

Wikipedia Recommender System with serendipity

Barszezak Yoann, Bricout Raphael, David Nicolas, Dupre Remi,
Fouche Aziz, Gourdel Garance, Kaddar Younesse, Mallem Maher,

Department of Computer science, ENS Paris-Saclay

November 2017

TO DO : Abstract

1 Introduction

2 Problem Overview

3 Retrival of Candidate articles

3.1 Neighbourhood

3.1.1 Naive distance

Definition 3.1. Let A_1 and A_2 be two articles. We call C_1 (resp. C_2) the set of categories linked to A_1 (resp. A_2). We define the naive distance between A_1 and A_2 the following quantity:

$$d(A_1, A_2) = \frac{\min(\text{Card}(C_1), \text{Card}(C_2))}{\text{Card}(C_1 \cap C_2)}$$

3.1.2 Ontology basis

3.2 serendipity

4 Candidate Ranking

4.1 possible input in neural network

4.1.1 Category vector

4.1.2 Word2Vect and Wikidata

4.2 Wide and deep Neural Network

4.2.1 Wide : memorization/focus

4.2.2 Deep : generalization/serendipity

5 System architecture

6 Experimentation

7 State of the Art