

Title (EN): Domain-Specific NER Adaptation

In the past years, the Named Entity Recognition (NER) technology has been under an active development and enjoy a significant increase in popularity and usage in the academic and industrial sphere. Nevertheless, vast majority of the developed NER systems have been developed as general-purpose systems. While they can perform well on multiple domains (macro level), on specific domains (micro level) their performance quality might be low. The ultimate goal of the thesis is to develop domain-specific NER models. Guidelines:

- Get familiar with the NER technology and available NER frameworks.
- Investigate possible datasets for domain-specific training of NER.
- Develop NER training datasets for several selected domains (e.g. sports, politics, music, etc.).
- Train a domain-specific NER model using existing frameworks, such as DBpedia Spotlight or StanfordNER.
- Validate and evaluate the developed domain-specific NER models.





**FACULTY  
OF INFORMATION  
TECHNOLOGY  
CTU IN PRAGUE**

Master's thesis

# **Domain-specific Name Entity Recognition**

*Bc. Bogoljub Jakovcheski*

Department of software engineering

Supervisor: Ing. Milan Dojchinovski, Ph.D.

May 21, 2018



---

# Acknowledgements

I would like to thank my family and friends for support during writing this thesis.



---

# Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended. In accordance with Article 46(6) of the Act, I hereby grant a nonexclusive authorization (license) to utilize this thesis, including any and all computer programs incorporated therein or attached thereto and all corresponding documentation (hereinafter collectively referred to as the “Work”), to any and all persons that wish to utilize the Work. Such persons are entitled to use the Work in any way (including for-profit purposes) that does not detract from its value. This authorization is not limited in terms of time, location and quantity. However, all persons that makes use of the above license shall be obliged to grant a license at least in the same scope as defined above with respect to each and every work that is created (wholly or in part) based on the Work, by modifying the Work, by combining the Work with another work, by including the Work in a collection of works or by adapting the Work (including translation), and at the same time make available the source code of such work at least in a way and scope that are comparable to the way and scope in which the source code of the Work is made available.

In Prague on May 21, 2018

.....

Czech Technical University in Prague

Faculty of Information Technology

© 2018 Bogoljub Jakovcheski. All rights reserved.

*This thesis is school work as defined by Copyright Act of the Czech Republic. It has been submitted at Czech Technical University in Prague, Faculty of Information Technology. The thesis is protected by the Copyright Act and its usage without author's permission is prohibited (with exceptions defined by the Copyright Act).*

### **Citation of this thesis**

Jakovcheski, Bogoljub. *Domain-specific Name Entity Recognition*. Master's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2018.



---

# Abstrakt

---

Klíčová slova

# Abstract

Keywords



---

# Contents

Citation of this thesis . . . . .	vi
<b>Introduction</b>	<b>1</b>
Motivation . . . . .	1
Goals of the thesis . . . . .	1
Thesis outline . . . . .	1
<b>1 Background and related work</b>	<b>3</b>
1.1 Information extraction . . . . .	3
1.2 NER . . . . .	3
1.2.1 Stanford NER . . . . .	4
1.3 RDF/NIF . . . . .	5
1.4 Domain specific NER . . . . .	5
1.5 DBpedia ontology . . . . .	5
1.6 Evaluation/training datasets . . . . .	5
<b>2 Domain specific NER</b>	<b>7</b>
2.1 Data pre-processing . . . . .	7
2.2 Types retrieval . . . . .	7
2.3 Domain specification . . . . .	7
2.4 Data transformation/output for Stanford . . . . .	7
<b>3 Experiments</b>	<b>9</b>
3.1 Goals of the experiments . . . . .	9
3.2 List of experiments . . . . .	9
3.2.1 Top 10 Links . . . . .	9
3.2.2 Top 20 Links . . . . .	14
3.2.3 Top 40 Links . . . . .	19
3.2.4 Top 100 Links . . . . .	22
3.2.5 Top 300 Links . . . . .	25
3.2.6 Top 400 Links . . . . .	30

3.2.7	Top 500 Links . . . . .	34
3.2.8	MIXED . . . . .	38
3.2.9	With lower links on model . . . . .	40
3.2.10	With higher links on model . . . . .	48
3.3	Summary of results . . . . .	58
3.3.1	Graphs . . . . .	58
<b>Conclusion</b>		<b>59</b>
<b>A Contents of CD</b>		<b>61</b>

---

## List of Figures

3.2	All 3 Domains Fine Grained Top 500 Links With All 3 Domains Fine Grained Top 500 Links And All 3 Domains Fine Grained Top 500 Links With Lower Page Rank . . . . .	38
3.1	All 3 Domains Fine Grained Top 300 With All 3 Domains Fine Grained Top 500 Links And All 3 Domains Fine Grained Top 500 Links With Lower PageRank . . . . .	38
3.3	All 3 Domains Fine Grained Top 500 Links With All 3 Domains Fine Grained Top 500 Links With Lower PageRank . . . . .	39
3.4	Transportation Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links And Transportation Fine Grained Top 300 Links . . . . .	39
3.5	All 3 Domains Fine Grained Top 10 Links Runned With All 3 Domains Fine Grained Top 100 Links . . . . .	40
3.6	All 3 Domains Fine Grained Top 10 Links Runned With All 3 Domains Fine Grained Top 300 Links . . . . .	40
3.7	All 3 Domains Fine Grained Top 10 Links Runned With All 3 Domains Fine Grained Top 500 Links . . . . .	41
3.8	All 3 Domains Fine Grained Top 10 Links Runned With All 3 Domains Fine Grained Top 500 Links With Lower PageRank . . . . .	41
3.9	All 3 Domains Fine Grained Top 100 Links Runned With All 3 Domains Fine Grained Top 500 Links . . . . .	42
3.10	All 3 Domains Fine Grained Top 100 Links Runned With All 3 Domains Fine Grained Top 500 Links With Lower PageRank . . . . .	42
3.11	All 3 Domains Fine Grained Top 100 Links Runned With Politics Fine Grained Top 500 Links . . . . .	43
3.12	All 3 Domains Fine Grained Top 100 Links Runned With Politics Fine Grained Top 500 Links With Lower PageRank . . . . .	43
3.13	All 3 Domains Fine Grained Top 100 Links Runned With Sport Fine Grained Top 500 Links . . . . .	43

3.14	All 3 Domains Fine Grained Top 100 Links Runned With Sport Fine Grained Top 500 Links With Lower PageRank . . . . .	44
3.15	All 3 Domains Fine Grained Top 100 Links Runned With Trans- portation Fine Grained Top 500 Links . . . . .	44
3.16	All 3 Domains Fine Grained Top 100 Links Runned With Trans- portation Fine Grained Top 500 Links With Lower PageRank . . .	44
3.17	All 3 Domains Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 500 Links . . . . .	45
3.18	All 3 Domains Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 500 Links With Lower PageRank . . .	45
3.19	All 3 Domains Fine Grained Top 300 Links Runned With Politics Fine Grained Top 500 Links . . . . .	46
3.20	All 3 Domains Fine Grained Top 300 Links Runned With Politics Fine Grained Top 500 Links With Lower PageRank . . . . .	46
3.21	All 3 Domains Fine Grained Top 300 Links Runned With Sport Fine Grained Top 500 Links . . . . .	46
3.22	All 3 Domains Fine Grained Top 300 Links Runned With Sport Fine Grained Top 500 Links With Lower PageRank . . . . .	47
3.23	All 3 Domains Fine Grained Top 300 Links Runned With Trans- portation Fine Grained Top 500 Links . . . . .	47
3.24	All 3 Domains Fine Grained Top 300 Links Runned With Trans- portation Fine Grained Top 500 Links With Lower PageRank . . .	47
3.25	All 3 Domains Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 10 Links . . . . .	48
3.26	All 3 Domains Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 100 Links . . . . .	48
3.27	All 3 Domains Fine Grained Top 300 Links Runned With Politics Fine Grained Top 10 Links . . . . .	48
3.28	All 3 Domains Fine Grained Top 300 Links Runned With Politics Fine Grained Top 100 Links . . . . .	49
3.29	All 3 Domains Fine Grained Top 300 Links Runned With Sport Fine Grained Top 10 Links . . . . .	49
3.30	All 3 Domains Fine Grained Top 300 Links Runned With Sport Fine Grained Top 100 Links . . . . .	49
3.31	All 3 Domains Fine Grained Top 300 Links Runned With Trans- portation Fine Grained Top 10 Links . . . . .	49
3.32	All 3 Domains Fine Grained Top 300 Links Runned With Trans- portation Fine Grained Top 100 Links . . . . .	50
3.33	All 3 Domains Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 10 Links . . . . .	50
3.34	All 3 Domains Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 100 Links . . . . .	50
3.35	All 3 Domains Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links . . . . .	51

3.36	All 3 Domains Fine Grained Top 500 Links Runned With Politics Fine Grained Top 10 Links . . . . .	51
3.37	All 3 Domains Fine Grained Top 500 Links Runned With Politics Fine Grained Top 100 Links . . . . .	51
3.38	All 3 Domains Fine Grained Top 500 Links Runned With Politics Fine Grained Top 300 Links . . . . .	52
3.39	All 3 Domains Fine Grained Top 500 Links Runned With Sport Fine Grained Top 10 Links . . . . .	52
3.40	All 3 Domains Fine Grained Top 500 Links Runned With Sport Fine Grained Top 100 Links . . . . .	52
3.41	All 3 Domains Fine Grained Top 500 Links Runned With Sport Fine Grained Top 300 Links . . . . .	53
3.42	All 3 Domains Fine Grained Top 500 Links Runned With Trans- portation Fine Grained Top 10 Links . . . . .	53
3.43	All 3 Domains Fine Grained Top 500 Links Runned With Trans- portation Fine Grained Top 100 Links . . . . .	53
3.44	All 3 Domains Fine Grained Top 500 Links Runned With Trans- portation Fine Grained Top 300 Links . . . . .	54
3.45	Politics Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 100 Links . . . . .	54
3.46	Politics Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 100 Links . . . . .	55
3.47	Politics Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links . . . . .	55
3.48	Sport Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 100 Links . . . . .	56
3.49	Sport Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 100 Links . . . . .	56
3.50	Sport Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links . . . . .	57
3.51	Transportation Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 100 Links . . . . .	57
3.52	Transportation Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 100 Links . . . . .	58
3.53	Transportation Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links . . . . .	58





---

## List of Tables

3.1	TABLE	10
3.2	TABLE	10
3.3	TABLE	10
3.4	TABLE	10
3.5	TABLE	11
3.6	TABLE	11
3.7	TABLE	12
3.8	TABLE	12
3.9	TABLE	12
3.10	TABLE	13
3.11	TABLE	13
3.12	TABLE	13
3.13	TABLE	14
3.14	TABLE	14
3.15	TABLE	15
3.16	TABLE	15
3.17	TABLE	15
3.18	TABLE	15
3.19	TABLE	16
3.20	TABLE	16
3.21	TABLE	16
3.22	TABLE	17
3.23	TABLE	17
3.24	TABLE	17
3.25	TABLE	18
3.26	TABLE	18
3.27	TABLE	18
3.28	TABLE	18
3.29	TABLE	19
3.30	TABLE	19

3.31	TABLE	19
3.32	TABLE	19
3.33	TABLE	20
3.34	TABLE	20
3.35	TABLE	20
3.36	TABLE	21
3.37	TABLE	21
3.38	TABLE	21
3.39	TABLE	21
3.40	TABLE	21
3.41	TABLE	22
3.42	TABLE	22
3.43	TABLE	22
3.44	TABLE	22
3.45	TABLE	22
3.46	TABLE	23
3.47	TABLE	23
3.48	TABLE	23
3.49	TABLE	24
3.50	TABLE	24
3.51	TABLE	24
3.52	TABLE	24
3.53	TABLE	25
3.54	TABLE	25
3.55	TABLE	25
3.56	TABLE	25
3.57	All 3 domain model with top 300 abstracts, tested with all 3 domain texts	26
3.58	TABLE	26
3.59	TABLE	26
3.60	TABLE	26
3.61	TABLE	27
3.62	TABLE	27
3.63	TABLE	28
3.64	TABLE	28
3.65	TABLE	28
3.66	TABLE	28
3.67	TABLE	29
3.68	TABLE	29
3.69	TABLE	29
3.70	TABLE	29
3.71	TABLE	30
3.72	TABLE	30
3.73	TABLE	30

---

3.74	TABLE	30
3.75	TABLE	31
3.76	TABLE	31
3.77	TABLE	32
3.78	TABLE	32
3.79	TABLE	32
3.80	TABLE	33
3.81	TABLE	33
3.82	TABLE	33
3.83	TABLE	33
3.84	TABLE	33
3.85	TABLE	34
3.86	TABLE	34
3.87	TABLE	34
3.88	TABLE	34
3.89	TABLE	35
3.90	TABLE	35
3.91	TABLE	36
3.92	TABLE	36
3.93	TABLE	36
3.94	TABLE	37
3.95	TABLE	37
3.96	TABLE	37
3.97	TABLE	37
3.98	TABLE	37



---

# Introduction

## Motivation

## Goals of the thesis

The main goal of the thesis is to show (test) does extracting information from a global domain gives us better results than extracting information from a specific domain, for example politics domain or any other domain.

## Thesis outline



---

# Background and related work

## 1.1 Information extraction

Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. In most of the cases this activity concerns processing human language texts by means of natural language processing (NLP). Recent activities in multimedia document processing like automatic annotation and content extraction out of images/audio/video could be seen as information extraction.

Information extraction, that is, automatically building a relational database from information contained in unstructured text. Unlike linear-chain models, general CRFs can capture long distance dependencies between labels.

PICTURE!!!!

## 1.2 NER

NER is the problem of identifying and classifying proper names in text, including locations, such as China; people, such as George Bush; and organizations, such as the United Nations. The named-entity recognition task is, given a sentence, first to segment which words are part of entities, and then to classify each entity by type (person, organization, location, and so on). The challenge of this problem is that many named entities are too rare to appear even in a large training set, and therefore the system must identify them based only on context. EVALUATION OF NER Precision, recall, F1 measurement The first step in most IE tasks is to find the proper names or named entities mentioned in a text. The task of named entity recognition (NER) is to find each mention of a named entity in the text and label its type. What constitutes a named entity type is application specific; these commonly include people, places, and organizations but also more specific entities from the names of genes and proteins (Cohen and Demner-Fushman, 2014) to the names of college courses (McCallum, 2005).

## 1. BACKGROUND AND RELATED WORK

---

Named-entity recognition (NER) (also known as entity identification, entity chunking and entity extraction) is a subtask of information extraction that seeks to locate and classify named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

Most research on NER systems has been structured as taking an unannotated block of text, such as this one:

Jim bought 300 shares of Acme Corp. in 2006.

And producing an annotated block of text that highlights the names of entities:

[Jim]Person bought 300 shares of [Acme Corp.]Organization in [2006]Time.

In this example, a person name consisting of one token, a two-token company name and a temporal expression have been detected and classified.

State-of-the-art NER systems for English produce near-human performance. For example, the best system entering MUC-7 scored 93.39% of F-measure while human annotators scored 97.60% and 96.95%. PICTURE FROM Stanford NER GUI!!!!

### 1.2.1 Stanford NER

Stanford NER is a Java implementation of a Named Entity Recognizer. Named Entity Recognition (NER) labels sequences of words in a text which are the names of things, such as person and company names, or gene and protein names. It comes with well-engineered feature extractors for Named Entity Recognition, and many options for defining feature extractors. Included with the download are good named entity recognizers for English, particularly for the 3 classes (PERSON, ORGANIZATION, LOCATION), and we also make available on this page various other models for different languages and circumstances, including models trained on just the CoNLL 2003 English training data.

Stanford NER is also known as CRFClassifier. The software provides a general implementation of (arbitrary order) linear chain Conditional Random Field (CRF) sequence models. That is, by training your own models on labeled data, you can actually use this code to build sequence models for NER or any other task. (CRF models were pioneered by Lafferty, McCallum, and Pereira (2001); see Sutton and McCallum (2006) or Sutton and McCallum (2010) for more comprehensible introductions.)



## 1.3 RDF/NIF

## 1.4 Domain specific NER

## 1.5 DBpedia ontology

The DBpedia Ontology is a shallow, cross-domain ontology, which has been manually created based on the most commonly used infoboxes within Wikipedia. The ontology currently covers 685 classes which form a subsumption hierarchy and are described by 2,795 different properties.

With the DBpedia 3.2 release, we introduced a new infobox extraction method based on hand-generated mappings of Wikipedia infoboxes to the DBpedia ontology. The mappings defined fine-granular rules on how to parse infobox values. The mappings also addressed weaknesses in the Wikipedia infobox system, like having different infoboxes for the same class, using different property names for the same property, and not having clearly defined datatypes for property values. Therefore, the instance data within the infobox ontology is much cleaner and better structured than the infobox data within the DBpedia infobox dataset which was generated using the old infobox extraction code.

With the DBpedia 3.5 release, we introduced a public wiki for writing infobox mappings, editing existing ones as well as editing the DBpedia ontology. This allows external contributors to define mappings for the infoboxes they are interested in and to extend the existing DBpedia ontology with additional classes and properties.

Since the DBpedia 3.7 release, the ontology is a directed-acyclic graph, not a tree. Classes may have multiple superclasses, which was important for the mappings to schema.org. A taxonomy can still be constructed by ignoring all superclasses except the one that is specified first in the list and is considered the most important.

## 1.6 Evaluation/training datasets



---

## Domain specific NER

There are parameters of the computer used for tests shown in Table ??.

### 2.1 Data pre-processing

How data are processed? How we choose those domains (Politics, Sport, Transportation)? Coarse grained Fine grained

### 2.2 Types retrieval

At the main file where we write text we retrieve all types that are available for that text. After that, that text is processed with coarse and fine grained types. In coarse grained we have politics, sport and transportation specific domains. In fine grained in politics domain we retrieve in total 26 types (write them all), which we sort in 11 more specific types like for example Ambassador, Mayor, Deputy etc. are fused in one specific domain Politician. We do the same for sport domain where we retrieve in total 171 types, so those types, same as politics domain, are specified in 8 types, like SportClub, SportsLeague, SportsTeam, Athlete, Coach, OrganizationMember, SportsManager, SportsEvent. We also do the same for transportation domain, where we retrieve in total 78 types and minimized in 14 more specific types like Aircraft, Automobile, Locomotive, MilitaryVehicle, Motorcycle, On-SiteTransportation, Rocket, Ship, SpaceShuttle, SpaceStation, Spacecraft, Train, PublicTransitSystem and Infrastructure.

### 2.3 Domain specification

### 2.4 Data transformation/output for Stanford

How we process data to be ready to use on Stanford NER system?



---

# Experiments

We have provide various types of experiments. In next sections we will discuss more about every provided experiment.

## 3.1 Goals of the experiments

We set a few goals of the experiments. First of all we waned to test does we will get better results if we run the model of all domains in coarse grained, against the model of all domains in fine grained. In this test we run the models also with all domains texts. Then we get those models and we run it with specific domain texts, in both fine and coarse grained. Also we make experiments with specific domain model runned with domain specific texts, for example, politics domain model in coarse grained is runned with politics domain text also annotated in coarse grained, politics domain model in fine grained is runned with politics domain text also annotated in fine grained, and the same for sport and transportation domains.

## 3.2 List of experiments

Make it like experiment - discussion

### 3.2.1 Top 10 Links

In this experiment we get the model that is trained with all 3 domains together with coarse grained and we make an experiment with the text that is trained, that means text from all 3 domains annotated with coarse grained. Here we have for every domain 10 abstracts, so in total 30 annotated abstracts. This is our smallest model. How we can see from the picture in politics domain F1 value is lower than in sport and transportation. This results with lower F1 value in total, but still that is very satisfying value, because we have only 1 false positive and 2 false negative words.

### 3. EXPERIMENTS

---

Table 3.1: TABLE

Entity	P	R	F1
POLITICS	0,9655	0,9333	0,9492
SPORT	1,0000	1,0000	1,0000
TRANSPORTATION	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9811</b>	<b>0,9630</b>	<b>0,9720</b>

In this experiment we get the previous trained model and test with politics coarse grained text. Result is not satisfying, because we have a very low recall, so that results with bad F1 value and very big false negative words. We can clearly say that this is not the model that we want to use in annotating. With those results we can say that all 3 domains together with all text perform better results than all 3 domain model tested with specified domain.

Table 3.2: TABLE

Entity	P	R	F1
POLITICS	0,9655	0,3636	0,5283
<b>Totals</b>	<b>0,9655</b>	<b>0,3636</b>	<b>0,5283</b>

Then we get the same model, and like previous, we tested with sport coarse grained text. So how we can see from the picture, we have perfect result, there is no false positive or false negative words, there are only true positive which results with F1 value of 1.000 (no lost words).

Table 3.3: TABLE

Entity	P	R	F1
SPORT	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

Finally we tested this model with our last domain, transportation domain with coarse grained text. Because of very small annotated words we have again perfect results without any false positive or false negative words, same as previous test, there are only true positive matches, which is what we want.

Table 3.4: TABLE

Entity	P	R	F1
TRANSPORTATION	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

After we finish testing of all 3 domains trained with coarse grained texts, we move on with new type of experiment, model that contains all our domains and is trained with fine grained texts. We run the same experiment like in

coarse grained, but now text is annotated with fine grained. So, in such small trained model is not surprising that we get the same results as in experiment with coarse grained annotated text.

Table 3.5: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Aircraft	1,0000	1,0000	1,0000
Athlete	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
PoliticalParty	0,9600	0,9231	0,9412
Politician	1,0000	1,0000	1,0000
PublicTransitSystem	1,0000	1,0000	1,0000
Ship	1,0000	1,0000	1,0000
SportsClub	1,0000	1,0000	1,0000
SportsEvent	1,0000	1,0000	1,0000
SportsLeague	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9811</b>	<b>0,9630</b>	<b>0,9720</b>

We have repeated the same experiments from previous, but now with fine grained annotation. So we tested our model with politics domain annotated in fine grained and we get exactly the same results like in coarse grained. In this case there is no difference if we use coarse or fine grained trained model, because the results are same.

Table 3.6: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Election	0,0000	0,0000	0,0000
PoliticalParty	0,9600	0,9231	0,9412
Politician	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9655</b>	<b>0,3636</b>	<b>0,5283</b>

The sport domain text annotated in fine grained provide also same result as coarse grained annotated domain.

### 3. EXPERIMENTS

---

Table 3.7: TABLE

Entity	P	R	F1
Athlete	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
SportsClub	1,0000	1,0000	1,0000
SportsEvent	1,0000	1,0000	1,0000
SportsLeague	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

Same as sport domain, the transportation domain Figure ?? text annotated with fine grained, provide also same results.

Table 3.8: TABLE

Entity	P	R	F1
Aircraft	1,0000	1,0000	1,0000
PublicTransitSystem	1,0000	1,0000	1,0000
Ship	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

After experimenting with models that are trained with all domains together, we move on with experiments where we create specific models, both coarse and fine grained. First we tested the model that is created with politics text annotated in coarse grained. How we can see from Figure ?? we have a better results than that experiment that we provide in all 3 domains. The F1 value is very big which is the results of low false positive and false negative annotated words.

Table 3.9: TABLE

Entity	P	R	F1
POLITICS	0,9737	0,9610	0,9673
<b>Totals</b>	<b>0,9737</b>	<b>0,9610</b>	<b>0,9673</b>

We also provide the same experiment like previous, but now the model is annotated in fine grained. From Figure ?? is clearly that we have improve our model, here the F1 value is bigger that in coarse grained (see Figure ??) and also false positive annotated words are decreased by 1. That means even we have a complete review of our entities and can see the precision of each entity, this domain, in total, is more precise that coarse grained model and all 3 domains in coarse and fine grained.



Table 3.10: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Election	1,0000	0,9333	0,9655
PoliticalParty	0,9600	0,9231	0,9412
Politician	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9867</b>	<b>0,9610</b>	<b>0,9737</b>

From here, with this small amount of data we can say that it's better to use a politics specific domain, in fine grained, than a global domain, because in global domain we have 28 words annotated as true positive, but in a politics specific coarse grained domain we have 74 words annotated as true positive, which is very big difference, if we want to be precise.

We repeated the experiment for sport specific domain, but in this case the results are exactly the same in coarse and fine grained (see Figure ?? and Figure ?? as well as in all 3 domains also in coarse and fine grained

Table 3.11: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
SPORT	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

Table 3.12: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Athlete	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
SportsClub	1,0000	1,0000	1,0000
SportsEvent	1,0000	1,0000	1,0000
SportsLeague	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

This model is really small, only 15 annotated words, that's because we cannot make a conclusion which model is better to use. But if we speak about performance, of course the better solution is specified domain.

The transportation model, like a sport domain, is also a very small, only 9 annotated words. We provide the experiments for coarse grained (see Figure ??) and we get same results as in all 3 domains (see Figure ??). Same is happening for the fine grained domain (see Figure ??). In such a small trained domain, we cannot clearly say which model is better to use.

### 3. EXPERIMENTS

---

Table 3.13: TABLE

Entity	P	R	F1
TRANSPORTATION	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

Table 3.14: TABLE

Entity	P	R	F1
Aircraft	1,0000	1,0000	1,0000
PublicTransitSystem	1,0000	1,0000	1,0000
Ship	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

In conclusion of this subsection, we can say that our strategy for creating a specific domain pays off, because in those domains we have a better results than in global domains. But let see in subsection 3.2.2 how domains will behave with a bigger data(text).

#### 3.2.2 Top 20 Links

In subsection 3.2.1 we provide various experiments with different models. Now in this subsection we will repeat those experiments, but now we doubled the number of text to every domain, so we have 20 abstracts for every domain, in total 60 abstract. Those models, like previous, are trained in both coarse and fine grained. The first experiment in Figure ?? is now rerun with a bigger model and text. From Figure ?? we can see that there are more annotated words and surprisingly better results than in previous domain. Then we tested the model with politics domain text in coarse grained annotation. Again the results are quite better than previous experiment (see Figure ??), also now there is no false positive words, but this is still not satisfying. After that we tested this model with sport domain text in coarse grained annotation (see Figure ??). And like in previous experiment (see Figure ??) we have a perfect results, so we can say that this kind of model with very small domain annotated words is good to use. And finally we tested the model with transportation domain text in coarse grained annotation (see Figure ??). Figure ?? show us that the model annotated some false positive word with SPORT, which even we have a perfect F1 value for transportation domain, in overall calculation, because of that word we have a lower F1 value than experiment in Figure ??.

Table 3.15: TABLE

Entity	P	R	F1
POLITICS	1,0000	0,9615	0,9804
SPORT	1,0000	1,0000	1,0000
TRANSPORTATION	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>0,9780</b>	<b>0,9889</b>

Table 3.16: TABLE

Entity	P	R	F1
POLITICS	1,0000	0,3906	0,5618
<b>Totals</b>	<b>1,0000</b>	<b>0,3906</b>	<b>0,5618</b>

Table 3.17: TABLE

Entity	P	R	F1
SPORT	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

Table 3.18: TABLE

Entity	P	R	F1
SPORT	0,0000	1,0000	0,0000
TRANSPORTATION	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9231</b>	<b>1,0000</b>	<b>0,9600</b>

Similarly like experiment with a coarse grained model, we repeated the same experiments, but now, of coarse, with fine grained model. We test the trained model with text from all 3 domains together in fine grained annotation, and how we can see from Figure ?? the results are quite worst that in experiment with coarse grained (see Figure ??), but better that the experiment where we have only 10 abstract (see Figure ??). Furthermore we test the model with politics text, and how we can see from Figure ?? again results are not even close to previous experiment with all 3 domains together where we have better results like overall also in politics domain. Then we move on with a sport text (see Figure ??) and again we have perfect results like in coarse grained (see Figure ??) and the experiment with 10 abstract (see Figure ??). And final experiment for this model is transportation text. Figure ?? shows us that results are worst that in coarse grained (see Figure ??). Now we do not have any false positive word annotated with SPORT domain, but we have a lower number of annotated words which results with a lower F1 value in total.

### 3. EXPERIMENTS

---

Table 3.19: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Aircraft	1,0000	0,5000	0,6667
Athlete	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
Infrastructure	1,0000	0,5000	0,6667
PoliticalParty	1,0000	0,9512	0,9750
Politician	1,0000	1,0000	1,0000
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	1,0000	1,0000	1,0000
SportsEvent	1,0000	1,0000	1,0000
SportsLeague	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>0,9560</b>	<b>0,9775</b>

Table 3.20: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Election	0,0000	0,0000	0,0000
PoliticalParty	1,0000	0,9512	0,9750
Politician	1,0000	0,2000	0,3333
<b>Totals</b>	<b>1,0000</b>	<b>0,3906</b>	<b>0,5618</b>

Table 3.21: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Athlete	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
SportsClub	1,0000	1,0000	1,0000
SportsEvent	1,0000	1,0000	1,0000
SportsLeague	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

Table 3.22: TABLE

Entity	P	R	F1
Aircraft	1,0000	0,5000	0,6667
Infrastructure	1,0000	0,5000	0,6667
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsTeam	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9091</b>	<b>0,8333</b>	<b>0,8696</b>

After we finish the experiments with models that are trained with texts from every domain, we do tests with models that are trained with texts from a particular domain (politics, sport and transportation). Figure ?? shows the output of politics coarse grain specified domain, where we can see that we have a way more better results than in experiment with a model trained with texts from all 3 domains (see Figure ?? and Figure ??). The politics specific domain finds 125 true positive words unlike the experiments with all 3 domains that gives us only 50 true positive annotated words. However we repeated this experiment, but now with model trained in fine grained. Figure ?? shows the result of experiment, where we can see that there is no false positive annotated word like in previous experiment (see Figure ??), which result with a better F1 value in total.

Table 3.23: TABLE

Entity	P	R	F1
POLITICS	0,9921	0,9766	0,9843
<b>Totals</b>	<b>0,9655</b>	<b>0,3636</b>	<b>0,5283</b>

Table 3.24: TABLE

Entity	P	R	F1
Election	1,0000	0,9688	0,9841
PoliticalParty	1,0000	0,9512	0,9750
Politician	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>0,9766</b>	<b>0,9881</b>

Sport specific domain do not disappointed us. Figure ?? show the output of experiment in coarse grained and Figure ?? show the output of experiment in fine grained. Both experiment same as experiments with all 3 domains together (see Figure ??, Figure ??, Figure ??, Figure ??) give us perfect results. But if you want to be more precise, we recommend to use the specific model, because is fastest of reading words.

### 3. EXPERIMENTS

---

Table 3.25: TABLE

Entity	P	R	F1
SPORT	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

Table 3.26: TABLE

Entity	P	R	F1
Athlete	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
SportsClub	1,0000	1,0000	1,0000
SportsEvent	1,0000	1,0000	1,0000
SportsLeague	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

At the end we make experiments with a transportation specific domain. In Figure ?? we can see that our model gives us a worst results that experiment in all 3 domains. For instance experiment in Figure ?? gives us a perfect annotation with 1.0000 value at F1, also experiment in Figure ?? gives the same result like previous experiment, but here because of false positive annotated word in SPORT domain in total we have a lower value in F1. In overall experiments in all 3 domains gives a better result that the experiment in specific domain. Of coarse we make an experiment with model that is trained with fine grained text. How Figure ?? shows that this model even gives worst result that the experiment with coarse grained model, and also worst results than the experiment in Figure ??.

Table 3.27: TABLE

Entity	P	R	F1
TRANSPORTATION	1,0000	0,8333	0,9091
<b>Totals</b>	<b>1,0000</b>	<b>0,8333</b>	<b>0,9091</b>

Table 3.28: TABLE

Entity	P	R	F1
Aircraft	1,0000	0,5000	0,6667
Infrastructure	1,0000	0,5000	0,6667
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>0,7500</b>	<b>0,8571</b>

### 3.2.3 Top 40 Links

Table 3.29: TABLE

Entity	P	R	F1
POLITICS	0,9890	0,9375	0,9626
SPORT	1,0000	1,0000	1,0000
TRANSPORTATION	1,0000	0,9846	0,9922
<b>Totals</b>	<b>0,9960</b>	<b>0,9724</b>	<b>0,9841</b>

Table 3.30: TABLE

Entity	P	R	F1
POLITICS	0,9890	0,3529	0,5202
<b>Totals</b>	<b>0,9890</b>	<b>0,3529</b>	<b>0,5202</b>

Table 3.31: TABLE

Entity	P	R	F1
SPORT	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

Table 3.32: TABLE

Entity	P	R	F1
SPORT	0,0000	1,0000	0,0000
TRANSPORTATION	1,0000	0,9846	0,9922
<b>Totals</b>	<b>0,9697</b>	<b>0,9846</b>	<b>0,9771</b>

### 3. EXPERIMENTS

---

Table 3.33: TABLE

Entity	P	R	F1
Aircraft	1,0000	1,0000	1,0000
Athlete	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
Infrastructure	1,0000	1,0000	1,0000
PoliticalParty	0,9863	0,9730	0,9796
Politician	1,0000	0,8182	0,9000
PublicTransitSystem	1,0000	1,0000	1,0000
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	1,0000	1,0000	1,0000
SportsEvent	1,0000	1,0000	1,0000
SportsLeague	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9960</b>	<b>0,9764</b>	<b>0,9861</b>

Table 3.34: TABLE

Entity	P	R	F1
Election	0,0000	0,0000	0,0000
PoliticalParty	0,9863	0,9730	0,9796
Politician	1,0000	0,1565	0,2707
<b>Totals</b>	<b>0,9890</b>	<b>0,3529</b>	<b>0,5202</b>

Table 3.35: TABLE

Entity	P	R	F1
Athlete	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
SportsClub	1,0000	1,0000	1,0000
SportsEvent	1,0000	1,0000	1,0000
SportsLeague	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>



Table 3.36: TABLE

Entity	P	R	F1
Aircraft	1,0000	1,0000	1,0000
Infrastructure	1,0000	1,0000	1,0000
PublicTransitSystem	1,0000	1,0000	1,0000
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsTeam	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9701</b>	<b>1,0000</b>	<b>0,9848</b>

Table 3.37: TABLE

Entity	P	R	F1
POLITICS	0,9921	0,9804	0,9862
<b>Totals</b>	<b>0,9921</b>	<b>0,9804</b>	<b>0,9862</b>

Table 3.38: TABLE

Entity	P	R	F1
Election	1,0000	0,9848	0,9924
PoliticalParty	0,9863	0,9730	0,9796
Politician	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9960</b>	<b>0,9882</b>	<b>0,9921</b>

Table 3.39: TABLE

Entity	P	R	F1
SPORT	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>1,0000</b>	<b>1,0000</b>

Table 3.40: TABLE

Entity	P	R	F1
Athlete	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
SportsClub	1,0000	0,9474	0,9730
SportsEvent	1,0000	1,0000	1,0000
SportsLeague	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>0,9890</b>	<b>0,9945</b>

### 3. EXPERIMENTS

---

Table 3.41: TABLE

Entity	P	R	F1
TRANSPORTATION	1,0000	0,9846	0,9922
<b>Totals</b>	<b>1,0000</b>	<b>0,9846</b>	<b>0,9922</b>

Table 3.42: TABLE

Entity	P	R	F1
Aircraft	1,0000	1,0000	1,0000
Infrastructure	1,0000	1,0000	1,0000
PublicTransitSystem	1,0000	0,9630	0,9811
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsTeam	0,0000	1,0000	0,0000
<b>Totals</b>	<b>1,0000</b>	<b>0,9846</b>	<b>0,9922</b>

#### 3.2.4 Top 100 Links

Table 3.43: TABLE

Entity	P	R	F1
POLITICS	0,9920	0,9612	0,9764
SPORT	0,9963	0,9926	0,9944
TRANSPORTATION	1,0000	0,9735	0,9865
<b>Totals</b>	<b>0,9952</b>	<b>0,9766</b>	<b>0,9856</b>

Table 3.44: TABLE

Entity	P	R	F1
POLITICS	0,9920	0,3615	0,5299
<b>Totals</b>	<b>0,9920</b>	<b>0,3615</b>	<b>0,5299</b>

Table 3.45: TABLE

Entity	P	R	F1
SPORT	0,9962	0,9888	0,9925
<b>Totals</b>	<b>0,9962</b>	<b>0,9888</b>	<b>0,9925</b>

Table 3.46: TABLE

Entity	P	R	F1
SPORT	0,0000	1,0000	0,0000
TRANSPORTATION	1,0000	0,9735	0,9865
<b>Totals</b>	<b>0,9821</b>	<b>0,9735</b>	<b>0,9778</b>

Table 3.47: TABLE

Entity	P	R	F1
Aircraft	1,0000	0,6957	0,8205
Athlete	1,0000	0,4167	0,5882
Automobile	1,0000	1,0000	1,0000
Coach	1,0000	0,6667	0,8000
Infrastructure	1,0000	1,0000	1,0000
PoliticalParty	0,8774	0,6700	0,7598
Politician	1,0000	0,7455	0,8542
PublicTransitSystem	0,9744	0,7308	0,8352
Ship	1,0000	0,6000	0,7500
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	0,9512	0,9398	0,9455
SportsEvent	0,9737	0,8605	0,9136
SportsLeague	0,9500	0,8636	0,9048
SportsManager	1,0000	1,0000	1,0000
SportsTeam	1,0000	0,6364	0,7778
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9452</b>	<b>0,7535</b>	<b>0,8385</b>

Table 3.48: TABLE

Entity	P	R	F1
Election	0,0000	0,0000	0,0000
PoliticalParty	0,8774	0,6700	0,7598
Politician	1,0000	0,1285	0,2278
SportsEvent	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,8985</b>	<b>0,2580</b>	<b>0,4009</b>

### 3. EXPERIMENTS

---

Table 3.49: TABLE

Entity	P	R	F1
Athlete	1,0000	0,4167	0,5882
Coach	1,0000	0,6667	0,8000
SportsClub	0,9506	0,9390	0,9448
SportsEvent	1,0000	0,8605	0,9250
SportsLeague	0,9500	0,8636	0,9048
SportsManager	1,0000	1,0000	1,0000
SportsTeam	1,0000	0,6250	0,7692
<b>Totals</b>	<b>0,9683</b>	<b>0,7985</b>	<b>0,8753</b>

Table 3.50: TABLE

Entity	P	R	F1
Aircraft	1,0000	0,6957	0,8205
Automobile	1,0000	1,0000	1,0000
Infrastructure	1,0000	1,0000	1,0000
PublicTransitSystem	0,9744	0,7308	0,8352
Ship	1,0000	0,6000	0,7500
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	0,0000	1,0000	0,0000
SportsTeam	0,0000	1,0000	0,0000
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9677</b>	<b>0,7965</b>	<b>0,8738</b>

Table 3.51: TABLE

Entity	P	R	F1
POLITICS	0,9956	0,9898	0,9927
<b>Totals</b>	<b>0,9956</b>	<b>0,9898</b>	<b>0,9927</b>

Table 3.52: TABLE

Entity	P	R	F1
Election	1,0000	0,9878	0,9939
PoliticalParty	0,9950	0,9852	0,9901
Politician	0,9937	0,9906	0,9922
<b>Totals</b>	<b>0,9956</b>	<b>0,9883</b>	<b>0,9920</b>

Table 3.53: TABLE

Entity	P	R	F1
SPORT	0,9963	0,9963	0,9963
<b>Totals</b>	<b>0,9963</b>	<b>0,9963</b>	<b>0,9963</b>

Table 3.54: TABLE

Entity	P	R	F1
Athlete	1,0000	0,9722	0,9859
Coach	1,0000	1,0000	1,0000
SportsClub	1,0000	0,9878	0,9939
SportsEvent	1,0000	0,9767	0,9882
SportsLeague	1,0000	1,0000	1,0000
SportsManager	1,0000	1,0000	1,0000
SportsTeam	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>0,9888</b>	<b>0,9944</b>

Table 3.55: TABLE

Entity	P	R	F1
TRANSPORTATION	1,0000	0,9912	0,9956
<b>Totals</b>	<b>1,0000</b>	<b>0,9912</b>	<b>0,9956</b>

Table 3.56: TABLE

Entity	P	R	F1
Aircraft	1,0000	1,0000	1,0000
Automobile	1,0000	1,0000	1,0000
Infrastructure	1,0000	1,0000	1,0000
PublicTransitSystem	1,0000	0,9808	0,9903
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	0,6667	0,8000
SpaceStation	1,0000	1,0000	1,0000
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>1,0000</b>	<b>0,9735</b>	<b>0,9865</b>

### 3.2.5 Top 300 Links

**MAIN EXPERIMENT** This is our main experiment where other experiments will be compared with this one. This model is trained with top 300 Wikipedia abstracts based on PageRank. The text is annotated in coarse grained and has around 2300 annotated words in total. First experiment that we do with this model is that we run it with the same text that model is

### 3. EXPERIMENTS

---

created. Results are not bad at all, we are above 95%, which is great number for such middle weight model.

Table 3.57: All 3 domain model with top 300 abstracts, tested with all 3 domain texts

Entity	P	R	F1
POLITICS	0,9872	0,9462	0,9662
SPORT	0,9846	0,9629	0,9736
TRANSPORTATION	0,9940	0,9823	0,9881
<b>Totals</b>	<b>0,9875</b>	<b>0,9625</b>	<b>0,9748</b>

Table 3.58: TABLE

Entity	P	R	F1
POLITICS	0,9839	0,4025	0,5713
TRANSPORTATION	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9792</b>	<b>0,4025</b>	<b>0,5705</b>

Table 3.59: TABLE

Entity	P	R	F1
POLITICS	0,0000	1,0000	0,0000
SPORT	0,9846	0,9628	0,9736
TRANSPORTATION	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9819</b>	<b>0,9628</b>	<b>0,9722</b>

Table 3.60: TABLE

Entity	P	R	F1
SPORT	0,0000	1,0000	0,0000
TRANSPORTATION	0,9940	0,9822	0,9880
<b>Totals</b>	<b>0,9861</b>	<b>0,9822</b>	<b>0,9841</b>

Table 3.61: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Aircraft	1,0000	1,0000	1,0000
Athlete	1,0000	0,9802	0,9900
Automobile	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
Infrastructure	1,0000	0,9820	0,9909
PoliticalParty	0,9860	0,9628	0,9743
Politician	1,0000	0,9353	0,9665
PublicTransitSystem	0,9919	0,9839	0,9879
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	0,9796	0,9683	0,9739
SportsEvent	1,0000	0,9242	0,9606
SportsLeague	0,9647	0,9805	0,9725
SportsManager	1,0000	0,9423	0,9703
SportsTeam	1,0000	0,9805	0,9902
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9880</b>	<b>0,9712</b>	<b>0,9795</b>

Table 3.62: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Election	0,0000	0,0000	0,0000
PoliticalParty	0,9860	0,9628	0,9743
Politician	1,0000	0,1849	0,3120
PublicTransitSystem	0,0000	1,0000	0,0000
Ship	0,0000	1,0000	0,0000
SportsLeague	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9825</b>	<b>0,4072</b>	<b>0,5758</b>

### 3. EXPERIMENTS

---

Table 3.63: TABLE

Entity	P	R	F1
Athlete	1,0000	0,9802	0,9900
Coach	1,0000	1,0000	1,0000
Politician	0,0000	1,0000	0,0000
SportsClub	0,9794	0,9680	0,9737
SportsEvent	1,0000	0,9242	0,9606
SportsLeague	0,9678	0,9805	0,9741
SportsManager	1,0000	0,9423	0,9703
SportsTeam	1,0000	0,9804	0,9901
Train	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9821</b>	<b>0,9716</b>	<b>0,9768</b>

Table 3.64: TABLE

Entity	P	R	F1
Aircraft	1,0000	1,0000	1,0000
Automobile	1,0000	1,0000	1,0000
Infrastructure	1,0000	0,9820	0,9909
Politician	0,0000	1,0000	0,0000
PublicTransitSystem	0,9918	0,9837	0,9878
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	0,0000	1,0000	0,0000
SportsTeam	0,0000	1,0000	0,0000
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9862</b>	<b>0,9881</b>	<b>0,9871</b>

Table 3.65: TABLE

Entity	P	R	F1
POLITICS	0,8039	0,6779	0,7355
<b>Totals</b>	<b>0,8039</b>	<b>0,6779</b>	<b>0,7355</b>

Table 3.66: TABLE

Entity	P	R	F1
Election	0,8240	0,6398	0,7203
PoliticalParty	0,8100	0,7006	0,7513
Politician	0,8599	0,7234	0,7858
<b>Totals</b>	<b>0,8354</b>	<b>0,6980</b>	<b>0,7606</b>



Table 3.67: TABLE

Entity	P	R	F1
SPORT	0,9432	0,8839	0,9126
<b>Totals</b>	<b>0,9432</b>	<b>0,8839</b>	<b>0,9126</b>

Table 3.68: TABLE

Entity	P	R	F1
Athlete	0,9713	0,8366	0,8989
Coach	1,0000	0,7500	0,8571
SportsClub	0,9453	0,9041	0,9242
SportsEvent	1,0000	0,7879	0,8814
SportsLeague	0,9418	0,8958	0,9182
SportsManager	1,0000	0,9615	0,9804
SportsTeam	0,9845	0,8301	0,9007
<b>Totals</b>	<b>0,9592</b>	<b>0,8750</b>	<b>0,9152</b>

Table 3.69: TABLE

Entity	P	R	F1
TRANSPORTATION	0,9583	0,9109	0,9340
<b>Totals</b>	<b>0,9583</b>	<b>0,9109</b>	<b>0,9340</b>

Table 3.70: TABLE

Entity	P	R	F1
Aircraft	0,9659	0,8333	0,8947
Automobile	1,0000	0,8000	0,8889
Infrastructure	0,9550	0,9550	0,9550
PublicTransitSystem	0,9662	0,9309	0,9482
Ship	1,0000	0,6429	0,7826
SpaceShuttle	1,0000	0,8333	0,9091
SpaceStation	0,0000	1,0000	0,0000
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9660</b>	<b>0,9010</b>	<b>0,9324</b>

### 3.2.6 Top 400 Links

Table 3.71: TABLE

Entity	P	R	F1
POLITICS	0,9804	0,9434	0,9615
SPORT	0,9832	0,9590	0,9709
TRANSPORTATION	0,9941	0,9754	0,9847
<b>Totals</b>	<b>0,9849</b>	<b>0,9584</b>	<b>0,9714</b>

Table 3.72: TABLE

Entity	P	R	F1
POLITICS	0,9754	0,4082	0,5756
SPORT	0,0000	1,0000	0,0000
TRANSPORTATION	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9531</b>	<b>0,4082</b>	<b>0,5716</b>

Table 3.73: TABLE

Entity	P	R	F1
POLITICS	0,0000	1,0000	0,0000
SPORT	0,9837	0,9588	0,9711
TRANSPORTATION	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9805</b>	<b>0,9588</b>	<b>0,9695</b>

Table 3.74: TABLE

Entity	P	R	F1
POLITICS	0,0000	1,0000	0,0000
SPORT	0,0000	1,0000	0,0000
TRANSPORTATION	0,9939	0,9762	0,9806
<b>Totals</b>	<b>0,9861</b>	<b>0,9822</b>	<b>0,9841</b>

Table 3.75: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Aircraft	1,0000	1,0000	1,0000
Athlete	1,0000	0,9899	0,9949
Automobile	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
Infrastructure	1,0000	0,9896	0,9948
PoliticalParty	0,9766	0,9486	0,9624
Politician	1,0000	0,9893	0,9946
PublicTransitSystem	0,9935	0,9776	0,9855
Ship	1,0000	0,9231	0,9600
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	0,9796	0,9658	0,9726
SportsEvent	1,0000	0,8636	0,9268
SportsLeague	0,9698	0,9835	0,9766
SportsManager	1,0000	0,9726	0,9861
SportsTeam	1,0000	0,9851	0,9925
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9870</b>	<b>0,9709</b>	<b>0,9789</b>

Table 3.76: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Aircraft	0,0000	1,0000	0,0000
Election	0,0000	0,0000	0,0000
PoliticalParty	0,9766	0,9484	0,9623
Politician	1,0000	0,2092	0,3460
PublicTransitSystem	0,0000	1,0000	0,0000
Ship	0,0000	1,0000	0,0000
SportsClub	0,0000	1,0000	0,0000
SportsLeague	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9619</b>	<b>0,4151</b>	<b>0,5799</b>

### 3. EXPERIMENTS

---

Table 3.77: TABLE

Entity	P	R	F1
Aircraft	0,0000	1,0000	0,0000
Athlete	1,0000	0,9899	0,9949
Coach	1,0000	1,0000	1,0000
PoliticalParty	0,0000	1,0000	0,0000
Politician	0,0000	1,0000	0,0000
SportsClub	0,9794	0,9654	0,9724
SportsEvent	1,0000	0,8636	0,9268
SportsLeague	0,9696	0,9834	0,9765
SportsManager	1,0000	0,9726	0,9861
SportsTeam	1,0000	0,9850	0,9924
Train	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9821</b>	<b>0,9721</b>	<b>0,9770</b>

Table 3.78: TABLE

Entity	P	R	F1
Aircraft	1,0000	1,0000	1,0000
Automobile	1,0000	1,0000	1,0000
Infrastructure	1,0000	0,9896	0,9948
PoliticalParty	0,0000	1,0000	0,0000
Politician	0,0000	1,0000	0,0000
PublicTransitSystem	0,9934	0,9773	0,9853
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	0,0000	1,0000	0,0000
SportsTeam	0,0000	1,0000	0,0000
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9866</b>	<b>0,9866</b>	<b>0,9866</b>

Table 3.79: TABLE

Entity	P	R	F1
POLITICS	0,9866	0,9479	0,9669
<b>Totals</b>	<b>0,9866</b>	<b>0,9479</b>	<b>0,9669</b>

Table 3.80: TABLE

Entity	P	R	F1
Election	0,9975	0,9590	0,9779
PoliticalParty	0,9767	0,9530	0,9647
Politician	0,9977	0,9920	0,9948
<b>Totals</b>	<b>0,9906</b>	<b>0,9717</b>	<b>0,9810</b>

Table 3.81: TABLE

Entity	P	R	F1
SPORT	0,9858	0,9676	0,9766
<b>Totals</b>	<b>0,9858</b>	<b>0,8676</b>	<b>0,9766</b>

Table 3.82: TABLE

Entity	P	R	F1
Athlete	1,0000	0,9731	0,9863
Coach	1,0000	1,0000	1,0000
SportsClub	0,9815	0,9715	0,9765
SportsEvent	1,0000	0,9091	0,9524
SportsLeague	0,9718	0,9787	0,9752
SportsManager	1,0000	0,9726	0,9850
SportsTeam	1,0000	0,9900	0,9796
<b>Totals</b>	<b>0,9865</b>	<b>0,9727</b>	<b>0,9796</b>

Table 3.83: TABLE

Entity	P	R	F1
TRANSPORTATION	0,9954	0,9747	0,9850
<b>Totals</b>	<b>0,9954</b>	<b>0,9747</b>	<b>0,9850</b>

Table 3.84: TABLE

Entity	P	R	F1
Aircraft	1,0000	0,9835	0,9917
Automobile	1,0000	0,9583	0,9787
Infrastructure	1,0000	0,9948	0,9974
PublicTransitSystem	0,9934	0,9773	0,9853
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	0,6667	0,8000
SpaceStation	1,0000	1,0000	1,0000
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9970</b>	<b>0,9807</b>	<b>0,9888</b>

### 3.2.7 Top 500 Links

Table 3.85: TABLE

Entity	P	R	F1
POLITICS	0,9788	0,9444	0,9613
SPORT	0,9850	0,9596	0,9721
TRANSPORTATION	0,9962	0,9750	0,9855
<b>Totals</b>	<b>0,9857</b>	<b>0,9587</b>	<b>0,9720</b>

Table 3.86: TABLE

Entity	P	R	F1
POLITICS	0,9734	0,4095	0,5765
SPORT	0,0000	1,0000	0,0000
TRANSPORTATION	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9549</b>	<b>0,4095</b>	<b>0,5732</b>

Table 3.87: TABLE

Entity	P	R	F1
POLITICS	0,0000	1,0000	0,0000
SPORT	0,9849	0,9594	0,9720
TRANSPORTATION	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9788</b>	<b>0,9594</b>	<b>0,9690</b>

Table 3.88: TABLE

Entity	P	R	F1
POLITICS	0,0000	1,0000	0,0000
SPORT	0,0000	1,0000	0,0000
TRANSPORTATION	0,9961	0,9769	0,9864
<b>Totals</b>	<b>0,9870</b>	<b>0,9769</b>	<b>0,9819</b>

Table 3.89: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Aircraft	1,0000	0,9929	0,9964
Athlete	1,0000	0,9896	0,9948
Automobile	1,0000	1,0000	1,0000
Coach	1,0000	1,0000	1,0000
Infrastructure	1,0000	0,9783	0,9890
PoliticalParty	0,9775	0,9403	0,9585
Politician	1,0000	0,9874	0,9937
PublicTransitSystem	0,9944	0,9807	0,9875
Ship	1,0000	0,9259	0,9615
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	0,9756	0,9553	0,9654
SportsEvent	1,0000	0,8796	0,9360
SportsLeague	0,9700	0,9810	0,9755
SportsManager	1,0000	0,9780	0,9889
SportsTeam	1,0000	0,9831	0,9915
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9866</b>	<b>0,9669</b>	<b>0,9766</b>

Table 3.90: TABLE

<b>Entity</b>	<b>P</b>	<b>R</b>	<b>F1</b>
Aircraft	0,0000	1,0000	0,0000
Election	0,0000	0,0000	0,0000
PoliticalParty	0,9774	0,9400	0,9583
Politician	1,0000	0,2171	0,3567
PublicTransitSystem	0,0000	1,0000	0,0000
Ship	0,0000	1,0000	0,0000
SportsClub	0,0000	1,0000	0,0000
SportsLeague	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9631</b>	<b>0,4138</b>	<b>0,5789</b>

### 3. EXPERIMENTS

---

Table 3.91: TABLE

Entity	P	R	F1
Aircraft	0,0000	1,0000	0,0000
Athlete	1,0000	0,9896	0,9948
Coach	1,0000	1,0000	1,0000
PoliticalParty	0,0000	1,0000	0,0000
Politician	0,0000	1,0000	0,0000
SportsClub	0,9753	0,9549	0,9650
SportsEvent	1,0000	0,8796	0,9360
SportsLeague	0,9717	0,9810	0,9763
SportsManager	1,0000	0,9780	0,9889
SportsTeam	1,0000	0,9831	0,9915
Train	0,0000	1,0000	0,0000
<b>Totals</b>	<b>0,9785</b>	<b>0,9690</b>	<b>0,9737</b>

Table 3.92: TABLE

Entity	P	R	F1
Aircraft	1,0000	0,9927	0,9963
Automobile	1,0000	1,0000	1,0000
Infrastructure	1,0000	0,9783	0,9890
PoliticalParty	0,0000	1,0000	0,0000
Politician	0,0000	1,0000	0,0000
PublicTransitSystem	0,9943	0,9804	0,9873
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	1,0000	1,0000
SpaceStation	1,0000	1,0000	1,0000
SportsClub	0,0000	1,0000	0,0000
SportsTeam	0,0000	1,0000	0,0000
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9884</b>	<b>0,9833</b>	<b>0,9858</b>

Table 3.93: TABLE

Entity	P	R	F1
POLITICS	0,9808	0,9450	0,9626
<b>Totals</b>	<b>0,9808</b>	<b>0,9450</b>	<b>0,9626</b>



Table 3.94: TABLE

Entity	P	R	F1
Election	0,9915	0,9393	0,9647
PoliticalParty	0,9777	0,9502	0,9637
Politician	0,9962	0,9877	0,9919
<b>Totals</b>	<b>0,9890</b>	<b>0,9648</b>	<b>0,9768</b>

Table 3.95: TABLE

Entity	P	R	F1
SPORT	0,9856	0,9706	0,9780
<b>Totals</b>	<b>0,9856</b>	<b>0,9706</b>	<b>0,9780</b>

Table 3.96: TABLE

Entity	P	R	F1
Athlete	1,0000	0,9791	0,9894
Coach	1,0000	1,0000	1,0000
SportsClub	0,9771	0,9630	0,9700
SportsEvent	1,0000	0,9074	0,9515
SportsLeague	0,9755	0,9848	0,9801
SportsManager	1,0000	0,9780	0,9889
SportsTeam	1,0000	0,9915	0,9957
<b>Totals</b>	<b>0,9861</b>	<b>0,9731</b>	<b>0,9796</b>

Table 3.97: TABLE

Entity	P	R	F1
TRANSPORTATION	0,9974	0,9756	0,9864
<b>Totals</b>	<b>0,9974</b>	<b>0,9756</b>	<b>0,9864</b>

Table 3.98: TABLE

Entity	P	R	F1
Aircraft	1,0000	0,9781	0,9889
Automobile	1,0000	0,8800	0,9362
Infrastructure	1,0000	0,9870	0,9934
PublicTransitSystem	0,9915	0,9804	0,9860
Ship	1,0000	1,0000	1,0000
SpaceShuttle	1,0000	0,6667	0,8000
SpaceStation	1,0000	1,0000	1,0000
Train	1,0000	1,0000	1,0000
<b>Totals</b>	<b>0,9961</b>	<b>0,9769</b>	<b>0,9864</b>

### 3. EXPERIMENTS

```
CRFClassifier tagged 760067 words in 2 documents at 3060,01 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,9735  0,6934  0,8099  147     4      65
Athlete  0,9101  0,6222  0,7391  415     41     252
Automobile 1,0000  0,4098  0,5814  25      0      36
Coach    1,0000  0,3000  0,4615  6       0      14
Infrastructure 0,8885  0,5218  0,6575  263     33     241
Locomotive 0,0000  0,0000  0,0000  0       0      2
Motorcycle 0,0000  0,0000  0,0000  0       0      1
OrganisationMember 0,0000  0,0000  0,0000  0       0      1
PoliticalParty 0,8393  0,7403  0,7867  935     179    328
Politician 0,9271  0,6593  0,7706  267     21     138
PublicTransitSystem 0,9027  0,7389  0,8126  464     50     164
Rocket    0,0000  0,0000  0,0000  0       0      5
Ship      0,9615  0,5682  0,7143  25      1      19
SpaceShuttle 1,0000  0,4375  0,6087  7       0      9
SpaceStation 1,0000  0,6667  0,8000  2       0      1
SportsClub 0,8722  0,6071  0,7159  771     113    499
SportsEvent 0,9000  0,4829  0,6286  99      11     106
SportsLeague 0,8622  0,7357  0,7939  682     109    245
SportsManager 0,9787  0,4946  0,6571  92      2      94
SportsTeam 0,9276  0,7514  0,8302  269     21     89
Train     1,0000  0,5455  0,7059  6       0      5
Totals    0,8844  0,6592  0,7553  4475    585    2314

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFiles domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv,domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecifiedWithLowerPageRank.tsv
```

Figure 3.2: All 3 Domains Fine Grained Top 500 Links With All 3 Domains Fine Grained Top 500 Links And All 3 Domains Fine Grained Top 500 Links With Lower Page Rank

#### 3.2.8 MIXED

```
CRFClassifier tagged 760067 words in 2 documents at 3091,92 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,9242  0,5755  0,7093  122     10     90
Athlete  0,8182  0,3778  0,5160  252     56     415
Automobile 0,9565  0,3607  0,5238  22      1      39
Coach    1,0000  0,2000  0,3333  4       0      16
Infrastructure 0,8457  0,2037  0,4359  148     27     356
Locomotive 0,0000  0,0000  0,0000  0       0      2
Motorcycle 0,0000  0,0000  0,0000  0       0      1
OrganisationMember 0,0000  0,0000  0,0000  0       0      1
PoliticalParty 0,7656  0,5819  0,6613  735     225    528
Politician 0,8925  0,4099  0,5618  166     20     239
PublicTransitSystem 0,8291  0,6178  0,7080  388     80     240
Rocket    0,0000  0,0000  0,0000  0       0      5
Ship      0,9375  0,3409  0,5000  15      1      29
SpaceShuttle 1,0000  0,3750  0,5455  6       0      10
SpaceStation 1,0000  0,3333  0,5000  1       0      2
SportsClub 0,8009  0,4276  0,5575  543     135    727
SportsEvent 0,9559  0,3171  0,4762  65      3      140
SportsLeague 0,8071  0,5912  0,6824  548     131    379
SportsManager 0,9643  0,2903  0,4463  54      2      132
SportsTeam 0,8856  0,5838  0,7037  209     27     149
Train     1,0000  0,5455  0,7059  6       0      5
Totals    0,8206  0,4837  0,6087  3284    718    3505

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFiles domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv,domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecifiedWithLowerPageRank.tsv
```

Figure 3.1: All 3 Domains Fine Grained Top 300 With All 3 Domains Fine Grained Top 500 Links And All 3 Domains Fine Grained Top 500 Links With Lower PageRank

### 3.2. List of experiments

```
CRFClassifier tagged 336613 words in 1 documents at 2685,15 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,6667 0,1111 0,1905 8       4       64
Athlete 0,4675 0,1268 0,1994 36      41      248
Automobile 0,0000 0,0000 0,0000 0       0       36
Coach 0,0000 0,0000 0,0000 0       0       14
Infrastructure 0,5352 0,1387 0,2203 38      33      236
Locomotive 0,0000 0,0000 0,0000 0       0       2
Motorcycle 0,0000 0,0000 0,0000 0       0       1
OrganisationMember 0,0000 0,0000 0,0000 0       0       1
PoliticalParty 0,5462 0,4097 0,4682 195     162     281
Politician 0,5962 0,1867 0,2844 31      21      135
PublicTransitSystem 0,6943 0,4098 0,5154 109     48      157
Rocket 0,0000 0,0000 0,0000 0       0       5
Ship 0,0000 0,0000 0,0000 0       1       17
SpaceShuttle 0,0000 0,0000 0,0000 0       0       9
SpaceStation 0,0000 0,0000 0,0000 0       0       1
SportsClub 0,6370 0,2675 0,3768 172     98      471
SportsEvent 0,2667 0,0412 0,0714 4       11      93
SportsLeague 0,6395 0,4125 0,5015 165     93      235
SportsManager 0,6000 0,0316 0,0600 3       2       92
SportsTeam 0,6316 0,2975 0,4045 36      21      85
Train 0,0000 0,0000 0,0000 0       0       5
Totals 0,5983 0,2670 0,3692 797     535     2188

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecifiedWithLowerPageRank.tsv
```

Figure 3.3: All 3 Domains Fine Grained Top 500 Links With All 3 Domains Fine Grained Top 500 Links With Lower PageRank

```
CRFClassifier tagged 344007 words in 2 documents at 4801,01 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,9950 0,9706 0,9826 198     1       6
Athlete 0,0000 0,0000 0,0000 0       0       202
Automobile 1,0000 0,8500 0,9180 34      0       6
Coach 0,0000 0,0000 0,0000 0       0       4
Infrastructure 1,0000 0,9820 0,9900 218     0       4
PoliticalParty 0,0000 0,0000 0,0000 0       0       511
Politician 0,0000 0,0000 0,0000 0       0       139
PublicTransitSystem 0,9835 0,9676 0,9755 478     8       16
Ship 1,0000 0,9655 0,9825 28      0       1
SpaceShuttle 1,0000 0,6667 0,8000 8       0       4
SpaceStation 1,0000 1,0000 1,0000 2       0       0
SportsClub 0,0000 0,0000 0,0000 0       0       347
SportsEvent 0,0000 0,0000 0,0000 0       0       66
SportsLeague 0,0000 0,0000 0,0000 0       0       307
SportsManager 0,0000 0,0000 0,0000 0       0       52
SportsTeam 0,0000 0,0000 0,0000 0       0       154
Train 1,0000 0,9091 0,9524 10      0       1
Totals 0,9909 0,9491 0,9513 976     9       1820

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-TransportationTop500LinksFineGrained.ser.gz -testFiles domainsWithOneAnnotation\Top300Links\All3DomainsTop300LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv,domainsWithOneAnnotation\Top300Links\TransportationFineGrainedTop300Links.tsv
```

Figure 3.4: Transportation Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links And Transportation Fine Grained Top 300 Links

### 3. EXPERIMENTS

#### 3.2.9 With lower links on model

```
CRFClassifier tagged 97089 words in 1 documents at 2905,12 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 0,0435 0,0833 1      0      22
Athlete 1,0000 0,0278 0,0541 1      0      35
Automobile 0,0000 0,0000 0,0000 0      0      3
Coach 1,0000 0,3333 0,5000 1      0      2
Infrastructure 0,0000 0,0000 0,0000 0      0      22
PoliticalParty 0,6458 0,3054 0,4147 62     34     141
Politician 1,0000 0,0727 0,1356 4      0      51
PublicTransitSystem 0,8750 0,1346 0,2333 7      1      45
Ship 1,0000 0,2000 0,3333 1      0      4
SpaceShuttle 0,0000 0,0000 0,0000 0      0      6
SpaceStation 0,0000 0,0000 0,0000 0      0      1
SportsClub 0,9000 0,1084 0,1935 9      1      74
SportsEvent 1,0000 0,0233 0,0455 1      0      42
SportsLeague 0,6667 0,2121 0,3218 14     7      52
SportsManager 0,0000 0,0000 0,0000 0      0      6
SportsTeam 1,0000 0,0303 0,0588 1      0      32
Train 0,0000 0,0000 0,0000 0      0      1
Totals 0,7034 0,1591 0,2595 102     43     539

C:\Dev\stanford-ner-2017-06-09>java -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top10Links\ner-All3DomainsTop10LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop100LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.5: All 3 Domains Fine Grained Top 10 Links Runned With All 3 Domains Fine Grained Top 100 Links

```
CRFClassifier tagged 270361 words in 1 documents at 3542,70 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 0,0098 0,0194 1      0      101
Athlete 1,0000 0,0050 0,0099 1      0      201
Automobile 0,0000 0,0000 0,0000 0      0      20
Coach 1,0000 0,2500 0,4000 1      0      3
Infrastructure 0,0000 0,0000 0,0000 0      0      111
PoliticalParty 0,5517 0,2192 0,3137 112     91     399
Politician 0,8000 0,0288 0,0556 4      1      135
PublicTransitSystem 0,6364 0,0282 0,0541 7      4      241
Ship 1,0000 0,0667 0,1250 1      0      14
SpaceShuttle 0,0000 0,0000 0,0000 0      0      6
SpaceStation 0,0000 0,0000 0,0000 0      0      1
SportsClub 0,8750 0,0403 0,0771 14     2      333
SportsEvent 1,0000 0,0152 0,0299 1      0      65
SportsLeague 0,6102 0,1173 0,1967 36     23     271
SportsManager 0,0000 0,0000 0,0000 0      0      52
SportsTeam 1,0000 0,0065 0,0129 1      0      153
Train 0,0000 0,0000 0,0000 0      0      6
Totals 0,5967 0,0781 0,1382 179     121     2112

C:\Dev\stanford-ner-2017-06-09>java -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top10Links\ner-All3DomainsTop10LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top300Links\All3DomainsTop300LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.6: All 3 Domains Fine Grained Top 10 Links Runned With All 3 Domains Fine Grained Top 300 Links

### 3.2. List of experiments

```
CRFClassifier tagged 423454 words in 1 documents at 3197,54 words per second.
Entity P R F1 TP FP FN
Aircraft 1,0000 0,0071 0,0142 1 0 139
Athlete 1,0000 0,0026 0,0052 1 0 382
Automobile 0,0000 0,0000 0,0000 0 0 25
Coach 1,0000 0,1667 0,2857 1 0 5
Infrastructure 0,0000 0,0000 0,0000 0 0 230
PoliticalParty 0,5385 0,1957 0,2870 154 132 633
Politician 0,8000 0,0167 0,0328 4 1 235
PublicTransitSystem 0,4118 0,0193 0,0369 7 10 355
Ship 1,0000 0,0370 0,0714 1 0 26
SpaceShuttle 0,0000 0,0000 0,0000 0 0 7
SpaceStation 0,0000 0,0000 0,0000 0 0 2
SportsClub 0,8148 0,0351 0,0673 22 5 605
SportsEvent 1,0000 0,0093 0,0183 1 0 107
SportsLeague 0,5556 0,0854 0,1480 45 36 482
SportsManager 0,0000 0,0000 0,0000 0 0 91
SportsTeam 1,0000 0,0295 0,0574 7 0 230
Train 0,0000 0,0000 0,0000 0 0 6
Totals 0,5701 0,0641 0,1153 244 184 3560

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top10Links\ner-All3DomainsTop10LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.7: All 3 Domains Fine Grained Top 10 Links Runned With All 3 Domains Fine Grained Top 500 Links

```
CRFClassifier tagged 336613 words in 1 documents at 2416,32 words per second.
Entity P R F1 TP FP FN
Aircraft 0,0000 0,0000 0,0000 0 0 72
Athlete 0,0000 0,0000 0,0000 0 0 284
Automobile 0,0000 0,0000 0,0000 0 0 36
Coach 0,0000 0,0000 0,0000 0 0 14
Infrastructure 0,0000 0,0000 0,0000 0 0 274
Locomotive 0,0000 0,0000 0,0000 0 0 2
Motorcycle 0,0000 0,0000 0,0000 0 0 1
OrganisationMember 0,0000 0,0000 0,0000 0 0 1
PoliticalParty 0,3916 0,1176 0,1809 56 87 420
Politician 0,0000 0,0000 0,0000 0 0 166
PublicTransitSystem 0,0000 0,0000 0,0000 0 10 266
Rocket 0,0000 0,0000 0,0000 0 0 5
Ship 0,0000 0,0000 0,0000 0 0 17
SpaceShuttle 0,0000 0,0000 0,0000 0 0 9
SpaceStation 0,0000 0,0000 0,0000 0 0 1
SportsClub 0,5455 0,0187 0,0361 12 10 631
SportsEvent 0,0000 0,0000 0,0000 0 0 97
SportsLeague 0,4348 0,0500 0,0897 20 26 380
SportsManager 0,0000 0,0000 0,0000 0 0 95
SportsTeam 1,0000 0,0331 0,0640 4 0 117
Train 0,0000 0,0000 0,0000 0 0 5
Totals 0,4089 0,0308 0,0573 92 133 2893

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top10Links\ner-All3DomainsTop10LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecifiedWithLowerPageRank.tsv
```

Figure 3.8: All 3 Domains Fine Grained Top 10 Links Runned With All 3 Domains Fine Grained Top 500 Links With Lower PageRank

### 3. EXPERIMENTS

```
CRFClassifier tagged 423454 words in 1 documents at 2649,15 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,9677 0,2143 0,3509 30      1      110
Athlete 1,0000 0,0392 0,0754 15      0      368
Automobile 1,0000 0,1600 0,2759 4       0      21
Coach 1,0000 0,3333 0,5000 2       0      4
Infrastructure 0,8065 0,1087 0,1916 25      6      205
PoliticalParty 0,7288 0,4473 0,5543 352     131    435
Politician 0,8851 0,3222 0,4724 77      10     162
PublicTransitSystem 0,8217 0,3564 0,4971 129     28     233
Ship 1,0000 0,1111 0,2000 3       0      24
SpaceShuttle 1,0000 0,8571 0,9231 6       0      1
SpaceStation 1,0000 0,5000 0,6667 1       0      1
SportsClub 0,8193 0,3254 0,4658 204     45     423
SportsEvent 0,9048 0,3519 0,5067 38      4      70
SportsLeague 0,8392 0,4953 0,6229 261     50     266
SportsManager 1,0000 0,0769 0,1429 7       0      84
SportsTeam 1,0000 0,1814 0,3071 43      0      194
Train 1,0000 0,5000 0,6667 3       0      3
Totals 0,8136 0,3155 0,4546 1200    275    2604

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top100Links\ner-All3DomainsTop100LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.9: All 3 Domains Fine Grained Top 100 Links Runned With All 3 Domains Fine Grained Top 500 Links

```
CRFClassifier tagged 336613 words in 1 documents at 3519,92 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,0000 0,0000 0,0000 0       1      72
Athlete 0,0000 0,0000 0,0000 0       0      284
Automobile 0,0000 0,0000 0,0000 0       0      36
Coach 0,0000 0,0000 0,0000 0       0      14
Infrastructure 0,0000 0,0000 0,0000 0       7      274
Locomotive 0,0000 0,0000 0,0000 0       0      2
Motorcycle 0,0000 0,0000 0,0000 0       0      1
OrganisationMember 0,0000 0,0000 0,0000 0       0      1
PoliticalParty 0,5022 0,2374 0,3224 113     112    363
Politician 0,5517 0,0964 0,1641 16      13     150
PublicTransitSystem 0,7255 0,1391 0,2334 37      14     229
Rocket 0,0000 0,0000 0,0000 0       0      5
Ship 0,0000 0,0000 0,0000 0       0      17
SpaceShuttle 0,0000 0,0000 0,0000 0       0      9
SpaceStation 0,0000 0,0000 0,0000 0       0      1
SportsClub 0,6093 0,1431 0,2317 92      59     551
SportsEvent 0,1667 0,0103 0,0194 1       5      96
SportsLeague 0,6092 0,2650 0,3693 106     68     294
SportsManager 0,0000 0,0000 0,0000 0       1      95
SportsTeam 0,8571 0,0496 0,0937 6       1      115
Train 0,0000 0,0000 0,0000 0       0      5
Totals 0,5690 0,1243 0,2040 371     281    2614

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top100Links\ner-All3DomainsTop100LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecifiedWithLowerPageRank.tsv
```

Figure 3.10: All 3 Domains Fine Grained Top 100 Links Runned With All 3 Domains Fine Grained Top 500 Links With Lower PageRank

### 3.2. List of experiments

```
CRFClassifier tagged 164060 words in 1 documents at 3254,58 words per second.
Entity P      R      F1      TP      FP      FN
Election 0,0000 0,0000 0,0000 0      0      494
PoliticalParty 0,7282 0,4483 0,5549 351    131    432
Politician 0,9398 0,0739 0,1371 78     5      977
SportsEvent 0,0000 1,0000 0,0000 0      1      0
SportsLeague 0,0000 1,0000 0,0000 0      5      0
Totals 0,7513 0,1840 0,2956 429    142    1903

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top100Links\ner-All3DomainsTop100LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\PoliticsFineGrainedTop500Links.tsv
```

Figure 3.11: All 3 Domains Fine Grained Top 100 Links Runned With Politics Fine Grained Top 500 Links

```
CRFClassifier tagged 117053 words in 1 documents at 3034,66 words per second.
Entity P      R      F1      TP      FP      FN
Election 0,0000 0,0000 0,0000 0      0      197
Infrastructure 0,0000 1,0000 0,0000 0      2      0
PoliticalParty 0,5045 0,2379 0,3233 113    111    362
Politician 0,6667 0,0245 0,0473 16     8      636
PublicTransitSystem 0,0000 1,0000 0,0000 0      3      0
SportsEvent 0,0000 1,0000 0,0000 0      2      0
SportsLeague 0,0000 1,0000 0,0000 0      3      0
Totals 0,5000 0,0974 0,1631 129    129    1195

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top100Links\ner-All3DomainsTop100LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\linksWithLowerPageRank\PoliticsFineGrainedTop500Links\linksWithLowerPageRank.tsv
```

Figure 3.12: All 3 Domains Fine Grained Top 100 Links Runned With Politics Fine Grained Top 500 Links With Lower PageRank

```
CRFClassifier tagged 142322 words in 1 documents at 3035,30 words per second.
Entity P      R      F1      TP      FP      FN
Athlete 1,0000 0,0392 0,0754 15     0      368
Coach 1,0000 0,3333 0,5000 2      0      4
Infrastructure 0,0000 1,0000 0,0000 0      1      0
PoliticalParty 0,0000 1,0000 0,0000 0      1      0
Politician 0,0000 1,0000 0,0000 0      1      0
SportsClub 0,8185 0,3269 0,4672 203    45     418
SportsEvent 0,9268 0,3519 0,5101 38     3      70
SportsLeague 0,8497 0,4952 0,6258 260    46     265
SportsManager 1,0000 0,0769 0,1429 7      0      84
SportsTeam 1,0000 0,1780 0,3022 42     0      194
Totals 0,8539 0,2878 0,4305 567    97     1403

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top100Links\ner-All3DomainsTop100LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\SportFineGrainedTop500Links.tsv
```

Figure 3.13: All 3 Domains Fine Grained Top 100 Links Runned With Sport Fine Grained Top 500 Links

### 3. EXPERIMENTS

```
CRFClassifier tagged 118469 words in 1 documents at 3761,28 words per second.
Entity P      R      F1      TP      FP      FN
Athlete 0,0000 0,0000 0,0000 0      0      271
Coach 0,0000 0,0000 0,0000 0      0      14
Infrastructure 0,0000 1,0000 0,0000 0      1      0
OrganisationMember 0,0000 0,0000 0,0000 0      0      1
Politician 0,0000 1,0000 0,0000 0      4      0
SportsClub 0,6216 0,1467 0,2374 92     56     535
SportsEvent 0,2500 0,0104 0,0200 1      3      95
SportsLeague 0,6082 0,2620 0,3662 104    67     293
SportsManager 0,0000 0,0000 0,0000 0      1      89
SportsTeam 0,8571 0,0504 0,0952 6      1      113
Totals 0,6042 0,1258 0,2082 203    133    1411

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top100Links\ner-All3DomainsTop100LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\inksWithLowerPageRank\SportFineGrainedTop500Links\inksWithLowerPageRank.tsv
```

Figure 3.14: All 3 Domains Fine Grained Top 100 Links Runned With Sport Fine Grained Top 500 Links With Lower PageRank

```
CRFClassifier tagged 117072 words in 1 documents at 3791,68 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,9677 0,2190 0,3571 30     1      107
Automobile 1,0000 0,1600 0,2759 4      0      21
Infrastructure 0,8333 0,1087 0,1923 25     5      205
Politician 0,0000 1,0000 0,0000 0      3      0
PublicTransitSystem 0,8217 0,3603 0,5010 129    28     229
Ship 1,0000 0,1875 0,3158 3      0      13
SpaceShuttle 1,0000 1,0000 1,0000 6      0      0
SpaceStation 1,0000 0,5000 0,6667 1      0      1
SportsClub 0,0000 1,0000 0,0000 0      1      0
SportsTeam 0,0000 1,0000 0,0000 0      1      0
Train 1,0000 0,6000 0,7500 3      0      2
Totals 0,8375 0,2580 0,3945 201    39     578

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top100Links\ner-All3DomainsTop100LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\TransportationFineGrainedTop500Links\inksWithLowerPageRank\SportFineGrainedTop500Links\inksWithLowerPageRank.tsv
```

Figure 3.15: All 3 Domains Fine Grained Top 100 Links Runned With Transportation Fine Grained Top 500 Links

```
CRFClassifier tagged 101091 words in 1 documents at 3831,82 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,0000 0,0000 0,0000 0      1      72
Automobile 0,0000 0,0000 0,0000 0      0      36
Infrastructure 0,0000 0,0000 0,0000 0      4      272
Locomotive 0,0000 0,0000 0,0000 0      0      2
Motorcycle 0,0000 0,0000 0,0000 0      0      1
PoliticalParty 0,0000 1,0000 0,0000 0      1      0
Politician 0,0000 1,0000 0,0000 0      1      0
PublicTransitSystem 0,7083 0,1399 0,2337 34     14     209
Rocket 0,0000 0,0000 0,0000 0      0      5
Ship 0,0000 0,0000 0,0000 0      0      12
SpaceShuttle 0,0000 0,0000 0,0000 0      0      9
SpaceStation 0,0000 0,0000 0,0000 0      0      1
SportsClub 0,0000 1,0000 0,0000 0      3      0
Train 0,0000 0,0000 0,0000 0      0      5
Totals 0,5862 0,0517 0,0950 34     24     624

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top100Links\ner-All3DomainsTop100LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\inksWithLowerPageRank\TransportationFineGrainedTop500Links\inksWithLowerPageRank.tsv
```

Figure 3.16: All 3 Domains Fine Grained Top 100 Links Runned With Transportation Fine Grained Top 500 Links With Lower PageRank



### 3.2. List of experiments

```
CRFClassifier tagged 423454 words in 1 documents at 2349,18 words per second.
Entity P R F1 TP FP FN
Aircraft 0,9573 0,8000 0,8716 112 5 28
Athlete 0,8889 0,5849 0,7055 224 28 159
Automobile 0,9565 0,8800 0,9167 22 1 3
Coach 1,0000 0,6667 0,8000 4 0 2
Infrastructure 0,9034 0,5696 0,6987 131 14 99
PoliticalParty 0,8854 0,7268 0,7983 572 74 215
Politician 0,9664 0,6025 0,7423 144 5 95
PublicTransitSystem 0,8966 0,7901 0,8399 286 33 76
Ship 1,0000 0,5556 0,7143 15 0 12
SpaceShuttle 1,0000 0,8571 0,9231 6 0 1
SpaceStation 1,0000 0,5000 0,6667 1 0 1
SportsClub 0,9018 0,6443 0,7516 404 44 223
SportsEvent 0,9688 0,5741 0,7209 62 2 46
SportsLeague 0,9079 0,7856 0,8423 414 42 113
SportsManager 0,9811 0,5714 0,7222 52 1 39
SportsTeam 0,9784 0,7637 0,8578 181 4 56
Train 1,0000 1,0000 1,0000 6 0 0
Totals 0,9124 0,6930 0,7877 2636 253 1168

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.17: All 3 Domains Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 500 Links

```
CRFClassifier tagged 336613 words in 1 documents at 3545,42 words per second.
Entity P R F1 TP FP FN
Aircraft 0,6667 0,1389 0,2299 10 5 62
Athlete 0,5000 0,0986 0,1647 28 28 256
Automobile 0,0000 0,0000 0,0000 0 0 36
Coach 0,0000 0,0000 0,0000 0 0 14
Infrastructure 0,5667 0,0620 0,1118 17 13 257
Locomotive 0,0000 0,0000 0,0000 0 0 2
Motorcycle 0,0000 0,0000 0,0000 0 0 1
OrganisationMember 0,0000 0,0000 0,0000 0 0 1
PoliticalParty 0,5191 0,3424 0,4127 163 151 313
Politician 0,5946 0,1325 0,2167 22 15 144
PublicTransitSystem 0,6846 0,3835 0,4916 102 47 164
Rocket 0,0000 0,0000 0,0000 0 0 5
Ship 0,0000 0,0000 0,0000 0 1 17
SpaceShuttle 0,0000 0,0000 0,0000 0 0 9
SpaceStation 0,0000 0,0000 0,0000 0 0 1
SportsClub 0,6043 0,2162 0,3184 139 91 504
SportsEvent 0,7500 0,0309 0,0594 3 1 94
SportsLeague 0,6009 0,3350 0,4302 134 89 266
SportsManager 0,6667 0,0211 0,0408 2 1 93
SportsTeam 0,5490 0,2314 0,3256 28 23 93
Train 0,0000 0,0000 0,0000 0 0 5
Totals 0,5822 0,2171 0,3163 648 465 2337

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\All3DomainsTop500LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecifiedWithLowerPageRank.tsv
```

Figure 3.18: All 3 Domains Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 500 Links With Lower PageRank

### 3. EXPERIMENTS

```
CRFClassifier tagged 164060 words in 1 documents at 4066,23 words per second.
Entity P      R      F1      TP      FP      FN
Athlete 0,0000 1,0000 0,0000 0       9       0
Election 0,0000 0,0000 0,0000 0       0       494
PoliticalParty 0,8853 0,7292 0,7997 571     74      212
Politician 0,9792 0,1336 0,2352 141     3       914
PublicTransitSystem 0,0000 1,0000 0,0000 0       2       0
Ship 0,0000 1,0000 0,0000 0       1       0
SportsClub 0,0000 1,0000 0,0000 0       1       0
SportsLeague 0,0000 1,0000 0,0000 0       2       0
SportsManager 0,0000 1,0000 0,0000 0       1       0
Totals 0,8845 0,3053 0,4539 712     93      1620

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\PoliticsFineGrainedTop500Links.tsv
```

Figure 3.19: All 3 Domains Fine Grained Top 300 Links Runned With Politics Fine Grained Top 500 Links

```
CRFClassifier tagged 117053 words in 1 documents at 4211,90 words per second.
Entity P      R      F1      TP      FP      FN
Athlete 0,0000 1,0000 0,0000 0       6       0
Election 0,0000 0,0000 0,0000 0       0       197
Infrastructure 0,0000 1,0000 0,0000 0       1       0
PoliticalParty 0,5200 0,3411 0,4122 162     149     313
Politician 0,5882 0,0307 0,0583 20      14      632
PublicTransitSystem 0,0000 1,0000 0,0000 0       9       0
SportsLeague 0,0000 1,0000 0,0000 0       2       0
SportsTeam 0,0000 1,0000 0,0000 0       1       0
Totals 0,5000 0,1375 0,2156 182     182     1142

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\LowerPageRank\PoliticsFineGrainedTop500Links\LowerPageRank.tsv
```

Figure 3.20: All 3 Domains Fine Grained Top 300 Links Runned With Politics Fine Grained Top 500 Links With Lower PageRank

```
CRFClassifier tagged 142322 words in 1 documents at 4466,12 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,0000 1,0000 0,0000 0       1       0
Athlete 0,9218 0,5849 0,7157 224     19      159
Automobile 0,0000 1,0000 0,0000 0       1       0
Coach 1,0000 0,6667 0,8000 4       0       2
PoliticalParty 0,0000 1,0000 0,0000 0       1       0
Politician 0,0000 1,0000 0,0000 0       3       0
SportsClub 0,0000 0,6441 0,7512 400     44      221
SportsEvent 0,9688 0,5741 0,7209 62      2       46
SportsLeague 0,9097 0,7867 0,8437 413     41      112
SportsManager 1,0000 0,5714 0,7273 52      0       39
SportsTeam 0,9783 0,7627 0,8571 180     4       56
Train 0,0000 1,0000 0,0000 0       1       0
Totals 0,9194 0,6777 0,7802 1335    117     635

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\SportFineGrainedTop500Links.tsv
```

Figure 3.21: All 3 Domains Fine Grained Top 300 Links Runned With Sport Fine Grained Top 500 Links

### 3.2. List of experiments

```
CRFClassifier tagged 118469 words in 1 documents at 4461,27 words per second.
Entity P R F1 TP FP FN
Athlete 0,5833 0,1033 0,1755 28 20 243
Coach 0,0000 0,0000 0,0000 0 0 14
OrganisationMember 0,0000 0,0000 0,0000 0 0 1
PoliticalParty 0,0000 1,0000 0,0000 0 3 0
Politician 0,0000 1,0000 0,0000 0 3 0
SportsClub 0,6096 0,2217 0,3251 139 89 488
SportsEvent 0,7500 0,0313 0,0600 3 1 93
SportsLeague 0,6000 0,3325 0,4279 132 88 265
SportsManager 0,6667 0,0225 0,0435 2 1 87
SportsTeam 0,5833 0,2353 0,3353 28 20 91
Totals 0,5961 0,2057 0,3058 332 225 1282

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\linksWithLowerPageRank\SportFineGrainedTop500Links\linksWithLowerPageRank.tsv
```

Figure 3.22: All 3 Domains Fine Grained Top 300 Links Runned With Sport Fine Grained Top 500 Links With Lower PageRank

```
CRFClassifier tagged 117072 words in 1 documents at 4190,42 words per second.
Entity P R F1 TP FP FN
Aircraft 0,9569 0,8102 0,8775 111 5 26
Automobile 1,0000 0,8800 0,9362 22 0 3
Infrastructure 0,9034 0,5696 0,6987 131 14 99
Politician 0,0000 1,0000 0,0000 0 2 0
PublicTransitSystem 0,8959 0,7933 0,8415 284 33 74
Ship 1,0000 0,8750 0,9333 14 0 2
SpaceShuttle 1,0000 1,0000 1,0000 6 0 0
SpaceStation 1,0000 0,5000 0,6667 1 0 1
SportsClub 0,0000 1,0000 0,0000 0 3 0
SportsTeam 0,0000 1,0000 0,0000 0 1 0
Train 1,0000 1,0000 1,0000 5 0 0
Totals 0,9082 0,7368 0,8136 574 58 205

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\TransportationFineGrainedTop500Links\linksWithLowerPageRank.tsv
```

Figure 3.23: All 3 Domains Fine Grained Top 300 Links Runned With Transportation Fine Grained Top 500 Links

```
CRFClassifier tagged 101091 words in 1 documents at 3283,24 words per second.
Entity P R F1 TP FP FN
Aircraft 0,6667 0,1389 0,2299 10 5 62
Athlete 0,0000 1,0000 0,0000 0 2 0
Automobile 0,0000 0,0000 0,0000 0 0 36
Infrastructure 0,5862 0,0625 0,1130 17 12 255
Locomotive 0,0000 0,0000 0,0000 0 0 2
Motorcycle 0,0000 0,0000 0,0000 0 0 1
PublicTransitSystem 0,6714 0,3868 0,4909 94 46 149
Rocket 0,0000 0,0000 0,0000 0 0 5
Ship 0,0000 0,0000 0,0000 0 1 12
SpaceShuttle 0,0000 0,0000 0,0000 0 0 9
SpaceStation 0,0000 0,0000 0,0000 0 0 1
SportsClub 0,0000 1,0000 0,0000 0 2 0
SportsLeague 0,0000 1,0000 0,0000 0 1 0
SportsTeam 0,0000 1,0000 0,0000 0 2 0
Train 0,0000 0,0000 0,0000 0 0 5
Totals 0,6302 0,1839 0,2847 121 71 537

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top500Links\linksWithLowerPageRank\TransportationFineGrainedTop500Links\linksWithLowerPageRank.tsv
```

Figure 3.24: All 3 Domains Fine Grained Top 300 Links Runned With Transportation Fine Grained Top 500 Links With Lower PageRank

### 3. EXPERIMENTS

```
CRFClassifier tagged 10336 words in 1 documents at 2532,71 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 1,0000 1,0000 1       0       0
Athlete 1,0000 1,0000 1,0000 1       0       0
Coach 1,0000 1,0000 1,0000 1       0       0
PoliticalParty 0,9545 0,8077 0,8750 21      1       5
Politician 1,0000 1,0000 1,0000 4       0       0
PublicTransitSystem 1,0000 0,8571 0,9231 6       0       1
Ship 1,0000 1,0000 1,0000 1       0       0
SportsClub 1,0000 1,0000 1,0000 7       0       0
SportsEvent 1,0000 1,0000 1,0000 1       0       0
SportsLeague 1,0000 0,7500 0,8571 3       0       1
SportsTeam 1,0000 1,0000 1,0000 1       0       0
Totals 0,9792 0,8704 0,9216 47      1       7

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top10Links\All3DomainsTop10LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.25: All 3 Domains Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 10 Links

#### 3.2.10 With higher links on model

```
CRFClassifier tagged 97089 words in 1 documents at 4145,56 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 1,0000 1,0000 23      0       0
Athlete 1,0000 1,0000 1,0000 36      0       0
Automobile 1,0000 1,0000 1,0000 3       0       0
Coach 1,0000 1,0000 1,0000 3       0       0
Infrastructure 1,0000 1,0000 1,0000 22      0       0
PoliticalParty 0,9848 0,9606 0,9726 195     3       8
Politician 1,0000 0,8909 0,9423 49      0       6
PublicTransitSystem 0,9800 0,9423 0,9608 49      1       3
Ship 1,0000 1,0000 1,0000 5       0       0
SpaceShuttle 1,0000 1,0000 1,0000 6       0       0
SpaceStation 1,0000 1,0000 1,0000 1       0       0
SportsClub 0,9639 0,9639 0,9639 80      3       3
SportsEvent 1,0000 0,9767 0,9882 42      0       1
SportsLeague 0,9403 0,9545 0,9474 63      4       3
SportsManager 1,0000 0,8333 0,9091 5       0       1
SportsTeam 1,0000 1,0000 1,0000 33      0       0
Train 1,0000 1,0000 1,0000 1       0       0
Totals 0,9825 0,9610 0,9716 616     11      25

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top10Links\All3DomainsTop10LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.26: All 3 Domains Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 100 Links

```
CRFClassifier tagged 4504 words in 1 documents at 2969,02 words per second.
Entity P      R      F1      TP      FP      FN
Election 0,0000 0,0000 0,0000 0       0       15
PoliticalParty 0,9545 0,8077 0,8750 21      1       5
Politician 1,0000 0,1111 0,2000 4       0       32
Totals 0,9615 0,3247 0,4854 25      1       52

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top10Links\PoliticsFineGrainedTop10Links.tsv
```

Figure 3.27: All 3 Domains Fine Grained Top 300 Links Runned With Politics Fine Grained Top 10 Links

### 3.2. List of experiments

```
CRFClassifier tagged 40673 words in 1 documents at 4019,47 words per second.
Entity P R F1 TP FP FN
Election 0,0000 0,0000 0,0000 0 0 164
PoliticalParty 0,9848 0,9606 0,9726 195 3 8
Politician 1,0000 0,1536 0,2663 49 0 270
Totals 0,9879 0,3557 0,5230 244 3 442

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\PoliticsFineGrainedTop100Links.tsv
```

Figure 3.28: All 3 Domains Fine Grained Top 300 Links Runned With Politics Fine Grained Top 100 Links

```
CRFClassifier tagged 2854 words in 1 documents at 2994,75 words per second.
Entity P R F1 TP FP FN
Athlete 1,0000 1,0000 1,0000 1 0 0
Coach 1,0000 1,0000 1,0000 1 0 0
SportsClub 1,0000 1,0000 1,0000 7 0 0
SportsEvent 1,0000 1,0000 1,0000 1 0 0
SportsLeague 1,0000 0,7500 0,8571 3 0 1
SportsTeam 1,0000 1,0000 1,0000 1 0 0
Totals 1,0000 0,9333 0,9655 14 0 1

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\SportFineGrainedTop100Links.tsv
```

Figure 3.29: All 3 Domains Fine Grained Top 300 Links Runned With Sport Fine Grained Top 10 Links

```
CRFClassifier tagged 31872 words in 1 documents at 3691,88 words per second.
Entity P R F1 TP FP FN
Athlete 1,0000 1,0000 1,0000 36 0 0
Coach 1,0000 1,0000 1,0000 3 0 0
SportsClub 0,9634 0,9634 0,9634 79 3 3
SportsEvent 1,0000 0,9767 0,9882 42 0 1
SportsLeague 0,9403 0,9545 0,9474 63 4 3
SportsManager 1,0000 0,8333 0,9091 5 0 1
SportsTeam 1,0000 1,0000 1,0000 32 0 0
Totals 0,9738 0,9701 0,9720 260 7 8

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\SportFineGrainedTop100Links.tsv
```

Figure 3.30: All 3 Domains Fine Grained Top 300 Links Runned With Sport Fine Grained Top 100 Links

```
CRFClassifier tagged 2978 words in 1 documents at 2863,46 words per second.
Entity P R F1 TP FP FN
Aircraft 1,0000 1,0000 1,0000 1 0 0
PublicTransitSystem 1,0000 0,8571 0,9231 6 0 1
Ship 1,0000 1,0000 1,0000 1 0 0
Totals 1,0000 0,8889 0,9412 8 0 1

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\TransportationFineGrainedTop100Links.tsv
```

Figure 3.31: All 3 Domains Fine Grained Top 300 Links Runned With Transportation Fine Grained Top 10 Links

### 3. EXPERIMENTS

```
CRFClassifier tagged 24544 words in 1 documents at 4180,55 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 1,0000 1,0000 23      0      0
Automobile 1,0000 1,0000 1,0000 3        0      0
Infrastructure 1,0000 1,0000 1,0000 22       0      0
PublicTransitSystem 0,9800 0,9423 0,9608 49       1      3
Ship 1,0000 1,0000 1,0000 5         0      0
SpaceShuttle 1,0000 1,0000 1,0000 6         0      0
SpaceStation 1,0000 1,0000 1,0000 1         0      0
SportsClub 0,0000 1,0000 0,0000 0         1      0
SportsTeam 0,0000 1,0000 0,0000 0         1      0
Train 1,0000 1,0000 1,0000 1         0      0
Totals 0,9735 0,9735 0,9735 110      3      3

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-All3DomainsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\TransportationFineGrainedTop100Links.tsv
```

Figure 3.32: All 3 Domains Fine Grained Top 300 Links Runned With Transportation Fine Grained Top 100 Links

```
CRFClassifier tagged 10336 words in 1 documents at 3459,17 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 1,0000 1,0000 1         0      0
Athlete 1,0000 1,0000 1,0000 1         0      0
Coach 1,0000 1,0000 1,0000 1         0      0
PoliticalParty 0,9545 0,8077 0,8750 21        1      5
Politician 1,0000 1,0000 1,0000 4         0      0
PublicTransitSystem 0,8333 0,7143 0,7692 5         1      2
Ship 1,0000 1,0000 1,0000 1         0      0
SportsClub 0,8750 1,0000 0,9333 7         1      0
SportsEvent 1,0000 1,0000 1,0000 1         0      0
SportsLeague 1,0000 1,0000 1,0000 4         0      0
SportsTeam 1,0000 1,0000 1,0000 1         0      0
Totals 0,9400 0,8704 0,9038 47        3      7

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop100LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.33: All 3 Domains Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 10 Links

```
CRFClassifier tagged 97089 words in 1 documents at 3675,94 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 1,0000 1,0000 23      0      0
Athlete 1,0000 1,0000 1,0000 36      0      0
Automobile 1,0000 1,0000 1,0000 3        0      0
Coach 1,0000 1,0000 1,0000 3         0      0
Infrastructure 1,0000 0,9545 0,9767 21       0      1
PoliticalParty 0,9799 0,9606 0,9701 195      4      8
Politician 1,0000 1,0000 1,0000 55       0      0
PublicTransitSystem 0,9796 0,9231 0,9505 48       1      4
Ship 1,0000 1,0000 1,0000 5         0      0
SpaceShuttle 1,0000 1,0000 1,0000 6         0      0
SpaceStation 1,0000 1,0000 1,0000 1         0      0
SportsClub 0,9529 0,9759 0,9643 81       4      2
SportsEvent 1,0000 0,8837 0,9383 38       0      5
SportsLeague 0,9275 0,9697 0,9481 64       5      2
SportsManager 1,0000 1,0000 1,0000 6         0      0
SportsTeam 1,0000 1,0000 1,0000 33       0      0
Train 1,0000 1,0000 1,0000 1         0      0
Totals 0,9779 0,9657 0,9717 619     14     22

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop100LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.34: All 3 Domains Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 100 Links

### 3.2. List of experiments

```
CRFClassifier tagged 270361 words in 1 documents at 2605,86 words per second.
Entity P R F1 TP FP FN
Aircraft 1,0000 1,0000 1,0000 102 0 0
Athlete 1,0000 0,9802 0,9900 198 0 4
Automobile 1,0000 1,0000 1,0000 20 0 0
Coach 1,0000 1,0000 1,0000 4 0 0
Infrastructure 1,0000 0,9640 0,9817 107 0 4
PoliticalParty 0,9718 0,9432 0,9573 482 14 29
Politician 1,0000 1,0000 1,0000 139 0 0
PublicTransitSystem 0,9918 0,9718 0,9817 241 2 7
Ship 1,0000 1,0000 1,0000 15 0 0
SpaceShuttle 1,0000 1,0000 1,0000 6 0 0
SpaceStation 1,0000 1,0000 1,0000 1 0 0
SportsClub 0,9709 0,9625 0,9667 334 10 13
SportsEvent 1,0000 0,8333 0,9091 55 0 11
SportsLeague 0,9551 0,9707 0,9628 298 14 9
SportsManager 1,0000 0,9615 0,9804 50 0 2
SportsTeam 1,0000 0,9805 0,9902 151 0 3
Train 1,0000 1,0000 1,0000 6 0 0
Totals 0,9822 0,9642 0,9731 2209 40 82

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top300Links\All3DomainsTop300LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.35: All 3 Domains Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links

```
CRFClassifier tagged 4504 words in 1 documents at 2429,34 words per second.
Entity P R F1 TP FP FN
Election 0,0000 0,0000 0,0000 0 0 15
PoliticalParty 0,9545 0,8077 0,8750 21 1 5
Politician 1,0000 0,1111 0,2000 4 0 32
Totals 0,9615 0,3247 0,4854 25 1 52

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\PoliticsFineGrainedTop100Links.tsv
```

Figure 3.36: All 3 Domains Fine Grained Top 500 Links Runned With Politics Fine Grained Top 10 Links

```
CRFClassifier tagged 40673 words in 1 documents at 4212,64 words per second.
Entity P R F1 TP FP FN
Election 0,0000 0,0000 0,0000 0 0 164
PoliticalParty 0,9799 0,9606 0,9701 195 4 8
Politician 1,0000 0,1724 0,2941 55 0 264
Totals 0,9843 0,3644 0,5319 250 4 436

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\PoliticsFineGrainedTop100Links.tsv
```

Figure 3.37: All 3 Domains Fine Grained Top 500 Links Runned With Politics Fine Grained Top 100 Links

### 3. EXPERIMENTS

```
CRFClassifier tagged 107491 words in 1 documents at 2506,26 words per second.
Entity P      R      F1      TP      FP      FN
Election 0,0000 0,0000 0,0000 0      0      322
PoliticalParty 0,9718 0,9432 0,9573 482    14      29
Politician 1,0000 0,1980 0,3305 136    0      551
PublicTransitSystem 0,0000 1,0000 0,0000 0      2      0
Ship 0,0000 1,0000 0,0000 0      1      0
SportsLeague 0,0000 1,0000 0,0000 0      1      0
Totals 0,9717 0,4066 0,5733 618    18      902

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top300Links\PoliticsFineGrainedTop300Links.tsv
```

Figure 3.38: All 3 Domains Fine Grained Top 500 Links Runned With Politics Fine Grained Top 300 Links

```
CRFClassifier tagged 2854 words in 1 documents at 2679,81 words per second.
Entity P      R      F1      TP      FP      FN
Athlete 1,0000 1,0000 1,0000 1      0      0
Coach 1,0000 1,0000 1,0000 1      0      0
SportsClub 0,8750 1,0000 0,9333 7      1      0
SportsEvent 1,0000 1,0000 1,0000 1      0      0
SportsLeague 1,0000 1,0000 1,0000 4      0      0
SportsTeam 1,0000 1,0000 1,0000 1      0      0
Totals 0,9375 1,0000 0,9677 15      1      0

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\SportFineGrainedTop100Links.tsv
```

Figure 3.39: All 3 Domains Fine Grained Top 500 Links Runned With Sport Fine Grained Top 10 Links

```
CRFClassifier tagged 31872 words in 1 documents at 4022,72 words per second.
Entity P      R      F1      TP      FP      FN
Athlete 1,0000 1,0000 1,0000 36      0      0
Coach 1,0000 1,0000 1,0000 3      0      0
SportsClub 0,9524 0,9756 0,9639 80      4      2
SportsEvent 1,0000 0,8837 0,9383 38      0      5
SportsLeague 0,9275 0,9697 0,9481 64      5      2
SportsManager 1,0000 1,0000 1,0000 6      0      0
SportsTeam 1,0000 1,0000 1,0000 32      0      0
Totals 0,9664 0,9664 0,9664 259      9      9

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\SportFineGrainedTop100Links.tsv
```

Figure 3.40: All 3 Domains Fine Grained Top 500 Links Runned With Sport Fine Grained Top 100 Links



### 3.2. List of experiments

```
CRFClassifier tagged 89224 words in 1 documents at 3135,95 words per second.
Entity P      R      F1      TP      FP      FN
Athlete 1,0000 0,9802 0,9900 198     0      4
Coach 1,0000 1,0000 1,0000 4        0      0
Politician 0,0000 1,0000 0,0000 0        2      0
SportsClub 0,9707 0,9622 0,9664 331     10     13
SportsEvent 1,0000 0,8333 0,9091 55      0      11
SportsLeague 0,9582 0,9707 0,9644 298     13     9
SportsManager 1,0000 0,9615 0,9804 50      0      2
SportsTeam 1,0000 0,9804 0,9901 150     0      3
Train 0,0000 1,0000 0,0000 0        1      0
Totals 0,9766 0,9628 0,9696 1086    26     42

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top300Links\SportFineGrainedTop300Links.tsv
```

Figure 3.41: All 3 Domains Fine Grained Top 500 Links Runned With Sport Fine Grained Top 300 Links

```
CRFClassifier tagged 2978 words in 1 documents at 2749,77 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 1,0000 1,0000 1        0      0
PublicTransitSystem 0,8333 0,7143 0,7692 5        1      2
Ship 1,0000 1,0000 1,0000 1        0      0
Totals 0,8750 0,7778 0,8235 7        1      2

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top10Links\TransportationFineGrainedTop10Links.tsv
```

Figure 3.42: All 3 Domains Fine Grained Top 500 Links Runned With Transportation Fine Grained Top 10 Links

```
CRFClassifier tagged 24544 words in 1 documents at 3821,86 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 1,0000 1,0000 23      0      0
Automobile 1,0000 1,0000 1,0000 3        0      0
Infrastructure 1,0000 0,9545 0,9767 21      0      1
PublicTransitSystem 0,9796 0,9231 0,9505 48      1      4
Ship 1,0000 1,0000 1,0000 5        0      0
SpaceShuttle 1,0000 1,0000 1,0000 6        0      0
SpaceStation 1,0000 1,0000 1,0000 1        0      0
SportsClub 0,0000 1,0000 0,0000 0        1      0
SportsTeam 0,0000 1,0000 0,0000 0        1      0
Train 1,0000 1,0000 1,0000 1        0      0
Totals 0,9730 0,9558 0,9643 108     3      5

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\TransportationFineGrainedTop100Links.tsv
```

Figure 3.43: All 3 Domains Fine Grained Top 500 Links Runned With Transportation Fine Grained Top 100 Links

### 3. EXPERIMENTS

```
CRFClassifier tagged 73646 words in 1 documents at 3209,68 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 1,0000 1,0000 102     0      0
Automobile 1,0000 1,0000 1,0000 20      0      0
Infrastructure 1,0000 0,9640 0,9817 107     0      4
Politician 0,0000 1,0000 0,0000 0        1      0
PublicTransitSystem 0,9917 0,9715 0,9815 239     2      7
Ship 1,0000 1,0000 1,0000 14      0      0
SpaceShuttle 1,0000 1,0000 1,0000 6        0      0
SpaceStation 1,0000 1,0000 1,0000 1        0      0
SportsClub 0,0000 1,0000 0,0000 0        3      0
SportsTeam 0,0000 1,0000 0,0000 0        1      0
Train 1,0000 1,0000 1,0000 5        0      0
Totals 0,9860 0,9782 0,9821 494     7      11

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-All3DomainsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top300Links\TransportationFineGrainedTop300Links.tsv
```

Figure 3.44: All 3 Domains Fine Grained Top 500 Links Runned With Transportation Fine Grained Top 300 Links

```
CRFClassifier tagged 97089 words in 1 documents at 7883,80 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,0000 0,0000 0,0000 0        0      23
Athlete 0,0000 0,0000 0,0000 0        0      36
Automobile 0,0000 0,0000 0,0000 0        0      3
Coach 0,0000 0,0000 0,0000 0        0      3
Election 0,0000 1,0000 0,0000 0        136     0
Infrastructure 0,0000 0,0000 0,0000 0        0      22
PoliticalParty 0,7831 0,7291 0,7551 148     41      55
Politician 0,1541 0,9273 0,2642 51      280     4
PublicTransitSystem 0,0000 0,0000 0,0000 0        0      52
Ship 0,0000 0,0000 0,0000 0        0      5
SpaceShuttle 0,0000 0,0000 0,0000 0        0      6
SpaceStation 0,0000 0,0000 0,0000 0        0      1
SportsClub 0,0000 0,0000 0,0000 0        0      83
SportsEvent 0,0000 0,0000 0,0000 0        0      43
SportsLeague 0,0000 0,0000 0,0000 0        0      66
SportsManager 0,0000 0,0000 0,0000 0        0      6
SportsTeam 0,0000 0,0000 0,0000 0        0      33
Train 0,0000 0,0000 0,0000 0        0      1
Totals 0,3034 0,3105 0,3069 199     457     442

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-PoliticsTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop100LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.45: Politics Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 100 Links

### 3.2. List of experiments

```
CRFClassifier tagged 97089 words in 1 documents at 5175,60 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,0000 0,0000 0,0000 0      0      23
Athlete 0,0000 0,0000 0,0000 0      0      36
Automobile 0,0000 0,0000 0,0000 0      0      3
Coach 0,0000 0,0000 0,0000 0      0      3
Election 0,0000 1,0000 0,0000 0      163      0
Infrastructure 0,0000 0,0000 0,0000 0      0      22
PoliticalParty 0,9652 0,9557 0,9604 194      7      9
Politician 0,1471 1,0000 0,2564 55      319      0
PublicTransitSystem 0,0000 0,0000 0,0000 0      0      52
Ship 0,0000 0,0000 0,0000 0      0      5
SpaceShuttle 0,0000 0,0000 0,0000 0      0      6
SpaceStation 0,0000 0,0000 0,0000 0      0      1
SportsClub 0,0000 0,0000 0,0000 0      0      83
SportsEvent 0,0000 0,0000 0,0000 0      0      43
SportsLeague 0,0000 0,0000 0,0000 0      0      66
SportsManager 0,0000 0,0000 0,0000 0      0      6
SportsTeam 0,0000 0,0000 0,0000 0      0      33
Train 0,0000 0,0000 0,0000 0      0      1
Totals 0,3374 0,3885 0,3611 249      489      392

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-PoliticsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop100LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.46: Politics Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 100 Links

```
CRFClassifier tagged 270361 words in 1 documents at 4572,62 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,0000 0,0000 0,0000 0      0      102
Athlete 0,0000 0,0000 0,0000 0      0      202
Automobile 0,0000 0,0000 0,0000 0      0      20
Coach 0,0000 0,0000 0,0000 0      0      4
Election 0,0000 1,0000 0,0000 0      337      0
Infrastructure 0,0000 0,0000 0,0000 0      0      111
PoliticalParty 0,9569 0,9550 0,9559 488      22      23
Politician 0,1518 0,9784 0,2628 136      760      3
PublicTransitSystem 0,0000 0,0000 0,0000 0      0      248
Ship 0,0000 0,0000 0,0000 0      0      15
SpaceShuttle 0,0000 0,0000 0,0000 0      0      6
SpaceStation 0,0000 0,0000 0,0000 0      0      1
SportsClub 0,0000 0,0000 0,0000 0      0      347
SportsEvent 0,0000 0,0000 0,0000 0      0      66
SportsLeague 0,0000 0,0000 0,0000 0      0      307
SportsManager 0,0000 0,0000 0,0000 0      0      52
SportsTeam 0,0000 0,0000 0,0000 0      0      154
Train 0,0000 0,0000 0,0000 0      0      6
Totals 0,3580 0,2724 0,3094 624      1119      1667

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-PoliticsTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top300Links\All3DomainsTop300LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.47: Politics Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links

### 3. EXPERIMENTS

```
CRFClassifier tagged 97089 words in 1 documents at 5454,44 words per second.
Entity P R F1 TP FP FN
Aircraft 0,0000 0,0000 0,0000 0 0 23
Athlete 0,2817 0,5556 0,3738 20 51 16
Automobile 0,0000 0,0000 0,0000 0 0 3
Coach 1,0000 0,6667 0,8000 2 0 1
Infrastructure 0,0000 0,0000 0,0000 0 0 22
PoliticalParty 0,0000 0,0000 0,0000 0 0 203
Politician 0,0000 0,0000 0,0000 0 0 55
PublicTransitSystem 0,0000 0,0000 0,0000 0 0 52
Ship 0,0000 0,0000 0,0000 0 0 5
SpaceShuttle 0,0000 0,0000 0,0000 0 0 6
SpaceStation 0,0000 0,0000 0,0000 0 0 1
SportsClub 0,8795 0,8795 0,8795 73 10 10
SportsEvent 0,9706 0,7674 0,8571 33 1 10
SportsLeague 0,8852 0,8182 0,8504 54 7 12
SportsManager 0,5000 1,0000 0,6667 6 6 0
SportsTeam 0,9583 0,6970 0,8070 23 1 10
Train 0,0000 0,0000 0,0000 0 0 1
Totals 0,7352 0,3292 0,4547 211 76 430

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-SportTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop100LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.48: Sport Fine Grained Top 300 Links Runned With All 3 Domains  
Fine Grained Top 100 Links

```
CRFClassifier tagged 97089 words in 1 documents at 5434,29 words per second.
Entity P R F1 TP FP FN
Aircraft 0,0000 0,0000 0,0000 0 0 23
Athlete 0,4359 0,9444 0,5965 34 44 2
Automobile 0,0000 0,0000 0,0000 0 0 3
Coach 1,0000 1,0000 1,0000 3 0 0
Infrastructure 0,0000 0,0000 0,0000 0 0 22
PoliticalParty 0,0000 0,0000 0,0000 0 0 203
Politician 0,0000 0,0000 0,0000 0 0 55
PublicTransitSystem 0,0000 0,0000 0,0000 0 0 52
Ship 0,0000 0,0000 0,0000 0 0 5
SpaceShuttle 0,0000 0,0000 0,0000 0 0 6
SpaceStation 0,0000 0,0000 0,0000 0 0 1
SportsClub 0,8889 0,9639 0,9249 80 10 3
SportsEvent 1,0000 0,8837 0,9383 38 0 5
SportsLeague 0,8889 0,9697 0,9275 64 8 2
SportsManager 0,4286 1,0000 0,6000 6 8 0
SportsTeam 0,9697 0,9697 0,9697 32 1 1
Train 0,0000 0,0000 0,0000 0 0 1
Totals 0,7835 0,4009 0,5304 257 71 384

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-SportTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop100LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.49: Sport Fine Grained Top 500 Links Runned With All 3 Domains  
Fine Grained Top 100 Links

### 3.2. List of experiments

```
CRFClassifier tagged 270361 words in 1 documents at 5259,02 words per second.
Entity P R F1 TP FP FN
Aircraft 0,0000 0,0000 0,0000 0 0 102
Athlete 0,5640 0,9604 0,7106 194 150 8
Automobile 0,0000 0,0000 0,0000 0 0 20
Coach 1,0000 1,0000 1,0000 4 0 0
Infrastructure 0,0000 0,0000 0,0000 0 0 111
PoliticalParty 0,0000 0,0000 0,0000 0 0 511
Politician 0,0000 0,0000 0,0000 0 0 139
PublicTransitSystem 0,0000 0,0000 0,0000 0 0 248
Ship 0,0000 0,0000 0,0000 0 0 15
SpaceShuttle 0,0000 0,0000 0,0000 0 0 6
SpaceStation 0,0000 0,0000 0,0000 0 0 1
SportsClub 0,9123 0,9597 0,9354 333 32 14
SportsEvent 1,0000 0,8636 0,9268 57 0 9
SportsLeague 0,9492 0,9739 0,9614 299 16 8
SportsManager 0,7937 0,9615 0,8696 50 13 2
SportsTeam 0,9684 0,9935 0,9808 153 5 1
Train 0,0000 0,0000 0,0000 0 0 6
Totals 0,8346 0,4758 0,6061 1090 216 1201

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-SportTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top300Links\All3DomainsTop300LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.50: Sport Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links

```
CRFClassifier tagged 97089 words in 1 documents at 4669,54 words per second.
Entity P R F1 TP FP FN
Aircraft 1,0000 0,6957 0,8205 16 0 7
Athlete 0,0000 0,0000 0,0000 0 0 36
Automobile 1,0000 1,0000 1,0000 3 0 0
Coach 0,0000 0,0000 0,0000 0 0 3
Infrastructure 0,8750 0,9545 0,9130 21 3 1
PoliticalParty 0,0000 0,0000 0,0000 0 0 203
Politician 0,0000 0,0000 0,0000 0 0 55
PublicTransitSystem 0,8696 0,7692 0,8163 40 6 12
Ship 1,0000 0,4000 0,5714 2 0 3
SpaceShuttle 1,0000 0,8333 0,9091 5 0 1
SpaceStation 0,0000 0,0000 0,0000 0 0 1
SportsClub 0,0000 0,0000 0,0000 0 0 83
SportsEvent 0,0000 0,0000 0,0000 0 0 43
SportsLeague 0,0000 0,0000 0,0000 0 0 66
SportsManager 0,0000 0,0000 0,0000 0 0 6
SportsTeam 0,0000 0,0000 0,0000 0 0 33
Train 1,0000 1,0000 1,0000 1 0 0
Totals 0,9072 0,1373 0,2385 88 9 553

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top300Links\ner-TransportationTop300LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop100LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.51: Transportation Fine Grained Top 300 Links Runned With All 3 Domains Fine Grained Top 100 Links

### 3. EXPERIMENTS

```
CRFClassifier tagged 97089 words in 1 documents at 5359,00 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 1,0000 0,9565 0,9778 22      0      1
Athlete 0,0000 0,0000 0,0000 0      0      36
Automobile 1,0000 1,0000 1,0000 3      0      0
Coach 0,0000 0,0000 0,0000 0      0      3
Infrastructure 1,0000 1,0000 1,0000 22      0      0
PoliticalParty 0,0000 0,0000 0,0000 0      0      203
Politician 0,0000 0,0000 0,0000 0      0      55
PublicTransitSystem 0,9412 0,9231 0,9320 48      3      4
Ship 1,0000 1,0000 1,0000 5      0      0
SpaceShuttle 1,0000 0,6667 0,8000 4      0      2
SpaceStation 1,0000 1,0000 1,0000 1      0      0
SportsClub 0,0000 0,0000 0,0000 0      0      83
SportsEvent 0,0000 0,0000 0,0000 0      0      43
SportsLeague 0,0000 0,0000 0,0000 0      0      66
SportsManager 0,0000 0,0000 0,0000 0      0      6
SportsTeam 0,0000 0,0000 0,0000 0      0      33
Train 1,0000 1,0000 1,0000 1      0      0
Totals 0,9725 0,1654 0,2827 106      3      535

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-TransportationTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop100LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.52: Transportation Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 100 Links

```
CRFClassifier tagged 270361 words in 1 documents at 3536,53 words per second.
Entity P      R      F1      TP      FP      FN
Aircraft 0,9900 0,9706 0,9802 99      1      3
Athlete 0,0000 0,0000 0,0000 0      0      202
Automobile 1,0000 0,8500 0,9189 17      0      3
Coach 0,0000 0,0000 0,0000 0      0      4
Infrastructure 1,0000 0,9820 0,9909 109      0      2
PoliticalParty 0,0000 0,0000 0,0000 0      0      511
Politician 0,0000 0,0000 0,0000 0      0      139
PublicTransitSystem 0,9795 0,9637 0,9715 239      5      9
Ship 1,0000 0,9333 0,9655 14      0      1
SpaceShuttle 1,0000 0,6667 0,8000 4      0      2
SpaceStation 1,0000 1,0000 1,0000 1      0      0
SportsClub 0,0000 0,0000 0,0000 0      0      347
SportsEvent 0,0000 0,0000 0,0000 0      0      66
SportsLeague 0,0000 0,0000 0,0000 0      0      307
SportsManager 0,0000 0,0000 0,0000 0      0      52
SportsTeam 0,0000 0,0000 0,0000 0      0      154
Train 1,0000 0,8333 0,9091 5      0      1
Totals 0,9879 0,2130 0,3504 488      6      1803

C:\Dev\stanford-ner-2017-06-09>java -Xmx11g -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier modelsWithOneAnnotation\Top500Links\ner-TransportationTop500LinksFineGrained.ser.gz -testFile domainsWithOneAnnotation\Top100Links\All3DomainsTop300LinksTextWithSameProcessedLinksOnAbstractFineGrainedSpecified.tsv
```

Figure 3.53: Transportation Fine Grained Top 500 Links Runned With All 3 Domains Fine Grained Top 300 Links

## 3.3 Summary of results

### 3.3.1 Graphs

---

## Conclusion





## Contents of CD