Analyzing and Summarizing User Contributions in Wikis

by

Fong Kin Fong

Master of Science in Software Engineering

2013

Faculty of Science and Technology
University of Macau

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

Analyzing and Summarizing User Contributions in Wikis

by

Fong Kin Fong

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Software Engineering

Faculty of Science and Technology
University of Macau

2013

Approved by _____
Supervisor

Date _____

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Macau, I agree that the Library and the Faculty of Science and Technology shall make its copies freely available for inspection. However, reproduction of this thesis for any purposes or by any means shall not be allowed without my written permission. Authorization is sought by contacting the author at

E-mail: peter.kf.fong@gmail.com

Signature _____

Date _____

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

University of Macau

Abstract

ANALYZING AND SUMMARIZING USER CONTRIBUTIONS IN WIKIS

by Fong Kin Fong

Thesis Supervisor: Assistant Professor, Dr. Robert P. Biuk-Aghai
Master of Science in Software Engineering

Wiki system, which allows users to collaboratively write articles online in a convenient way, has become popular in recent years. By design, wiki encourage a large number of users to make many edits rapidly. This characteristic makes it impractical to identify editor's contributions of an article by examining each edit manually. However, finding major contributors are, in many cases, useful for wiki community members. Therefore, a method to automatically identify and summarize user contribution can be helpful. In this thesis, we proposed a framework to programmatically analyze wiki article edits, present it in a way that is close to human understanding, and measure the significance of the edit using a metric. With the significance value of each edit on hand, we can then perform different kinds of aggregation to find out who is the major contributor of an article, what topic writers are interested, what type of contribution are made by them, and many others.

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

Article: A content page in a wiki.

Contributor: A person who edits an article in a wiki.

Edit: Modification of the content in a page, which creates a new version of the page.

Namespace: Namespaces are used in wikis to separate content by purpose. For example, articles are stored in the main namespace, while discussions are stored in the Talk namespace.

Page: The basic element in MediaWiki, which can be an article, a category or other object, which are differentiated by their namespace. A page is uniquely identified by its page ID.

Version: A unique state of a document, which represents the content on the document at a certain date and time.

Wikitext: A document written in a wiki markup language.

ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to my supervisor, Dr. Robert P. Biuk-Aghai, who kept guiding me and encouraging me along the way of this research. With his great vision and experience, he helped me to find my research direction, inspired me with different new ideas, pointed the way to solve the problems and suggested numerous possible improvements. It is impossible for me to complete this thesis without his leadership and experience. Moreover, he has put a lot of effort in helping me to improve my academic writing skills. I truly appreciate his effort, patience and help during these years of supervision.

I would also like to thank teachers and students, both former and current, in the Faculty of Science and Technology, especially Mr. Miguel Gomes da Costa Junior and Mr. Pang Cheong-Iao, Patrick, for generously sharing their knowledge and experience related to this research work.

Last but not least, I would like to thank my parents and all my friends for their support and encouragement on both my research and my life.