NOTICE: This is the author's version of a work accepted for publication by Springer. Changes resulting from the publishing process, including editing, corrections, structural formatting and other quality control mechanisms, may not be reflected in this document. Changes may have been made to this work since it was submitted for publication in Conference: Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML 2014), 2014. The final publication is available at Springer.

Deriving the Employee-perceived Application Quality in Enterprise IT Infrastructures using Information from Ticketing Systems

Susanna Schwarzmann, Thomas Zinner, and Matthias Hirth

University of Würzburg, Institute of Computer Science, Würzburg, Germany

The need for a less complex maintenance of applications and the IT infrastructure for huge enterprises lead to the centralization of applications and services within data centers. Employees at sites and branches are connect to data centers via the Internet using a thin-client architectures resulting in additional failure sources beside the end devices, namely the transport network and hardware components in the data centers. To provide a good application quality to the employers using a multitude of different applications and access networks has thus become a complex task [3].

In order to evaluate the quality of an application, subjective metrics like Quality of Experience (QoE) [1] are often used. Ongoing research in the field of QoE typically tries to understand the impact of technical systems on the subjective perception of specific applications. Main influence factors are deduced and appropriate models allowing an estimation of the QoE for varying parameters like bandwidth, packet loss, or jitter are developed. The QoE for applications like web browsing, video streaming, VoIP, and office products are well understood. This, however, does not hold for enterprise applications like resource planning and management or data warehouse applications, which are not covered by current research. Time-consuming user surveys in the employers working environment highly affect the day-to-day business and thus are not practicable. Nevertheless, a profound knowledge of the application quality and availability is required to enable good conditions of work and a high working efficiency. For that, enterprises may rely on support systems like a hotline or a ticketing system. Particular the latter is a huge database collecting complaints and problems of the users over a long period of time and thus are an interesting starting point to identify performance problems.

Using this data source, we propose an approach to automatically identify tickets indicating problematic applications and reflecting the user experience. To this end, our approach first groups similar tickets and afterwards tags the resulting groups with adequate keywords. For the grouping process, we rely on the information from the free-text fields of the tickets, which include a summary and a detailed description of the reported issue, and calculate the lexicographical distance between the tickets using the Jaccard index [2]. The keywords for the groups are based on word frequencies within the groups. The tagged groups can finally be evaluated further to identify issues in the IT system.

We evaluate the accuracy of our approach using 12,000 tickets accumulated in June 2013 at the ticketing system of a company. These tickets were manually

categorized in tickets covering application performance issues (303 performance tickets), respectively tickets addressing other issues and serve as gold standard data for the categorization results. The performance of the approach is measured by two different metrics, (1) the overall share of correctly classified tickets and (2) the share of performance relevant tickets in groups tagged as performance relevant. A parameter study was conducted to investigate the impact of different Jaccard similarity thresholds on the misclassification rate (Type I and Type II errors). Based on the specific threshold, the ratio of correctly classified performance tickets varied between 44 % and 57 %, whereas, the number of false classified non-performance tickets was between 55 and 325. Hence, a higher hit ratio also results in more manual overhead for checking the tickets within the performance ticket groups and removing the wrongly classified tickets.

Even though not all performance tickets can be detected by our algorithm, the number of correctly classified tickets is sufficient to draw conclusions about temporal performance problems. To this end, we compared the number of actual performance tickets per day with the correctly classified number of performance tickets per day identified using our approach. Both time series show similar trends, although the identified number of perfomance tickets per day is always lower than the actual number. Nevertheless, daily trends are preserved and presented approach detected at minimum 20% of the daily performance tickets. Consequently, the algorithm cannot be used if the exact number of tickets is required, however, it is possible to identify trends temporal occurring performance problems. The root course of the issue can then be evaluated further by technical staff using the tickets in the performance ticket groups.

The preliminary results of the proposed algorithm are promising, but a lot of optimization potential remains. In the next steps, the impact of other similarity metrics, e.g., the Cosine-measure, will be evaluated. Further, more sophisticated methods for evaluating ticket similarity will integrated, e.g., considering n-grams or content based evaluations.

Acknowledgement

This work is supported by the Deutsche Forschungsgemeinschaft (DFG) under Grants HO TR 257/41–1 “Trade-offs between QoE and Energy Efficiency in Data Centers”. The authors alone are responsible for the content.

References

1. Qualinet White Paper on Definitions of Quality of Experience (2012) Version 1.2. *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, March 2013.
2. W. W. Cohen, P. Ravikumar, and S. E. Fienberg. A comparison of string distance metrics for name-matching tasks. In *Proc. Workshop on Information Integration on the Web*, Acapulco, Mexico, August 2003.
3. T. Hoßfeld, R. Schatz, M. Varela, and C. Timmerer. Challenges of QoE Management for Cloud Applications. *IEEE Communications Magazine*, April 2012.