B.Sc. in Computer Science and Engineering Thesis

Building A Knowledge Graph Using Twitter Data

Submitted by

Sukanta Kundu
201205068

Anindya Biswas
201405006

Qazi Fahim Farhan
201405052

Supervised by

Dr. Muhammad Masroor Ali

Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology

Dhaka, Bangladesh

February 2018

CANDIDATES' DECLARATION

This is to certify that the work presented in this thesis, titled, "Building A Knowledge Graph Using Twitter Data", is the outcome of the investigation and research carried out by us under the supervision of Dr. Muhammad Masroor Ali.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Sukanta Kundu
201205068

Anindya Biswas
201405006

Qazi Fahim Farhan
201405052

CERTIFICATION

This thesis titled, **"Building A Knowledge Graph Using Twitter Data"**, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Computer Science and Engineering in February 2018.

Group Members:

Sukanta Kundu

Anindya Biswas

Qazi Fahim Farhan

Supervisor:

Dr. Muhammad Masroor Ali
Professor
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology

ACKNOWLEDGEMENT

We are indebted to our supervisor Professor Dr. Muhammad Mashroor Ali got introduing us to the Knowledge Graph. We've learned how to think critically like a researcher, carry on a research, read scientific papers, and how to write , speak and present properly. We show our gratitude for his dedication and patience in correcting our mistakes, reviewing our ideas, methodologies, proofs, etc. We couldn't have done it without him.

Dhaka
February 2018

Sukanta Kundu

Anindya Biswas

Qazi Fahim Farhan

Contents

List of Figures

List of Tables

List of Algorithms

ABSTRACT

In this paper, we wish to work on creating a Knowledge Graph using Twitter micro-data. A Knowledge Graph is a representation of human knowledge in the form of a graph. Just like any graph, it has nodes (entities), attributes and edges (relations). Since tweets give us limited data (maximum length of 300 characters), it becomes very difficult to extract adequate information to create a knowledge graph.

Chapter 1

Introduction

This is our introduction

Chapter 2

Related Works

Previously there have been many researches conducted on knowledge graphs. Most of them primarily identify major terms or keywords, remove noise, fetching missing information and disambiguate them to appropriate topic.

To disambiguate and relate a concept with similar concept, researchers have used Probabilistic Soft Logic (PSL) withontological constraints, identify coreferences etc, as well as machine learning, deep neural networks. There are many links among different concepts. Mihalcea annd Csomai et al worked onlink detection. Their work is on linking phrases to appropriate concept. To disambiguate in link detetion, they took help of the surrounding words (the words themselves and their parts of speech), and compare with training data. After that, in order to organize the documents, they have used topic indexing. In order to make link detection efficient, Mihalcea annd Csomai et al used (a) commonness (prior probability of each sense) and (b) how the sense relate with the nearby context.

Next we have an important work of Larry Heck and Hongzhao Huang. First they developed a robust way to represent concepts. Many NLP tools use string concatenation to represent ideas. But most of the time it fails to represent appropriate concepts. For example 'Madrid' is a place, 'Real' means something authentic or true, but 'Real Madrid' referres to a football club. This is a counter example where string concatenation fails. A possible alternative is to use word hashing with n-gram (tri-gram) word representation. This way can be used to represent any words that are not even seen before. After that they did neural embedding of knowledge graph. To do so, they tracked a concept and its corresponding subgraph, encode the knowledge as featured vector. then they trained Deep Neural Network to get semantic relationship.

Chapter 3

Another Chapter

3.1 A Section

Some text.

3.1.1 This is a Subsection

And some more.

This is a Subsubsection

Yet some more.

3.2 And Another Section

Here are some dummy texts.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras et ultricies massa. Nulla a sapien lobortis, dignissim nibh in, aliquet mauris. Integer at dictum metus. Quisque in tortor congue ipsum ultricies tristique. Maecenas ut tortor dapibus, sagittis enim at, tincidunt massa. Ut sollicitudin sagittis ipsum, ac tincidunt quam gravida ac. Nullam quis faucibus purus. Aliquam vel pretium turpis. Aliquam a quam non ex interdum sagittis id vitae quam. Nullam sodales ligula malesuada maximus consequat. Proin a justo eget lacus vulputate maximus luctus vitae enim. Aliquam libero turpis, pharetra a tincidunt ac, pulvinar sit amet urna. Pellentesque eget rutrum diam, in faucibus sapien. Aenean sit amet est felis. Aliquam dolor eros, porttitor quis volutpat eget, posuere a ligula. Proin id velit ac lorem finibus pellentesque.

Maecenas vitae interdum mi. Aenean commodo nisl massa, at pharetra libero cursus vitae. In hac habitasse platea dictumst. Suspendisse iaculis euismod dui, et cursus diam. Nullam euismod, est ut dapibus condimentum, lorem eros suscipit risus, sit amet hendrerit justo tortor nec lorem. Morbi et mi eget erat bibendum porta. Ut tristique ultricies commodo. Nullam iaculis ligula sed lacinia ornare.

Sed ultricies cursus nisi at vestibulum. Aenean laoreet viverra efficitur. Ut eget sapien lorem. Mauris malesuada, augue in pulvinar consectetur, ex tortor tristique ligula, sit amet faucibus metus lectus interdum nisl. Nam eget turpis vitae ligula pulvinar bibendum a ut ipsum. Mauris fringilla lacinia malesuada. Fusce id orci velit. Donec tristique rhoncus urna, a hendrerit arcu vehicula imperdiet. Integer tristique erat at gravida condimentum. Sed ornare cursus quam, eget tincidunt enim bibendum sed. Aliquam elementum ligula scelerisque leo sagittis, quis convallis elit dictum. Donec sit amet orci aliquam, ultricies sapien nec, gravida nisi. Etiam et pulvinar diam, et pellentesque arcu. Nulla interdum metus sed aliquet consequat.

Proin in mi id nulla interdum aliquet ac quis arcu. Duis blandit sapien commodo turpis hendrerit pharetra. Phasellus sit amet justo orci. Proin mattis nisl dictum viverra fringilla. Interdum et malesuada fames ac ante ipsum primis in faucibus. Curabitur facilisis euismod augue vestibulum tincidunt. Nullam nulla quam, volutpat vitae efficitur eget, porta sit amet nunc. Phasellus pharetra est eget urna ornare volutpat. Aenean ultrices, libero eget porttitor fringilla, purus tortor accumsan neque, sit amet viverra felis tortor eget justo. Nunc id metus a purus tempus euismod condimentum non lacus. Nam vitae diam aliquam, facilisis diam quis, pharetra nunc. Nulla eget vestibulum tellus, ut cursus tellus. Vestibulum euismod pellentesque sodales.

Maecenas at mi interdum, faucibus lorem sed, hendrerit nisi. In vitae augue consequat diam commodo porta sit amet eu purus. Mauris mattis condimentum feugiat. Nulla commodo molestie risus vitae maximus. Proin hendrerit neque malesuada urna laoreet convallis. Etiam a diam pulvinar, auctor sem ac, hendrerit risus. Ut urna urna, venenatis ac tellus non, scelerisque tristique ligula. Vestibulum sollicitudin vel leo malesuada accumsan. Donec sit amet erat diam. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Vivamus odio dui, scelerisque et lorem egestas, posuere ullamcorper nunc. Integer varius nunc nec velit tincidunt commodo. Mauris rhoncus ultrices sapien non suscipit.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras et ultricies massa. Nulla a sapien lobortis, dignissim nibh in, aliquet mauris. Integer at dictum metus. Quisque in tortor congue ipsum ultricies tristique. Maecenas ut tortor dapibus, sagittis enim at, tincidunt massa. Ut sollicitudin sagittis ipsum, ac tincidunt quam gravida ac. Nullam quis faucibus purus. Aliquam vel pretium turpis. Aliquam a quam non ex interdum sagittis id vitae quam. Nullam sodales ligula malesuada maximus consequat. Proin a justo eget lacus vulputate maximus luctus vitae enim. Aliquam libero turpis, pharetra a tincidunt ac, pulvinar sit amet urna. Pellentesque eget rutrum diam, in faucibus sapien. Aenean sit amet est felis. Aliquam dolor eros, porttitor quis

volutpat eget, posuere a ligula. Proin id velit ac lorem finibus pellentesque.

Maecenas vitae interdum mi. Aenean commodo nisl massa, at pharetra libero cursus vitae. In hac habitasse platea dictumst. Suspendisse iaculis euismod dui, et cursus diam. Nullam euismod, est ut dapibus condimentum, lorem eros suscipit risus, sit amet hendrerit justo tortor nec lorem. Morbi et mi eget erat bibendum porta. Ut tristique ultricies commodo. Nullam iaculis ligula sed lacinia ornare.

Sed ultricies cursus nisi at vestibulum. Aenean laoreet viverra efficitur. Ut eget sapien lorem. Mauris malesuada, augue in pulvinar consectetur, ex tortor tristique ligula, sit amet faucibus metus lectus interdum nisl. Nam eget turpis vitae ligula pulvinar bibendum a ut ipsum. Mauris fringilla lacinia malesuada. Fusce id orci velit. Donec tristique rhoncus urna, a hendrerit arcu vehicula imperdiet. Integer tristique erat at gravida condimentum. Sed ornare cursus quam, eget tincidunt enim bibendum sed. Aliquam elementum ligula scelerisque leo sagittis, quis convallis elit dictum. Donec sit amet orci aliquam, ultricies sapien nec, gravida nisi. Etiam et pulvinar diam, et pellentesque arcu. Nulla interdum metus sed aliquet consequat.

Proin in mi id nulla interdum aliquet ac quis arcu. Duis blandit sapien commodo turpis hendrerit pharetra. Phasellus sit amet justo orci. Proin mattis nisl dictum viverra fringilla. Interdum et malesuada fames ac ante ipsum primis in faucibus. Curabitur facilisis euismod augue vestibulum tincidunt. Nullam nulla quam, volutpat vitae efficitur eget, porta sit amet nunc. Phasellus pharetra est eget urna ornare volutpat. Aenean ultrices, libero eget porttitor fringilla, purus tortor accumsan neque, sit amet viverra felis tortor eget justo. Nunc id metus a purus tempus euismod condimentum non lacus. Nam vitae diam aliquam, facilisis diam quis, pharetra nunc. Nulla eget vestibulum tellus, ut cursus tellus. Vestibulum euismod pellentesque sodales.

Maecenas at mi interdum, faucibus lorem sed, hendrerit nisi. In vitae augue consequat diam commodo porta sit amet eu purus. Mauris mattis condimentum feugiat. Nulla commodo molestie risus vitae maximus. Proin hendrerit neque malesuada urna laoreet convallis. Etiam a diam pulvinar, auctor sem ac, hendrerit risus. Ut urna urna, venenatis ac tellus non, scelerisque tristique ligula. Vestibulum sollicitudin vel leo malesuada accumsan. Donec sit amet erat diam. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Vivamus odio dui, scelerisque et lorem egestas, posuere ullamcorper nunc. Integer varius nunc nec velit tincidunt commodo. Mauris rhoncus ultrices sapien non suscipit.

Chapter 4

Index Creation

4.1 BUET

Bangladesh University of Engineering and Technology, abbreviated as BUET, is one of the most prestigious institutions for higher studies in the country. About 5500 students are pursuing undergraduate and postgraduate studies in engineering, architecture, planning and science in this institution. At present, BUET has sixteen teaching departments under five faculties and it has three institutes. Every year the intake of undergraduate students is around 900, while the intake of graduate students in Master's and PhD programs is around 1000. A total of about five hundred teachers are teaching in these departments and institutes. There are additional teaching posts like Dr. Rashid Professor, Professor Emeritus and Supernumerary Professors.

4.2 Campus

The BUET campus is in the heart of Dhaka — the capital city of Bangladesh. It has a compact campus with halls of residence within walking distances of the academic buildings. The physical expansion of the University over the last three decades has been impressive with construction of new academic buildings, auditorium complex, halls of residence, etc.

4.3 History

BUET is the oldest institution for the study of Engineering and Architecture in Bangladesh. The history of this institution dates back to the days of Dhaka Survey School which was established at Nalgola, in Old Dhaka in 1876 to train Surveyors for the then Government of Bengal of British India. As the years passed, the Survey School became the Ahsanullah School of En-

gineering offering three-year diploma courses in Civil, Electrical and Mechanical Engineering. In recognition of the generous financial contribution from the then Nawab of Dhaka, it was named after his father Khawja Ahsanullah. It moved to its present premises in 1912. In 1947, the School was upgraded to Ahsanullah Engineering College as a Faculty of Engineering under the University of Dhaka, offering four-year bachelors courses in Civil, Electrical, Mechanical, Chemical and Metallurgical Engineering. In order to create facilities for postgraduate studies and research, Ahsanullah Engineering College was upgraded to the status of a University in 1962 and was named East Pakistan University of Engineering and Technology. After the War of Liberation in 1971, Bangladesh became an independent state and the university was renamed as the Bangladesh University of Engineering and Technology.

4.4 Students

Till today, it has produced around 25,000 graduates in different branches of engineering and architecture, and has established a good reputation all over the world for the quality of its graduates, many of whom have excelled in their profession in different parts of the globe. It was able to attract students from countries like India, Nepal, Iran, Jordan, Malaysia, Sri Lanka, Pakistan and Palestine.

4.5 Departments

Both Undergraduate and Postgraduate studies and research are now among the primary functions of the University. Eleven departments under five faculties offer Bachelor Degrees, while most of the departments and institutes offer Master's Degrees and some of the departments have Ph.D. programs. In addition to its own research programs, the university undertakes research programs sponsored by outside organizations like European Union, UNO, Commonwealth, UGC, etc. The expertise of the University teachers and the laboratory facilities of the University are also utilized to solve problems and to provide up-to-date engineering and technological knowledge to the various organizations of the country.

Chapter 5

k-safe Labeling of Petersen Graph

In 1898, Petersen produced a trivalent graph with no leaves, now called the Petersen graph [1]. In this chapter we study k-safe labeling for the Petersen graph. We also give upper bound for the span of the Petersen graph. We provide necessary proof for the upper bound.

References

[1] D. A. Holton and J. Sheehan, *The Petersen Graph*, vol. 7. Cambridge University Press, 1993.

Index

Appendix A

Algorithms

A.1 Sample Algorithm

In Algorithm 1 we show how to calcute $y = x^n$.

Algorithm 1 Calculate $y = x^n$

Require: $n \geq 0 \vee x \neq 0$
Ensure: $y = x^n$
  $y \leftarrow 1$
  if $n < 0$ **then**
    $X \leftarrow 1/x$
    $N \leftarrow -n$
  else
    $X \leftarrow x$
    $N \leftarrow n$
  end if
  while $N \neq 0$ **do**
    if N is even **then**
      $X \leftarrow X \times X$
      $N \leftarrow N/2$
    else $\{N$ is odd$\}$
      $y \leftarrow y \times X$
      $N \leftarrow N - 1$
    end if
  end while

Appendix B

Codes

B.1 Sample Code

We use this code to find out...

```
#include <stdio.h>
int Fibonacci(int);

main()
{
  int n, i = 0, c;

  printf("Enter the value of n: ");
  scanf("%d",&n);

  printf("\nFibonacci series\n");

  for (c = 1 ; c <= n ; c++)
    {
      printf("%d\n", Fibonacci(i));
      i++;
    }

  return 0;
}

int Fibonacci(int n)
{
```

```
24   if (n == 0)
25     return 0;
26   else if (n == 1)
27     return 1;
28   else
29     return (Fibonacci(n-1) + Fibonacci(n-2));
30 }
```

B.2 Another Sample Code

```
1  SELECT associations2.object_id, associations2.term_id,
2         associations2.cat_ID, associations2.term_taxonomy_id
3  FROM (SELECT objects_tags.object_id, objects_tags.term_id,
4       wp_cb_tags2cats.cat_ID, categories.term_taxonomy_id
5  FROM (SELECT wp_term_relationships.object_id,
6       wp_term_taxonomy.term_id, wp_term_taxonomy.term_taxonomy_id
7  FROM wp_term_relationships
8  LEFT JOIN wp_term_taxonomy ON
9       wp_term_relationships.term_taxonomy_id =
10      wp_term_taxonomy.term_taxonomy_id
11 ORDER BY object_id ASC, term_id ASC)
12 AS objects_tags
13 LEFT JOIN wp_cb_tags2cats ON objects_tags.term_id =
14      wp_cb_tags2cats.tag_ID
15 LEFT JOIN (SELECT wp_term_relationships.object_id,
16      wp_term_taxonomy.term_id as cat_ID,
17      wp_term_taxonomy.term_taxonomy_id
18 FROM wp_term_relationships
19 LEFT JOIN wp_term_taxonomy ON
20      wp_term_relationships.term_taxonomy_id =
21      wp_term_taxonomy.term_taxonomy_id
22 WHERE wp_term_taxonomy.taxonomy = 'category'
23 GROUP BY object_id, cat_ID, term_taxonomy_id
24 ORDER BY object_id, cat_ID, term_taxonomy_id)
25 AS categories on wp_cb_tags2cats.cat_ID = categories.term_id
26 WHERE objects_tags.term_id = wp_cb_tags2cats.tag_ID
27 GROUP BY object_id, term_id, cat_ID, term_taxonomy_id
28 ORDER BY object_id ASC, term_id ASC, cat_ID ASC)
29 AS associations2
30 LEFT JOIN categories ON associations2.object_id =
```

```
31      categories.object_id
32 WHERE associations2.cat_ID <> categories.cat_ID
33 GROUP BY object_id, term_id, cat_ID, term_taxonomy_id
34 ORDER BY object_id, term_id, cat_ID, term_taxonomy_id
```

Generated using Undegraduate Thesis LaTeX Template, Version 1.4. Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

This thesis was generated on Tuesday 20th November, 2018 at 11:46am.