

Predicting the Programming Languages Tags in Stack Overflow

Authors

Abstract—

Index Terms—Stack Overflow, Machine learning, Tag, Code Snippet, Programming Languages and Natural Language Processing.

I. INTRODUCTION

II. BACKGROUND AND RELATED WORK

Predicting a programming language from a given text and from code snippets has been a rising topic of interest in the research community. Despite of a lot of interesting contributions to this field, we are still lacking behind in predicting the exact programming language just on the basis of text or the code snippet.

III. METHODOLOGY

IV. RESULTS AND DISCUSSIONS

V. THREATS TO VALIDITY

VI. CONCLUSIONS

APPENDIX

REFERENCES

- [1] A. K. Saha, R. K. Saha, and K. A. Schneider, A discriminative model approach for suggesting tags automatically for stack overflow questions, in Proceedings of the International Workshop on Mining Software Repositories. IEEE Press, 2013, pp. 7376.
- [2] C. Stanley and M. D. Byrne. Predicting Tags for StackOverflow Posts. In Proceedings of ICCM 2013 (12th International Conference on Cognitive Modeling), 2013.
- [3] S. M. Nasehi, J. Sillito, F. Maurer, and C. Burns, What makes a good code example? A study of programming Q&A in StackOverflow, in Proc. ICSM, pages 2535, 2012.
- [4] J. Brandt, P. J. Guo, J. Lewenstein, M. Dontcheva, and S. R. Klemmer, Two Studies of Opportunistic Programming: Interleaving Web Foraging, Learning, and Writing Code, in Proceedings of CHI 2009, New York, NY, USA, 2009, pp. 15891598.
- [5] C. Treude, O. Barzilay, and M.-A. Storey, How Do Programmers Ask and Answer Questions on the Web? in Proceedings of ICSE 2011, New York, NY, USA, 2011, pp. 804807.
- [6] C. McMillan, M. Grechanik, D. Poshyvanyk, Q. Xie, and C. Fu, Portfolio: A Search Engine for Finding Functions and Their Usages, in Proceedings of ICSE 2011, 2011, pp. 10431045.
- [7] M. Revelle, B. Dit, and D. Poshyvanyk, Using Data Fusion and Web Mining to Support Feature Location in Software, in Proceedings of ICPC 2010, 2010, pp. 1423.
- [8] R. Holmes, R. J. Walker, and G. C. Murphy, Strathcona Example Recommendation Tool, in ACM SIGSOFT Software Engineering Notes, New York, NY, USA, 2005, pp. 237240.
- [9] B. Seaman, The Information Gathering Strategies of Software Maintainers, in Proceedings of ICSM 2002, 2002, pp. 141 149.
- [10] V. S. Rekha, N. Divya, and P. S. Bagavathi. A hybrid auto-tagging system for stackoverflow forum questions. In Proceedings of the 2014 International Conference on Interdisciplinary Advances in Applied Computing, ICONIAAC 14, pages 56:156:5, New York, NY, USA, 2014. ACM.
- [11] Algorithmia, ProgrammingLanguageIdentification tool”, 2017. [Online]. Available: <https://www.algorithmia.com>.
- [12] E.Loper, S. Bird, NLTK: The natural language toolkit, in Proc. Interact. Present. Sessions Association for Computational Linguistics, 2006, pp. 6972.
- [13] F.Pedregosa,G.Varoquaux,A.Gramfort,V.Michel,B.Thirion,O.Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research, 2011.
- [14] L. Mamykina, B. Manoim, M. Mittal, G. Hripcsak, and B. Hartmann, Design Lessons from the Fastest Q&A Site in the West, in Proceedings of the 2011 annual conference on Human factors in computing systems, New York, NY, USA, 2011, pp. 28572866.
- [15] M. Asaduzzaman, A. S. Mashiyat, C. K. Roy, and K. A. Schneider. Answering questions about unanswered questions of stack overflow. In MSR, 2013.
- [16] SoStats, 2017. [Online]. Available: <https://sostats.github.io/last30days/>
- [17] Quantcast, 2017. [Online]. Available: <https://www.quantcast.com>