Extracting Knowledge about Cognitive Style

The Use of Sensory Vocabulary in Forums: A Text Mining Approach

Gudrun KELLNER
Institute of Software Technology and Interactive Systems
Vienna University of Technology
Vienna, Austria
gudrun.kellner@tuwien.ac.at

Bettina BERENDT
Department of Computer Science
K.U. Leuven
Leuven, Belgium
bettina.berendt@cs.kuleuven.be

Abstract:

According to psycholinguistic research, any text contains a lot of implicit information about its writer. The Internet provides an incredible amount of text produced by users. The information potential of texts that can be directly linked to a user (which is especially the case for forum and blog posts) is not yet sufficiently examined. Natural language processing techniques are in some cases used for getting information on the user's personality, moods, affects and sentiments, but not yet for getting information on the user's cognitive style concerning sensory preference. This paper explores the potential of such an idea.
Therefore, a corpus with more than 1.000.000 forum posts was analyzed for the occurrence of expressions that are directly linked to a sensory system. We found that users differ significantly in their use of sensory expressions and that most users have preferred patterns for the use of sensory expressions. Furthermore we found a correlation between the sensory vocabulary of a post and the sensory preference of the users who answered this post.

Keywords:
Cognitive Style, Sensory Preference, User Modeling, Text Mining, Lexical Analysis

I. INTRODUCTION

Digital media permits to present information in manifold ways. An adequate mode of information presentation helps the user with information processing, e.g. user A prefers to read new information whereas user B definitely prefers listening to it. Research shows that information can be more easily understood if its presentation is adapted to the cognitive preference of the user that needs to be reached. One aspect of cognitive preference consists in the preferred mode of perception. Such information might be of interest for every user model used in a setting where the user's interest needs to be captured and/or the user's process of information perception and organization shall be supported. Therefore, it might be an interesting extension for adaptive hypermedia, especially for the field of flexible content and interface design. Such personalization could be useful for a wide range of applications including, but not limited to, e-commerce and e-learning.

But how can such a preference be elicited? In the field of user modeling, there exist a huge number of models and systems that perform very well without any direct interaction with the user in form of questionnaires or tests. It would thus not make much sense to suggest a new variable for user models that needs intense interaction with the user for being elicited. For this reason, we propose to use existing text, written by the target user and published in the internet, as a resource of information on his or her perceptional preference, and to use natural language processing techniques for text mining and lexical analysis. We tested our approach with a forum corpus that contains more than one million forum posts.

To ground further investigation in the relation between the use of sensory vocabulary and sensory preference, several basic questions about sensory word use in forums need to be answered: a) Is sensory vocabulary used sufficiently often to be considered as an indicator for individual preference? b) Does the use of sensory vocabulary stay consistent within the posts of one user? c) How can the use of sensory vocabulary be modeled?

First results on the given topic were published in [17]. This paper extends our previous work by a literature review, two more hypotheses, and a more detailed analysis of data.

The paper is organized as follows: Section II searches for existing answers to those questions in related research. Section III explains the methods we used to find our own answers to those questions. Section IV presents the results, which are discussed in section V. Section VI provides a conclusion and an outlook to further research.

II. RELATED RESEARCH

This section explores related research. First, we take a look at user models with a focus on personality and/or preferences. Second, we give an overview of cognitive styles, especially those that reflect upon sensory preference. The third part shows in what manner text may be used as a source of implicit information about its author.

A. User models with a focus on personality/preferences

In the field of user modeling, two sorts of variables are modeled in order to describe a user: rather permanent and rather transitory attributes. Permanent attributes stay the same over a long time and can concern user personality, preferences, or fields of interest. Transitory attributes are state-dependent, may change after a short period of time, and represent moods, plans and goals. The choice and number of variables used in such a model are strongly dependent on the field of application

and the purpose of the user model. User models are usually designed for a specific goal, e.g., to make people learn more easily or buy more. One method to evoke such effects is to personalize the choice, amount and presentation mode of information.

Two subfields of user modeling should be considered within this context: The subfield of cognitive user modeling combines psychological theories and representations about aspects of human cognition. Different applications for this field are described in [15]. The subfield focusing on educational application seeks to collect information about learner characteristics like knowledge, skills, and personality traits in order to select the best learning environment for a particular student to optimize learning outcome [26]. In this paper, we suggest a new variable for the field of cognitive styles which combines aspects of cognitive user modeling and user modeling in e-learning environments. We think that this variable could be a useful extension for the subfields of cognitive user modeling and e-learning.

B. Cognitive style

Cognitive style "may be defined as an individual's consistent approach to organising and processing information during thinking" [25]. In education, cognitive or learning styles describe different preferences in how learners perceive and retain information. Quite some models have one or more attributes that express sensory preferences: Several learning style models describe dimensions of information representation that can be classified as sensory preference for visual input as opposed to verbal input (which might be considered as preference for auditory input if one assumes content is communicated orally): e.g., the dimension "imager vs. verbalizer" as proposed by *Riding* [9; 25], or the dimension "visual vs. verbal" as proposed by *Felder and Silverman* [11]. *Dunn and Dunn*'s model includes the "modality preferences" visual, auditory, kinesthetic and tactile preference [9; 10]. An application of this concept to the field of language learning was published by *Reid* [23]. In her model on learning styles she differentiates between four basic perceptual learning channels, visual, auditory, kinesthetic and tactile modalities, and presents a questionnaire focusing on the learner's perceptual preference and his or her preference for individual vs. group learning. According to [10], students achieve significantly better learning results when the students' perceptual preference is matched with educational methods. As this concerns mainly the processes of receiving and processing information – and therefore processes that we would like to evoke and ease for our users – it might make sense to use this concept in a broader context.

In the field of e-learning, several methods are used to obtain information on the user's learning style: besides the "classical" approach via questionnaire, the learner's behavior in web based learning [2; 6] or learning context features [8] can be used as indicators to detect learning styles. We propose to follow the tradition of psycholinguistics by analyzing the individual's use of expression in order to deduce perceptual preference.

C. Using text as a source of implicit information on its author

Every text contains a message that might sometimes be easily understood, but sometimes needs more context to be interpreted in the right way. A text also reveals a great deal about its author. In recent years, the field of web text mining has done considerable research on what texts can tell us about authors by analyzing spontaneous utterances of users in texts such as blogs and forums.

To retrieve information on the user's affects, words and terms expressing such are collected in an affective lexicon and used as a basic instrument to determine expression of affect. *Valitutti et al.* published a model of lexical representation of affective knowledge using WordNet [27; 28]. Information on the users' sentiments and opinions is gathered analogically [3]. Methods in this field are quite well developed: specialized algorithms have been developed to handle different sorts of text available on the internet concerning average text length and genre, such as blogs [1], news [12] or social networks [20], giving only some examples for each category. There exist already fusion approaches [20] and multilingual solutions [5]. The user's mood might also be influenced by the topic he or she is writing about [4; 18]. There exist also some, but considerably fewer attempts to derive information on the user's personality from text. Based on lexical analysis, [21] extracts information on personality differences like neuroticism, extraversion, openness, agreeableness, and conscientiousness out of blog texts. [22] evaluates word use in depression forums in English and Spanish.

III. METHOD

Our work proposes an extension to existing user models. We present a modeling technique that implicitly acquires sensory vocabulary data by analyzing forum text.

A. Sensory lexicon

There is no real agreement on the number of sensory systems one should take into account in the field of perception research [13; 14; 19]. We decided to use the most common model and postulate: Sensory systems S = {visual, auditory, kinesthetic, olfactory, gustatory}

We created our lexicon of sensory vocabulary as follows: It is based on the list of stems of sensory vocabulary collected by [16]. We expanded this list by incorporating a list of color terms. We worked with four categories of sensory vocabulary: visual, auditory, kinesthetic and olfactory/gustatory. Olfactory and gustatory expressions were combined because of the high overlap of words within those two fields. This overlap might be explained by the similar body regions that are active when perceiving odor and taste, wherefore they are often referred jointly as "chemical senses" [13]. Hence, the lexicon of sensory vocabulary may be divided into 4 disjoint sets, namely the lexicon of visual vocabulary LV, of auditory vocabulary LA, of kinesthetic vocabulary LK, and of olfactory and gustatory vocabulary L{OG}. The cardinalities of these sets are $|LV|$=87, $|LA|$=92, $|LK|$=112, $|L\{OG\}|$=65. Naturally, this list does not yet cover all existing terms in the German language. As this is

a very first attempt, we concentrated more on precision than on recall.

B. Measures

The use of sensory vocabulary in a document p is expressed as a four-dimensional vector $vprofile(p) = [p.V, p.A, p.K, p.OG]$ based on the frequency of sensory vocabulary per sensory system. Thus, the visual component of the vector is defined as

$$p.V = \frac{|\{ t \epsilon\ p' \mid v(t) = 1\}|}{|\{ t \epsilon\ p' \mid v(t) = 1\ or\ a(t) = 1\ or\ k(t) = 1\ or\ og(t) = 1\}|} \quad (1)$$

where t are the terms in p', which is p modeled as a bag of words, and $v(t)$ (etc.) are the sensory indicators of the term:

$$indicator\ level\ s(t) = 1\ \text{if stem(t) is in the lexicon of} \quad (2)$$
$$\text{sensory vocabulary}$$
$$= 0\ \text{in all other cases.}$$

Profiles were not only calculated for each document (forum post), but also for each author by concatenating all posts of this author to one new pseudo-document and calculating its profile as described in (1). The similarity $vsim(p1,p2)$ between two posts was measured as the cosine similarity of their profiles.

C. Hypotheses

We think that every user has one or two preferred sensory systems that influence his or her choice of expressions. In this paper we are not (yet) trying to prove the relation between a preferred sensory system and the use of sensory expression. Being the first to investigate this topic, we focus instead of the use of sensory expression itself in order to develop a sense of the potential of such an approach.

To evaluate whether such a preference within the use of sensory vocabulary exists, three assumptions need to be made and will be tested with our hypotheses: (a) The choice of expression stays stable as long as the discussed topics remain within the same field. (One can expect that, e.g., in a discussion about photography there is an overall higher use of visual expressions than in a discussion about cooking. To bypass this problem, we chose a forum corpus on a very narrow topic: a forum dedicated to the discussion of the car BMW model E30.) (b) Individuals have different preferences regarding their choice of sensory expression, there is no complete dominance of one sensory system over all others (as sometimes stated, e.g., under the heading of the "age of visuality"). Also within a certain pattern of preferences, there may be differences concerning the intensity of the preference. (c) The use of sensory vocabulary cannot be referred directly to the use of overall vocabulary.

These three assumptions may be combined in the following statement: Variance in the use of sensory expression is higher between texts written by different users than between texts written by the same user. This assumption might be altered by a tendency to adjust one's personal style of expression to the conversational partner. Both options cannot be explained only with the use of overall vocabulary. We propose to test this idea with 3 hypotheses which examine different parameters:

Hypothesis A states that every user has a preference profile for sensory modalities, expressed as a profile of usage of sensory vocabulary. Hence, posts written by one author should be more similar to one another than to posts written by somebody else. Of course, this will not hold for every two such pairs of posts, but only on average. We therefore posit as Hypothesis A:

$$avg_{p1, p2\ \epsilon\ P,\ p1.author = p2.author}\ vsim(p1,p2) > \quad (3)$$
$$avg_{p1, p2\ \epsilon\ P,\ p1.author \neq p2.author}\ vsim(p1,p2)$$

where the pi are posts, P is the set of all posts, $pi.author$ is the post's author, and $vsim$ is the similarity between the VAKOG profiles of its two arguments. We restrict the comparison to those posts that have a distinguishable VAKOG profile in the first place, operationalized as a total number of VAKOG words larger than or equal to 3.

Hypothesis B states that the answers to a post p1 tend to have the same distribution of sensory vocabulary as p1. Again, we expect this to hold on average, restrict attention to posts with at least 3 VAKOG words, and state with a notation analogous to that in (3):

$$avg_{p1, p2\ \epsilon\ P,\ p2.refersTo = p1}\ vsim(p1,p2) > \quad (4)$$
$$avg_{p1, p2\ \epsilon\ P,\ p2.refersTo \neq p1}\ vsim(p1,p2)$$

Hypothesis C varies Hypothesis B. It starts from the assumption that an author will feel "attracted" to answer a post if this post matches his/her general sensory preferences, but that (for various reasons) the produced answer post may deviate from this general pattern. This leads to the expectation that the answering author's profile (rather than the answer post's profile in Hypothesis B) will correspond to the answered post. We restrict attention to those authors who have a distinguishable profile in the first place, operationalized as a total number of posts larger than or equal to 5:

$$avg_{p1, p2\ \epsilon\ P,\ p2.refersTo = p1}\ vsim(p1,p2.author) > \quad (5)$$
$$avg_{p1, p2\ \epsilon\ P,\ p2.refersTo \neq p1}\ vsim(p1,p2.author)$$

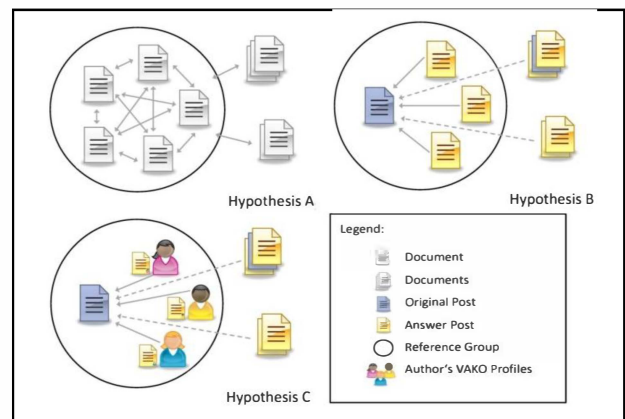An overview of the 3 hypotheses and their methods of testing is given in Fig. 1.

Figure 1. Hypotheses A–C and their testing

D. Data

Text Corpus: To test the hypotheses, we chose *Richling*'s forum corpus [24] which can be accessed freely after registration via web interface.[1] It is a corpus built on posts from discussion forums on the car type BMW E30, published in the years 2000 until 2007. This very narrow topic helps to exclude result variation due to discussion of different topics. The corpus is monolingual and consists of more than one billion posts in German, each post text accompanied by information on the author, the header, the reference post, and the date.

Data preparation and analysis procedure(s): The following four steps were executed in order to obtain analyzable data:

(1) We used Unix shell and PHP scripts to store the forum corpus, initially plain text with an xml-like tagging structure, in a MySQL database.

(2) The algorithm proposed by [16] for identifying and counting sensory vocabulary in text splits a text into chunks of words. Each word is looked up in the dictionary of sensory vocabulary stems and, if not found, is recursively reduced by one letter and looked up again. As soon as a match is found, the surrounding "leftover" letters are looked up in a German dictionary including all word forms and in several lists containing grammatical word formation components. If all letters around an identified sensory stem can be matched with existing components in compliance with the rules of German word formation, the counter of the respective sensory system for the text is augmented by 1. We implemented the algorithm within PHP and improved it by introducing a dictionary that contains all expressions already tagged as either visual, auditory, kinesthetic, olfactory or gustatory, or other. As a result, only words that are not yet included in the dictionary need to be analyzed as described above. Every time a new word is classified, the dictionary is expanded automatically. As a result, the process of scanning a text with regard to the occurrence of sensory vocabulary is much faster than if every word needs to run through the whole process of classification. The sensory profiles thus obtained for each forum post (*vprofile(p)*) are based only on the post body and do not take post headers into account. This was done because headers very often repeat the answered post's header and can therefore not be considered as original text written by the author of the post. For the same reason, we excluded empty posts from further analysis.

(3) The sensory profiles were then split into different subsets according to the needs of each hypothesis. For the subsets of hypotheses B and C, an answer tree structure was reconstructed with help of a PHP script running recursively from one post to its predecessor, following the reference number of a post. Every post was tagged as either being an initial post, a direct answer to an initial post or an answer to an answer post. In cases where referenced posts were missing (which might be the case for messages deleted by users or forum administrators), the post was handled as initial post. Several constraints were applied when creating subsets: Whenever comparing one *vprofile* (of an author or a post) to another, all profiles with fewer than 3 identified sensory words were excluded. When looking at author profiles, only authors with at least 5 non-empty posts were taken into account. An overview of the reduction steps necessary for each subset and the resulting amount of data is shown in Table 1. In view of the huge number of pairs in the complete set ("ALL") and the comparatively small percentage of pairs within the hypotheses sets, we chose to work with the same comparison set (including all pairs) for all three hypotheses instead of using a different comparison set (including all other pairs) for each hypothesis.

(4) All hypotheses were tested in two ways. First, a comparison of means was used to test the hypothesis directly. However, the result of such a comparison may yield a larger VAKOG similarity of, for example, a post and an answer to it, simply because these two documents will discuss similar content and therefore probably also use similar language. We decided to control for this by treating the full-text similarity of the two posts as a covariate. The hypotheses then are refined to "if two pairs of posts each have the same full-text similarity, the pair in which one answers the other will have higher VAKOG similarity than the unrelated pair" (analogously for the other hypotheses). Full-text similarity was operationalized as the cosine similarity between the two posts modeled as bags of words (BOW) by the WEKA[2] StringToWordVector filter, using standard parameters: word counts of a maximum of 100,000 words weighted by TF.IDF.

Since similarities were not normally distributed, we opted for a non-parametric test of these statements, choosing loglinear modeling for 3-way contingency tables. The values of the three dimensions were: (i) pairwise full-text similarity, binned into n equal-sized intervals (in our case of $n=10$: $[0, 0.1)$, $[0.1, 0.2)$, etc.), (ii) pairwise VAKOG similarity, binned into the same n intervals, and (iii) the 2 categorical values of the variable of interest (e.g. *answer-post-relation* vs. *non-answer-post-relation*).

The statistical analysis was done in Excel and with the help of an online tool for the comparison of 3-way-contingency-tables[3].

TABLE I. REDUCTION CRITERIA AND THE RESULTING AMOUNT OF DATA PER SUBSET

	Reduction criteria	Posts left for further analysis	Pairs of posts left for further analysis
ALL	• number of VAKOG terms (n_v) per post>2	97.951	4.797.150.225
Hypothesis A	• min. 2 posts with n_v>2 per author	54.182 posts of 5.879 authors	9.587.622
Hypothesis B	• min. 1 initial post and 1 corresponding answer post with VAKOG-sum per post >2	34.419	22.796
Hypothesis C	• min. 1 initial post with VAKOG-sum >2 • 1 corresponding answer post from an author with min. not-empty posts per author >4	187.273	158.885

TABLE II.   THE DISTRIBUTION OF SENSORY EXPRESSION IN THE E30 CORPUS

Posts	1,053,841
Original posts	223,973
Answer posts	829,868
Not-empty posts (neP)	646,455
Av. nr of terms per neP	40.69
Authors	30,021
Av. nr of posts per author	35.10
E30 Dictionary*	
Different terms	474,264
Different visual terms	6,798
Different auditory terms	5,047
Different kinesthetic terms	7,518
Different olfactory+ gustatory terms	1,674
Sensory terms	785,303
Visual terms	318,305
Auditory terms	248,896
Kinesthetic terms	192,566
Olfactory+gustatory terms	25,536

*Terms are handled as different as soon as they differ in one letter, including typos versions

IV.   RESULTS: THE USE OF SENSORY VOCABULARY IN FORUMS

The E30 forum corpus consists of 1,053,841 posts, written by 30,021 different authors. The corpus contains 26,305,285 terms (defined as strings of characters and/or numbers, separated by blanks and/or punctuation marks). A detailed distribution can be found in Table 2.

We categorized 21,037 different terms as sensory expression, 6,798 of them as visual, 5,047 as auditory, 7,518 as kinesthetic and 1,674 as olfactory and gustatory and stored them in a dictionary. Naturally, most of the categorized terms appeared more than once in the corpus. The corpus contains 785,303 terms classified as sensory expression, 318,305 of it visual, 248,896 auditory, 192,566 kinesthetic and 25,536 olfactory and gustatory terms. 336,055 posts (52.0% of the total amount of non-empty posts) contained at least 1 sensory term; 97,951 posts (15.2% of the total amount of non-empty posts) contained at least 3 sensory terms which was used as filter criterion when comparing single posts.

Concerning our hypotheses, we applied two methods of testing.
(1) Comparison of distributions with Mann-Whitney's U test (Fig.2): Values for each set and for both full-text and VAKOG similarity distribution were calculated. Significance testing against the null hypothesis of equal distribution was calculated separately for full-text and for VAKOG similarity and for each hypothesis, comparing the hypothesis set with the set of all post pairs. All comparisons were statistically relevant with a p-value<.0001.
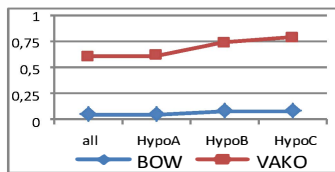

Figure 2.   Full-text and VAKOG similarity in our samples

(2) To combine the full-text and VAKOG similarities, we used loglinear modeling for 3-way contingency tables (Fig.3). Results were calculated comparing all 3 tables at once and comparing pairs of 2 tables collapsed across the levels of the third. All comparisons were statistically significant with a p-value <.0001.
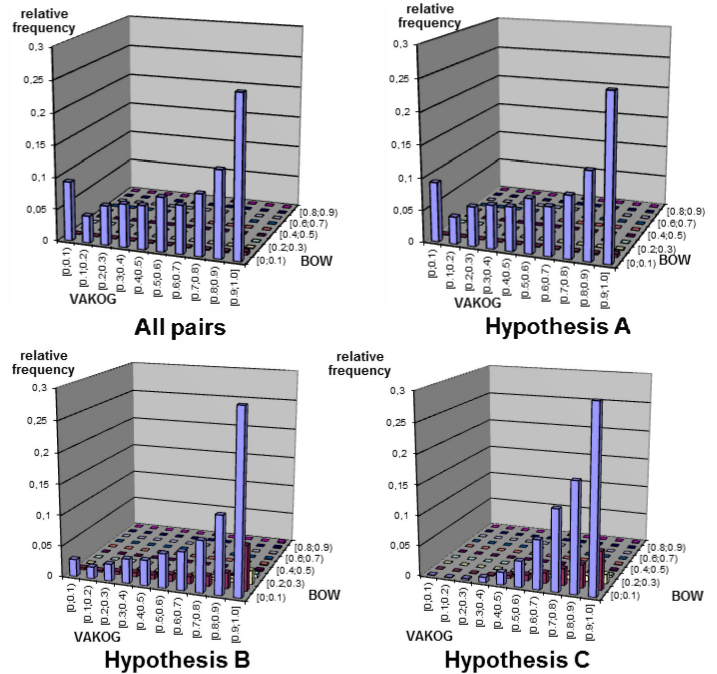

Figure 2.   Distribution of full-text and VAKOG similarity within the samples

V.   DISCUSSION

According to the findings of this study, the questions we raised in the introduction are answered as follows: (a) Sensory vocabulary is used sufficiently often to be considered as an indicator for individual preference. (b) Our results show that there are individual tendencies concerning the use of sensory expression. Text can be used as a source of implicit information about its author through word by word analysis and identification of all sensory expressions. (c) The result of our analysis of word use is a vector with four dimensions for each user. The vector describes how much a user expresses his or her thoughts in visual, auditory, kinesthetic and olfactory and gustatory terms.

Looking at the distribution of sensory expression within the E30 forum corpus, around one out of 30 terms could be classified as sensory expression. (By manual analysis we found that there is still a large number of non-systematic "dummy terms" within the corpus replacing original links, images, references to other postings etc. The occurrence of sensory expression might thus be higher in text corpora containing only "real" words). Even though the used number of sensory terms might not seem very high, it is sufficient to obtain relevant information on the average user (who publishes 35 posts).

The hypotheses testing on the E30 forum corpus shows a significantly higher VAKOG similarity within the hypotheses subsets than within the set of all post pairs. These results were consistent both for the comparison of means and the combined full-text and VAKOG similarities. This confirms all three hypotheses and leads to the following conclusions: Authors of forum posts have a tendency to use sensory expressions in similar distributions over time. Hence, that distribution can be considered as an interesting extension to user descriptions for user modeling. The corpus analysis showed further a more similar distribution of sensory vocabulary within answer threads than in randomly chosen pairs of posts. This might be explained either by a tendency of users to adapt their use of expression to the dialog partner or as an effect of topic relatedness. In any case, there is a tendency for authors to rather answer posts that correspond to their personal sensory preference.

Limitations: Our corpus of sensory expression is still far from covering all expressions with a direct link to a sensory system that exist in German. Therefore, fewer words are classified as sensory expression than are effectively within the analyzed texts. The method of matching letters around an identified sensory stem with components used in German word formation helps to classify also words that are not spelled correctly and ad hoc word creations. On the other hand, it might sometimes allow compositions that should not be classified as sensory expression. In every such case we found, we implemented exception handling. Nevertheless it is possible that due to this, some more words are classified as sensory expression than are effectively within the analyzed texts. However, all these misclassifications are constant during the whole corpus analysis and should therefore not affect any of the presented results. We are working on expanding our corpus. In this paper, we only examined the use of sensory vocabulary within a narrow discussion topic. Further research needs to take into account also different topics, in order to investigate the influence of topic on the use of sensory expression.

VI. CONCLUSIONS AND OUTLOOK

In this paper, we have proposed a new dimension for user modeling based on the use of sensory expressions. Based on findings from cognitive information processing and learning styles, we investigated the potential of the idea to analyze the use of sensory expression as an individual preference that might indicate sensory preference.

We opted for an implicit approach to data acquisition concerning the use of sensory vocabulary by means of analyzing forum text. Starting from a list of 356 stems of German sensory expressions, we implemented a testing pipeline in order to classify more than 1,000,000 forum posts with regard to the use of sensory vocabulary. We identified more than 20,000 different sensory terms (by counting difference in at least one letter as distinction criterion). We found that authors tend to use sensory expressions in similar distributions when writing new posts. Furthermore, similarity concerning the use of sensory vocabulary is higher within answer threads. Both of those results support our idea of treating the use of sensory vocabulary as an interesting new dimension in user modeling.

The obtained results are quite encouraging: Our next steps are to enlarge the corpus of sensory expression, investigate the relation between preferred sensory system(s) and the use of sensory expression by combining forum text analysis with user tests on sensory preference, and examine the influence of topic on the use of sensory expression.

REFERENCES

[1] Andreevskaia, A., Bergler, S., Urseanu, M.: All Blogs Are Not Made Equal: Exploring Genre Differences in Sentiment Tagging of Blogs. In: Proceedings of the Int. Conf. on Weblogs and Social Media (ICWSM07), Boulder, Colorado, pp. 195–198, 2007.

[2] Atman, N et al: Learning Styles Diagnosis Based on Learner Behaviors in Web Based Learning. In: Gervasi, O. et al. (Eds.): ICCSA 2009, Part II, pp. 900–909, 2009.

[3] Baccianella, S., Esuli, A., Sebastiani, F.: SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. Proceedings of the 7th conference on International Language Resources and Evaluation LREC10, pp. 2200–2204, 2008.

[4] Balog, K. de Rijke, M.: How to Overcome Tiredness: Estimating Topic-Mood Associations. In: Proceedings of the Int. Conf. on Weblogs and Social Media (ICWSM07), pp. 199–202, 2007.

[5] Bautin, M., Vijayarenu, L., Skiena, S.: International Sentiment Analysis for News and Blogs. In: in Proceedings of the International Conference on Weblogs and Social Media (ICWSM), pp. 19–26, 2008.

[6] Chang, Y., Kao, W., Chu, C., Chiu, C.: A learning style classification mechanism for e-learning. In: Computers & Education 53, pp. 273–285, 2009.

[7] Chen, S.Y., Magoulas, G.D., Dimakopoulos, D.: A Flexible Interface Design for Web Directories to Accommodate Different Cognitive Styles. In: Journal of the American Society for information science and Technology 56/1, pp. 70–83, 2005.

[8] Chi, M., VanLehn, K., Litman, D., Jordan, P.: Inducing Effective Pedagogical Strategies Using Learning Context Features. In: De Bra, P., Kobsa, A., Chin, D. (Eds.): UMAP 2010. Springer, Berlin; pp. 147–158, 2010.

[9] Coffield, F., Moseley, D., Hall, E., Ecclestone, K.: Learning styles and pedagogy in post-16 learning. A systematic and critical review. LSRC, London , 2004.

[10] Dunn, R.: Commentary: Teaching Students Through Their Perceptual Strengths or Preferences. In: Journal of Reading 31/4, pp. 304–309, 1988.

[11] Felder, R., Silverman, L: Learning and Teaching Styles In Engineering Education. In: Engr. Education, 78/7, pp. 674–68, 1988.

[12] Godbole, N., Srinivasaiah, M., Skiena, S.: Large-Scale Sentiment Analysis for News and Blogs. In: Proceedings of the ICWSM07, Boulder, Colorado, pp. 219–222, 2007.

[13] Goldstein, E.B.: Sensation & Perception. Thomson Wadsworth, Belmont, 2007.

[14] Gregory, R.L., Colman, A.M.: Sensation and Perception. Longman Essential Psychology, London, 1995.

[15] Heinath, M., Dzaack, J., Wiesner, A., Urbas, L.: Applications for Cognitive User Modeling. In: Conati, C., McCoy, K., Paliouras, G.

(Eds.): UM 2007. LNCS, vol. 4511, pp. 127–136. Springer, Heidelberg, 2007.

[16] Kellner, G.: Wege der Kommunikationsoptimierung. Anwendung von NLP im Bereich der Künstlichen Intelligenz. VDM, Saarbrücken, 2010.

[17] Kellner, G., Berendt, B.: Towards a new dimension for user modeling: The use of sensory vocabulary. In: Advances in User Modeling: Selected papers from the UMAP 2011 workshops. Springer, Berlin, in print.

[18] Keshtkar, F., Inkpen, D.: Using Sentiment Orientation Features for Mood Classification in Blogs. Proceedings of the IEEE NLP-KE 2009, 2009.

[19] Laming, D.: The Measurement of Sensation. Oxford University Press, New York, 1997.

[20] Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.X.: Topic sentiment mixture: modeling facets and opinions in weblogs. In Proceedings of the international conference on World Wide Web, pp. 171–180, 2007.

[21] Nowson, S.: The Language of Weblogs: A study of genre and individual differences. Doctoral Thesis. University of Edinburgh, 2006.

[22] Ramirez-Esparza, N., Chung, C.K., Kacewicz, E., Pennebaker, J.W.: The psychology of word use in depression forums in English and in Spanish: Testing two text analytic approaches. In: Proceedings of the Int. Conference on Weblogs and Social Media (ICWSM08), Seattle, Washington, pp. 102–108, 2008.

[23] Reid, J.M.: The Learning Style Preferences of ESL Students. In: TESOL Quarterly, 21/1, pp. 87–111, 1987.

[24] Richling, J.: Die Sprache in Foren und Newsgroups. VDM, Saarbrücken, 2008.

[25] Sadler-Smith, E., Riding, R.: Cognitive style and instructional preferences. In: Instructional Science 27, pp. 355–371, 1999.

[26] Shute, V., Towle, B.: Adaptive E-Learning. In: Educational Psychologist 38/2, pp. 105–114, 2003.

[27] Strappavara, C., Valitutti, A., Stock, O.: The AffectiveWeight of Lexicon. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC), Genoa, pp. 423–426, 2006.

[28] Valitutti, A., Strappavara, C., Stock, O.: Developing Affective Lexical Resources. In: PsychNology 2/1, pp. 61–83, 2004.