Gaze-tracked Crowdsourcing

Jakub Simko
Institute of Informatics and Software Engineering
Slovak University of Technology
Bratislava, Slovakia
Email: jakub.simko@stuba.sk

Maria Bielikova
Institute of Informatics and Software Engineering
Slovak University of Technology
Bratislava, Slovakia
Email: maria.bielikova@stuba.sk

***Abstract*—When creating intelligent systems, we often need proper knowledge bases and resources annotated with metadata. Sometimes, we have no other option, than to utilize crowdsourcing, to acquire the data in necessary quantity. Crowdsourcing is a costly endeavor, always with space for improvements in task solving quantity and quality. Studies show that consideration of implicit feedback (behavior of workers during task solving) helps to improve the overall crowd output. Gaze-tracking is a powerful source of implicit feedback as it records user's activity outside typical feedback channels (e.g. clicking, scrolling, typing) and reveals a great deal of person's cognitive processes. This paper argues that gaze-tracking represents a potent feedback source even for crowdsourcing, as gaze-tracking technology becomes available for wider worker pools. The paper also presents an example case study demonstrating the use of the gaze-tracking during a typical crowdsourcing task – acquisition of training dataset for automated word sense disambiguation. Normally in such task, the worker explicitly selects a corresponding sense for a given word located in a text snippet and thus contributes to the dataset. With gaze-tracking involved, worker also shares us other information useful for dataset enrichment: worker's reading pattern (which may indicate confidence) and important sense-distinguishing words (e.g. contextual words that trigger worker's decisions).**

I. MOTIVATION

Nowadays, a plethora of methods and technologies are focused on automated processing of information spaces, especially Web. Although their ultimate goal are fully automated processes, in the meantime, they need a plenty of human-generated input: creation of training and experimental datasets, preparation of domain models, ad-hoc human interventions during the information processing or post-hoc validations of outputs of automated processes.

Therefore, much research is devoted on how to *effectively* employ humans into creation of information and knowledge needed for automated processing of Web resources. This type of human work is largely covered by the research field of *human computation*, which promotes the use of human mind labor for solving computational tasks that are hard or impossible to be solved by machines. Depending on a "type of the job", human computation may have either character of (1) *expert work* (when the knowledge is created by small number of highly capable domain experts) or (2) *crowdsourcing* (when a broad group of workers of unknown qualification is organized by a software platform to acquire knowledge) [1]. For certain human computation tasks, the use of experts is irreplaceable (e.g. creation of an ontology core). However, their cost is too high and availability too low for creation of extensive datasets [2]. Here, the crowdsourcing is a more typical option, because crowd can provide the necessary quantity, while maintaining the sufficient quality of provided information.

Despite scalability and result quality, crowdsourcing also has its effectiveness issues. Part of the crowd's problem solving capacity must always be spent on quality assurance. Not all task solutions provided by the workers are right, either due to mistakes or deliberately harmful behavior (e.g. spam). One of the reasons, why the solution quality assessment is hard in crowdsourcing, is the lack of transparency in the task solving process. Crowdsourcing platforms usually treat their workers as black-boxes: they put in tasks and expect valid solutions, yet they have no means (in general) of automated solution validation. Neither they are aware of the process of task solving itself. Yet, the worker behavior, manifesting as a stream of user actions (implicit feedback), may contain useful traits indicating the quality of the resulting solution. For example, more skilled workers [3] or spammers [4] can be detected. Moreover, the gain from an already correct answer can be improved. For example if we ask the worker to categorize articles and request her to read them sentence by sentence, we not only receive article categories, but implicitly also particular sentences, which convinced her to make the decision.

Gaze-tracking is an unique source of implicit user feedback in human-computer interaction. Abstracting from its technological realization, it records sequences of user's gaze fixations during a computer session. The gaze-tracked sequence implicitly carries much information about attention, interest and even mental state of the user. These are, at the same time, hard to obtain from "traditional" implicit feedback sources (such as cursor movement, clicking, scrolling or server logs). The steadily decreasing costs of gaze-tracking technologies make them available as a potent source of implicit feedback even for crowdsourcing. *It is therefore viable to prepare for this trend in advance and research opportunities of harnessing gaze tracking for crowdsourcing*.

The goal of this paper is to advocate the future fusion of crowdsourcing and gaze-tracking. First, we analyze fields of crowdsourcing and gaze tracking and show their complementary properties. Then, we discuss the fusion of these fields, define open challenges and lay out possible research directions. Lastly, we present an example case study in the domain of training dataset preparation for term meaning disambiguation algorithms. During the study, we tracked workers' gaze and identified possibilities to enrich task results using gaze data.

II. Towards synergy between human computation and gaze-tracking

A. Crowdsourcing

The crowdsourcing is useful for a wide spectrum of tasks, in which human computation is needed in large scale. These include creation of knowledge bases, preparation of training datasets (for automated text analysis, computer vision, etc.), mapping of geographical areas or civic activism [5]. At the same time, crowdsourcing comes in a variety of forms and can be classified from multiple perspectives [6]. These include composition of the crowd, motivation and recruitment schemes, task decomposition and distribution, quality assurance or abuse prevention. Typically, generic crowdsourcing platforms (such as Amazon Mechanical Turk or CrowdFlower) are used to convey tasks to workers, who are recruited from general Web users and motivated to participate by micropayments for each solved task.

One of the drawbacks of generic crowdsourcing platforms is the lack of implicit feedback acquisition from the workers. Although workers deliver their task solutions (correctness of which can be evaluated by other workers), the process of task solving itself remains obscured. It is possible to record user actions such as clicks, scrolling or keypressing in the work environment, however, this is mostly not enough to uncover cognitive processes of the worker. For example, using "traditional" feedback actions, we can track, when the worker was stuck on a particular question in a questionnaire (indicated by a long time that passed since previous question was answered). But without additional information, we cannot say whether it was due to difficulty of the question, fatigue of the worker or due to the fact that the worker took a break or was disturbed.

Several studies show us, how more implicit feedback from users can help to improve the crowdsourcing's effectiveness. In their study [4], Chen et al. detected spammers in community question answering systems. A fraud type, common for these systems, is when spammers try to place recommendations on commercial products into answers against interests of the community. Chen et al. detect these users by analyzing their behavior: repetition of similar posts, non-authentic voting for their own answers or use of same sentence constructions.

In another study [3], Rzeszotarski and Kittur introduced a general method for tracking of the worker behaviour through low level user actions (clicks, cursor movements, scrolling, keypressing). Authors took streams of these actions, mapped them to quality metrics (e.g. accuracy, understandability) of respective task solutions and trained models for predicting these metrics. They evaluated this method on content classification, generation and comprehension tasks.

Analyzing the current state-of-the-art in crowdsourcing quality management, we conclude that *analysis of implicit worker feedback, harvested during crowd-task solving, possitively contributes to the effectiveness of the crowdsourcing processes*.

B. Gaze-tracking

The gaze-tracking is not new – first automated approaches started to appear almost seventy decades ago and were implemented by various technological means (including special "pointy" contact lenses or measurement of electrical potential of eye movements). Today, the most widespread technique of eye-tracking (on which the gaze-tracking builds) is based on non-invasive analysis of reflection of infrared rays, projected on the eyes [7].

Despite not being new, the use of gaze-tracking has long been limited to limited-scale lab work. This had several reasons: technology costs (not only of tracking hardware, but also data analysis software), weak portability, specific environment requirements and overall difficulty of using of the technology. Yet, the today's technological advances gradually improve all of these aspects. The price of the technology is dropping[1], devices become portable and robust and their driving software easier to use. Vendors are competing and some of them already declared the will to deliver gaze-tracking as built-in a technology into computing devices. We can therefore expect a wider penetration of this technology and its breakout outside the labs, into business and retail.

The gaze-tracking is used in many research and practical areas, in which implicit feedback from people is of interest. In marketing, gaze-tracking is used to measure interest of a viewer in specific regions of advertisement graphics. In human-computer interface design, a detection of confusing interface sections and scenarios, is supported by "gaze plots" [8]. In such use cases, qualitative, small-scale studies are typical, with gaze-tracking as one of many sources of feedback from experiment participants (others being, for example, loud thinking or decision analysis). An essential role in qualitative studies is played by human analysts, who must be exprienced-enough to interpret the individual feedback action streams.

Even the "manual" (qualitative) analysis of gaze data must be supported by automated preprocessing – for aggregation and visualization. Mostly used visualizations include heat maps, which show the aggregated interest of participants in certain object in the interface, and gaze plots, which display the sequences of participant's gaze fixations and saccades (movements between fixations). These visualizations are widely used, but sometimes, more specific ones are needed. One such was created by Rakoczi and Pohl, who analyzed pathological behaviour of students in online learning system [9].

With wider penetration of gaze-tracking technologies in the future, the qualitative approach, where each user-session is manually examined by a human examinator, would not scale. Some research was already undertook to explore the quantitative approaches. For example Pan et al. studied behaviour of people on web pages using only quantitative aggregations (study with 30 participants) [10].

Much research attention in automated (quantitative) processing of gaze data, is devoted towards *text reading* (i.e. what exactly and when the user reads). Primary issue that is being solved, is the actual transformation of the raw gaze data sequence (usually a sequence of timespans coupled with screen coordinates) into a sequence of fixations on text objects (letters, words, sentences). This non-trivial task suffers especially from insufficient accuracy of gaze tracking hardware or its imperfect calibration. In his gaze-based text-reading, quantitative analysis framework, Martinez-Gomez [11] used the text itself for

[1] http://www.economist.com/news/technology-quarterly/21567195-computer-interfaces-ability-determine-location-persons-gaze

help: using information about center positions of the words in text (on which the fixations are usually focused) and lengths of words (which correlate with lengths of fixations), he was able to confront expected sequences with measured ones and thus overcome the systematic measurement bias.

Once the reading sequence is known on the lexical level, it can be utilized further. Martinez-Gomez et al [12] used their framework for identification of those sequences of texts, which are comparatively harder to read (regardless of the reason), by identifying snippets which were read repeatedly, or on which the readers got regularly "stuck". Other study [13], of Xu et al., focused on text summarization. Based on the reading sequences (which apart from the actual order of words read contained also the lengths of fixations on the words), the authors computed an aggregated gaze "intensity" for each word of the document and propagated this intensity to other words according to their semantic relatedness. Next, they aggregated the intensities per sequence, sorted them and created the summarization from the top-ones.

Another domain for utilizing of the automated gaze-tracking analysis is personalized Web. Buscher et al. [14] used the gaze data and text analysis for acquisition of information on short-term context (interests) of the web users and used it for expansion of their search queries. Their method extracts keywords from the web search result snippets to build short-term user model, but only from the snippets, on which the user was "gazing" the most (taking the fixation time and repetition into account). Gaze-tracking can also help in assessing long-term Web user interests, which are typically used for recommendation. In another study, Xu et al. [15] gaze-tracked the users for some time and estimated their interests (used later for content recommendation) according to reading of texts, video watching and interest in advertisement objects.

Last but not least, the gaze-tracking (eye-tracking in general) can help to determine the mental state of the observee (as was shown by McDonald, the cognitive state of the human is reflected in eye's gaze and movements [16]). Practical implications of this can be found, for example in technology enhanced learning, where empathetic software agents intervene upon detection of problematic states of students [17].

None of the examples of quantitative analysis of gaze data we listed, was designed as a crowdsourcing task, yet many of the approaches are naturally close to it (in various aspects).

C. Fusion into a new field

The number of gaze-tracking applications for gathering of implicit feedback encourages its use also in crowdsourcing scenarios. Many existing gaze-tracking applications already resemble tasks typical for mass human computation. For example, the mentioned text summarization typically needs to be done manually during creation of training samples for machine learning algorithms. Such manual task is tedious, but using gaze-tracking in a scenario similar to work of Xu et al. [13], the work could be done faster and with more comfort for the worker. Similar opportunities can be found in other typical crowd tasks, such as multimedia resource tagging. For example, in the image tagging scenario, where worker's task is to identify objects depicted in the images using tags, we can utilize the gaze data to guess additional information, e.g. where exactly in the images the objects are depicted.

There are also other tasks, in which gaze data are analysed mostly manually, for example user experience studies or evaluation of marketing resources. These too can potentially be redesigned towards more automation and quantitative evaluation, if enough users can be gaze-tracked.

From an optimistic viewpoint it seems that only obstacle left is the still small penetration of the eye-tracking hardware, which will, however, gradually become less of a problem.

Yet, as a completely new discipline is opening, new challenges lie ahead and we should be preparing for them already. We should create new methods that utilize gaze-tracking in crowdsourcing scenarios and evaluate them, at least in laboratory conditions (one example study is presented in this paper). The new field will gradually develop its own fundamentals, challenges and questions, as more and more approaches will be created. We expect some of the open questions to be of the following:

Method of task design. How will a typical approach of designing the gaze-tracked crowd task look like? Although we cannot rule out genuinely new task designs (which may be brought in by new technological options), we expect that currently, the most natural way of designing gaze-tracked tasks leads through modifications of existing regular crowdsourcing tasks. *We should systematically re-think our crowd task designs and identify those, where additional gaze information can help in improving quality of the task solutions.*

Task types. What types of tasks can benefit most from gaze tracking? We see no prior limitations except technical ones: the worker must have the means for properly setting up the eye-tracking environment (which, for example, greatly limits tasks designed for mobile platforms). However, we see application of gaze-tracking for certain task types as more straightforward than other. For example: image categorization, text reading or resource annotation. These are tasks, where workers spent a great portion of their time by studying visual stimuli, during which we don't have any means of tracking his behavior – except the gaze-tracking. On the other hand, we see less oportunities in tasks, that are not so dependent on visual stimuli, for example survey question answering or creative tasks (e.g. image drawing contests).

Specificity of the approach. Are there any universally applicable task design schemes we can use? Each task design will probably have to be unique for the particular problem it tries to solve (just as in any other field). We also expect the task designs be narrow, because semantic interpretation of gaze data strongly depends on the case. Often, gaze data are hard to be interpreted even by humans (not mentioning algorithms), even if the task (e.g. interface the user is working with) is even moderately complex.

On the other hand, we also see some principles, which may potentially be more broadly applicable. For example, it is useful to know, in what general mood is the worker. Is he nervous, stressed, discomforted, focused or relaxed? Knowledge about this may serve as useful predictor in task-result quality assessment. And, eye-tracking technology can greatly help with measurement of these factors (for example, pupila dilatation, which is an eye metric commonly available on eye trackers, is directly connected to person's excitement).

Diagnostic vs. interactive system. According to Duchowski [18], systems utilizing gaze-tracking can be split into two categories: (1) diagnostic (in which gaze tracking does not directly influence the user's activities, is only measured and the data are analyzed post-hoc) and (2) interactive (in which user directly controls the application through his gaze). The gaze-tracked crowdsourcing systems inherit this dichotomy. The diagnostic paradigm is more straightforward (consider all the previous examples and the modificatory task design philosophy we advocate). On the other hand, interactive approach offers a great potential in improving ergonomy of tedious crowdsourcing tasks. For example, in an user interface of a classification task, selection of a particular category by gaze may be faster than selection using pointing device.

III. CASE STUDY: WORD SENSE DISAMBIGUATION

To demonstrate the potential of gaze-tracking support for crowdsourcing, we conducted a qualitative study in a typical crowdsourcing domain: a dataset preparation. In automated text processing, the *word sense disambiguation* is an essential step. It often involves machine learning, trained by extensive datasets which need to be created by humans in crowdsourcing scenarios [19]. A training set is typically a set of triples $(w_t, \bar{c}, s)$ where w_t denotes the task word, sense of which should be determined, $\bar{c} = (w_1, w_2, ..., w_n)$ denotes a vector of words that represent the snippet in which w_t is placed and $s \in S_t$ denotes the actual sense that should be predicted for given w_t and $\bar{c}$. The worker's task is defined simply: read (skim) through a given snippet of text ($\bar{c}$) and for the given w_t select the corresponding s from S_t (which denotes possible senses for the word w_t).

Our research question was to explore *whether the gaze-tracking of workers during word disambiguation task solving can disclose useful information that can improve the task output.* More specifically, we postulated that for each example, a subvector $\bar{c}_d \subset \bar{c}$ of the context vector exists, in which all words have relatively high sense distinguishing value when related to the examined w_t (e.g. they are semantically related to the sense s). We hypothesized that the words from $\bar{c}_d$ will convince the workers to stop reading the snippets and make their decisions immediately. Such behavior can be gaze-tracked and $\bar{c}_d$ words disclosed. Consequently, these "important" words would receive more weight in the resulting training set.

Five participants in our study were assigned with the same task: to identify the sense of a single word ("president") in 10 different contexts (which were snippets few sentences long). There were 3 possible senses available (head of a company, head of state and U.S. president). We used random examples from SemEval2 dataset, used specifically for sense disambiguation experiments [19]. All participants consecutively went through all 10 tasks. Tasks were presented on slides and participants answered them verbally. We have also encouraged the participants to think aloud and share their thoughts about the solving process. Throughout the experiment, the participants were gaze-tracked using Tobii X2-30 device[2].

[2]http://www.tobii.com/en/eye-tracking-research/global/

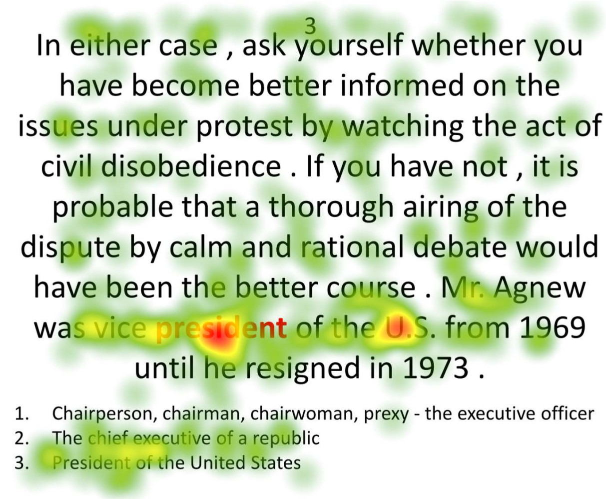

Fig. 1. Example of gaze-tracked disambiguation task for word "president". The sense distinguishing word "U.S." received much worker gaze attention.

Together, the participants commenced 50 sense assignments, making 5 individual errors. By applying a majority vote, 9 of 10 examples were assigned correctly (which was in line with experiments of Snow et al [19], who previously tested the crowd performance on the same task and data).

We manually evaluated the gaze data (using video-replay gaze-plot and heatmap visualizations). First, we sought for words the participants looked lastly prior to making their (verbal) decision about a particular snippet. In 54% of cases, the participants made their last gaze fixations on one of the "sense distinguishing words" found in $\bar{c}_d$ (which we manually assessed prior to the experiment). In rest of the cases, the participants either read the snippet all to the end (27%) or stopped on words not from $\bar{c}_d$ (19%). From this, we conclude, that apart from gaining the primary result of the crowd-assignment (i.e. correct picking of s), we can, to some extent, identify sense distinguishing words (provided we have an algorithmic tool for identifying the "lastly read words") and use this knowledge to further enrich the dataset.

The a posterior analysis of gaze data has furthermore disclosed other behavioral patterns. The participants used two strategies of reading of the snippets. 3 of them have properly read from the beginning, either to the end or to the moment they found a strong enough "sense distinguishing word". Rest of the participants preferred fast text skimming (starting not necessarily from the beginning of the snippet, but rather in close proximity to the task word w_t). They only resorted to "proper reading" when they were unsure about their answer. Repeated reading also occurred with "proper readers", especially with examples that were hard to solve (as reported by participants). This was most obvious in the only wrongly evaluated case (where 3 participants chose incorrect s). This finding is also useful for gaze-tracked crowdsourcing: by detecting repeated snippet reading, we may track the confidence of workers about their answers.

IV. LOOKING AHEAD

We believe that pricing trends in gaze-tracking technologies will soon open up a joint research area of gaze-tracking in crowdsourcing. We have argued, why gaze-tracking complements the crowdsourcing. We presented one example of how the intersection of these fields may look like: a training dataset preparation task (for term sense disambiguation), where gaze data can disclose important contextual words.

We believe, that similar tasks may be straightforwardly drawn up (e.g. in an image categorization task, distinct image regions could be identified). Some examples of human-computation (outside crowdsourcing) tasks already exist (e.g. for text summarization [13] or hard-to-read text detection [12]).

We expect that various research questions will be raised on how to design the gaze-tracked crowdsourcing tasks effectively. Many of the future methods will probably be based on existing crowdsourcing scenarios. In fact, as researchers, we should systematically revise existing crowdsourcing tasks and analyze possibilities of enhancing them with gaze-tracking data. Many of them will be problem-specific, but we can also arguably expect more general principles to emerge (e.g. approaches for worker confidence detection). The most promising application domains will be those, where workers visually process some content (e.g. image processing, text processing). Effort will also be needed for automating the existing (currently mostly qualitative) approaches of gaze data analysis.

Acknowledgments

This work was partially supported by the Scientific Grant Agency of the Slovak Republic, grant No. VG 1/0646/15 and was created with the support of the Ministry of Education, Science, Research and Sport of the Slovak Republic within the Research and Development Operational Programme for the project ITMS 26240220084, co-funded by the European Regional Development Fund.

References

[1] J. Simko and M. Bielikova, *Semantic Acquisition Games: Harnessing Manpower for Creating Semantics*. Springer Publishing Company, Incorporated, 2014.

[2] M. Sabou, K. Bontcheva, and A. Scharl, "Crowdsourcing research opportunities: Lessons from natural language processing," in *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, ser. i-KNOW '12. New York, NY, USA: ACM, 2012, pp. 17:1–17:8. [Online]. Available: http://doi.acm.org/10.1145/2362456.2362479

[3] J. M. Rzeszotarski and A. Kittur, "Instrumenting the crowd: Using implicit behavioral measures to predict task performance," in *Proc. of the 24th Annual ACM Symp. on User Interface Software and Technology*, ser. UIST '11. New York, USA: ACM, 2011, pp. 13–22. [Online]. Available: http://doi.acm.org/10.1145/2047196.2047199

[4] C. Chen, K. Wu, V. Srinivasan, and K. B. R, "The best answers? think twice: Online detection of commercial campaigns in the cqa forums," *CoRR*, vol. abs/1208.1448, 2012. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr1208.html#abs-1208-1448

[5] C. A. Le Dantec, M. Asad, A. Misra, and K. E. Watkins, "Planning with crowdsourced data: Rhetoric and representation in transportation planning," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '15. New York, NY, USA: ACM, 2015, pp. 1717–1727. [Online]. Available: http://doi.acm.org/10.1145/2675133.2675212

[6] A. J. Quinn and B. B. Bederson, "Human computation: A survey and taxonomy of a growing field," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 1403–1412. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979148

[7] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2009.30

[8] N. H. M. Zain, F. H. A. Razak, A. Jaafar, and M. F. Zulkipli, "Eye tracking in educational games environment: Evaluating user interface design through eye tracking patterns," in *Proceedings of the Second International Conference on Visual Informatics: Sustaining Research and Innovations - Volume Part II*, ser. IVIC'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 64–73. [Online]. Available: http://dl.acm.org/citation.cfm?id=2074670.2074678

[9] G. Rakoczi and M. Pohl, "Visualisation and analysis of multiuser gaze data: Eye tracking usability studies in the special context of e-learning," in *Proceedings of the 2012 IEEE 12th International Conference on Advanced Learning Technologies*, ser. ICALT '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 738–739. [Online]. Available: http://dx.doi.org/10.1109/ICALT.2012.15

[10] B. Pan, H. A. Hembrooke, G. K. Gay, L. A. Granka, M. K. Feusner, and J. K. Newman, "The determinants of web page viewing behavior: An eye-tracking study," in *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, ser. ETRA '04. New York, NY, USA: ACM, 2004, pp. 147–154. [Online]. Available: http://doi.acm.org/10.1145/968363.968391

[11] P. Martinez-Gomez, "Quantitative analysis and inference on gaze data using natural language processing techniques," in *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, ser. IUI '12. New York, NY, USA: ACM, 2012, pp. 389–392. [Online]. Available: http://doi.acm.org/10.1145/2166966.2167055

[12] P. Martínez-Gómez, T. Hara, C. Chen, K. Tomita, Y. Kano, and A. Aizawa, "Synthesizing image representations of linguistic and topological features for predicting areas of attention," in *Proc. of the 12th Pacific Rim Int. C. on Trends in Artificial Intelligence*, ser. PRICAI'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 312–323. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-32695-0_29

[13] S. Xu, H. Jiang, and F. C. Lau, "User-oriented document summarization through vision-based eye-tracking," in *Proc. of the 14th Int. C. on Intelligent User Interfaces*, ser. IUI '09. New York, NY, USA: ACM, 2009, pp. 7–16. [Online]. Available: http://doi.acm.org/10.1145/1502650.1502656

[14] G. Buscher, A. Dengel, and L. van Elst, "Query expansion using gaze-based feedback on the subdocument level," in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '08. New York, NY, USA: ACM, 2008, pp. 387–394. [Online]. Available: http://doi.acm.org/10.1145/1390334.1390401

[15] S. Xu, H. Jiang, and F. C. Lau, "Personalized online document, image and video recommendation via commodity eye-tracking," in *Proceedings of the 2008 ACM Conference on Recommender Systems*, ser. RecSys '08. New York, NY, USA: ACM, 2008, pp. 83–90. [Online]. Available: http://doi.acm.org/10.1145/1454008.1454023

[16] S. McDonald and R. Shillcock, "Eye movements reveal the on-line computation of lexical probabilities during reading," *Psychological Science*, vol. 14, no. 6, pp. 648–652, 2003, cited By 82. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0346991530&partnerID=40&md5=0e86ff5198e854cdc6bd5049009787db

[17] H. Wang, M. Chignell, and M. Ishizuka, "Empathic tutoring software agents using real-time eye tracking," in *Proceedings of the 2006 Symposium on Eye Tracking Research &Amp; Applications*, ser. ETRA '06. New York, NY, USA: ACM, 2006, pp. 73–78. [Online]. Available: http://doi.acm.org/10.1145/1117309.1117346

[18] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003.

[19] R. Snow, B. O'Connor, D. Jurafsky, and A. Y. Ng, "Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks," in *Proc. of the C. on Empirical Methods in NLP*, ser. EMNLP '08. Stroudsburg, PA, USA: ACL, 2008, pp. 254–263. [Online]. Available: http://dl.acm.org/citation.cfm?id=1613715.1613751