



# UNIVERSITÀ DI TRENTO

Dipartimento di Ingegneria e Scienza dell'Informazione

Corso di Laurea in  
Informatica

ELABORATO FINALE

## TITOLO

*Sottotitolo (alcune volte lungo - opzionale)*

Supervisore  
.....

Laureando  
Gandelli Alessio

Anno accademico .../...

# Ringraziamenti

*...thanks to...*

# Contents

<b>Sommario</b>	<b>2</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Main Project . . . . .	3
1.2 My Contribution . . . . .	4
1.3 Related Work . . . . .	4
1.4 Context . . . . .	4
<b>2 Background</b>	<b>5</b>
<b>3 Methods</b>	<b>6</b>
3.1 Dataset . . . . .	6
3.2 Approach . . . . .	6
<b>4 Results and Discussion</b>	<b>7</b>
<b>5 Infrastructure</b>	<b>8</b>
<b>6 Conclusions</b>	<b>9</b>
<b>Bibliografia</b>	<b>9</b>
<b>A Titolo primo allegato</b>	<b>11</b>
A.1 Titolo . . . . .	11
A.1.1 Sottotitolo . . . . .	11
<b>B Titolo secondo allegato</b>	<b>12</b>
B.1 Titolo . . . . .	12
B.1.1 Sottotitolo . . . . .	12

# Sommario

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

Sommario è un breve riassunto del lavoro svolto dove si descrive l'obiettivo, l'oggetto della tesi, le metodologie e le tecniche usate, i dati elaborati e la spiegazione delle conclusioni alle quali siete arrivati.

Il sommario dell'elaborato consiste al massimo di 3 pagine e deve contenere le seguenti informazioni:

- contesto e motivazioni
- breve riassunto del problema affrontato
- tecniche utilizzate e/o sviluppate
- risultati raggiunti, sottolineando il contributo personale del laureando/a

# Chapter 1

## Introduction

Wikipedia (WP) is the biggest source of information you can find on the internet, there are more than 6 million articles and they are all maintained by volunteers. The value of Wikipedia is all in the hands of the editors.

Many articles mean many users and therefore many potential conflicts. Avoiding these conflicts is the best way for this encyclopedia to develop.

Each Wikipedia page has four different sections:

- Article: the actual content of the page.
- Talk Page: a forum where people can talk about edits.
- History: a place where everyone can see the older versions of the pages.
- Source: in this section users can edit the page.

Conflicts could happen both on the Talk page, through a discussion, and in the Article, through an edit war. It is important to analyze all these aspects

### 1.1 Main Project

The project in collaboration between Eurecat and the Wikimedia Foundation is named: "Community Health Metrics: Understanding Editor Drop-off" and this is a part of the project idea (?):

"The primary value of Wikipedia is the editors. When an editor leaves the project, we lose their participation and contribution to the community, This could be related to multiple factors, also external to the project, but it could signal an issue related to internal dynamics and to the health of the community. While a big effort was dedicated to retain new editors, we lack knowledge and initiatives focused on understanding and preventing drop-off for experienced editors."

As stated in the description of the project, the focus is on the experienced users, which are the core of Wikipedia: there are 41,741,926 Wikipedia accounts but the active users are only 132,916 namely 3% of all users.

Focusing on this category of users and understanding the reasons that lead to a drop-off can give a big help to WP. Several people are working on this project, this work is just a part of the whole. In the

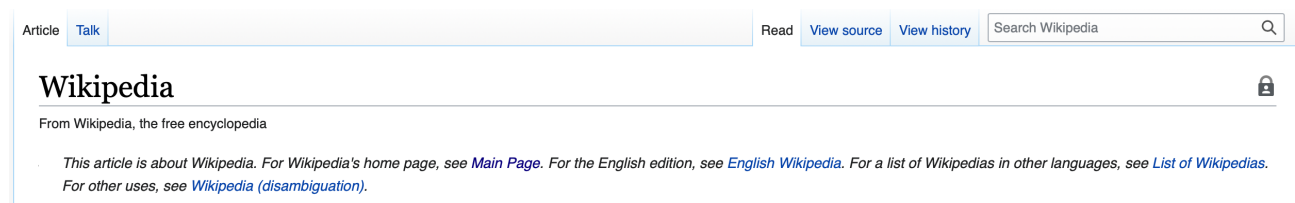


Figure 1.1: page structure

team, everyone is working on a specific topic and the idea is to combine the different results to obtain an analysis of the phenomenon from different points of view in order to have a greater understanding of the life cycle of users.

The prevention of the drop-off is not the only goal of the project, improving the community health is also important to let users be in a good environment without being held back from editing.

## **1.2 My Contribution**

The topic explored in this study is the revert analysis - i.e., when the version of a page is restored to that of a specific date - for all the articles of Wikipedia.

This project consisted of the analysis of the edit history of different language editions of Wikipedia to study patterns of reverts and edit wars to understand their potential effect on individual user activity.

I implemented state-of-the-art metrics of controversy based on reverts and mutual reverts and developed a new metric based on revert chains. Metrics have been computed per page and per user monthly.

## **1.3 Releated Work**

There are different works that involves reverts: An interesting tool that allows to visualize the conflicts

## **1.4 Context**

[?]

[1] <https://stats.wikimedia.org/EN/TablesWikipediansEditsGt5.htm>

## Chapter 2

# Background

ciao

# Chapter 3

## Methods

### 3.1 Dataset

The dataset used is the wikimedia history dumps which is a large dataset derived from - TSV - each line is an event - event types: revision page user

i'm interested in revision event  
here you the schema overview  
more detailed info there

### 3.2 Approach

Considering the huge dimension of the dataset and the fact that a large portion of its content was useless, smaller datasets have been computed with the aim of expediting the analysis even for future usages.

the first skimming was to filter the rows, after this only the revisions which made o received a revert is kept. then i ordered it by page, this dataset which is 10 % of the original one is the one i used to compute the new dataset...

while this is just the biggest one without some rows, the structure of the computed one is different. there are 2 modules

- Chains
- Group

image slide 5



## Chapter 4

# Results and Discussion

## Chapter 5

# Infrastructure

## Chapter 6

# Conclusions

# Bibliography

# Allegato A    Titolo primo allegato

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

## A.1    Titolo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

### A.1.1    Sottotitolo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

# Allegato B      Titolo secondo allegato

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

## B.1      Titolo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.

### B.1.1      Sottotitolo

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec sed nunc orci. Aliquam nec nisl vitae sapien pulvinar dictum quis non urna. Suspendisse at dui a erat aliquam vestibulum. Quisque ultrices pellentesque pellentesque. Pellentesque egestas quam sed blandit tempus. Sed congue nec risus posuere euismod. Maecenas ut lacus id mauris sagittis egestas a eu dui. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque at ultrices tellus. Ut eu purus eget sem iaculis ultricies sed non lorem. Curabitur gravida dui eget ex vestibulum venenatis. Phasellus gravida tellus velit, non eleifend justo lobortis eget.