device system architecture"*, in: Proceedings of the 15th HCII. pp. 439-448

Croft, W., Cruse, D. A. (2004), *" Cognitive Linguistics"*. 1st ed. Cambridge: Cambridge University Press.

Fillmore, Charles J. (1985), *"Frames and the semantics of understanding"*, in: Quaderni di Semantica 6. pp. 222-254

Hoffmann, C. R. (2010), *"From Monologue to Dialogue? Cohesive Interaction in Personal Weblogs"*. Ph.D. thesis, Augsburg University.

Jakobs, E.-M., Digmayer, C., Trevisan, B. (2014/in preparation), *"IBK als multifaktoriell geprägtes sprachliches Handeln"*, in: Reihe Empirische Linguistik. Akademie-Verlag

Kowalewski, S., Borg, A., Kluge, J., Himmel, S., Trevisan, B., Eraßme, D., Ziefle, M., Jakobs, E.-M. (2014), *"Modelling the influence of human factors on the perception of renewable energies. Taking geothermics as an example"*, in: Proceedings of the 1st International IBM Symposium on Human Factors, Software, and Systems Engineering (AHFE 2014)

Neunerdt, M., Trevisan, B., Reyer, M., Mathar, R. (2013), *"Part-of-Speech Tagging for Social Media Texts"*, in: Gurevych, I., Biemann, C., Zesch, T. (Ed.). pp. 139-150

Statistisches Bundesamt, www.destatis.de (23.02.2014)

Trevisan, B. (2014/in press), *"Bewerten in Blogkommentaren. Mehrebenenannotation sprachlichen Bewertens"*, Ph.D. thesis, RWTH Aachen University

Trevisan, B., Eraßme, D., Jakobs, E.-M. (2013), *"Web Comment-based Trend Analysis on Deep Geothermal Energy"*, in:

Proceedings of the IPCC 2013
Trevisan, B., Neunerdt, M., Jakobs, E.-M. (2012), *“A Multi-level Annotation Model for Fine-grained Opinion Detection in Blog Comments”*, in: Proceedings of KONVENS 2012. pp. 179-188
Trevisan, B., Jakobs, E.-M. (2010), *“Talking about Mobile Communication Systems. Verbal Comments in the Web as a Source for Acceptance Research in Large-scale Technologies”*, in: Proceedings of the IPCC 2010. pp. 93-100