

Automatic Semantic Annotation of Sport News Using Knowledge Base and Extraction Patterns

Nguyen Quang Minh, Ngo Hong Son, Cao Tuan Dung

Hanoi University of Science and Technology, No. 1, Dai Co Viet, Hai Ba Trung, Hanoi, Viet Nam

Received: April 17, 2018; Accepted: June 29, 2018

Abstract

The World Wide Web is currently one of the most popular platforms for publishing, disseminating and consuming news. However, the huge number of daily published news items brings new challenges for both readers and publishers of web news systems in the process of finding or arranging information. Aiming to change the representation of data in a machine-readable semantic annotation, the semantic web technology promise to address these obstacles. Therefore, finding the solution for creating annotation with valuable semantics is a key point in the development of our news aggregation system. In this paper, we present a method for generating automatically semantic annotations of sport news items. It combines the results obtained through our continuous study of capturing different kinds of semantics which having from simple to more complex representation structure. Our approach relies on the detection of named entities as the ontology instances using knowledge base on sport. The instances are matched with pre-defined patterns to extract semantics. Experiments on corpus of sport news validates the advantages of the proposed method and shows that semantic annotations are generated with high precision and coverage.

Keywords: semantic annotation, semantic web, knowledge base, named entity recognition

1. Introduction

Thanks to the availability, accessibility, the Web is now one of the most popular platforms for publishing, disseminating and consuming news. It is a trend that news agencies and television channels use the Web as the main mean of distributing emerging news covering different domains as for example sport, business, entertainment, etc. Unfortunately, new issues appear not only to readers but also the news publisher. As most information on web page is designed only for human understanding by mixing content with presentation, the huge number of daily published news items makes the process of finding the ones relevant difficult to a reader. In addition, tasks of arranging, aggregating and linking news items become harder for the editor. Therefore, it is important to make machines aware of more functionality in a web news system, especially information retrieval and manipulation.

Among efforts to evolve the Web to its extreme potential, the Semantic Web aims to enable machine to support human better in data interpretation, aggregation and usage. The proposed approach is to change the representation of data in a machine-readable manner [2]. Certain research groups have worked on the development of semantic web portals

and content management system with better accuracy when processing information [5].

Inspired by the promising potential of Semantic Web technology, our previous work [9] focus on the construction of BKSport, a news aggregation system that help readers finding news items relevant to their needs using semantic search. The main idea is to make every news item "more intelligent" by associating them with metadata represented in a machine-understandable way. These metadata, as known as semantic annotations, are the basis for implementing semantic-based functionalities including news searching and recommending. As a result, the solution for creating annotation with valuable semantics is a key point in the development of our news aggregation system.

In this paper, we present a method for generating automatically semantic annotations of sport news items. It combines the results obtained through our continuous study of capturing different kinds of semantics which having from simple to more complex representation structure. Our approach relies on the detection of named entities as the ontology instances using knowledge base on sport. The instances are matched with pre-defined patterns to extract semantics from the text and formalize them using RDF language. The rest of the paper is structured as follows. Section 2 provides the background of semantic annotation. Subsequently, Section 3 elaborates on the proposed method and its

* Corresponding author: Tel: (+84) 926816659
Email: minh.nguyenquang@hust.edu.vn

implementation. The experimental results are presented in Section 4, followed by some related works. Last, Section 6 concludes the paper and discusses directions for future work.

2. Background of semantic annotation

Fernández [6] described the term "Semantic Annotation" as "the action and results of describing (part of) an electronic resource by means of metadata whose meaning is formally specified in an ontology". According to the classification given by [1], we can consider semantic annotation as a kind of formal metadata, which is machine understandable. Using ontology vocabulary, an annotation links an entity in the news item to its semantic description [11].

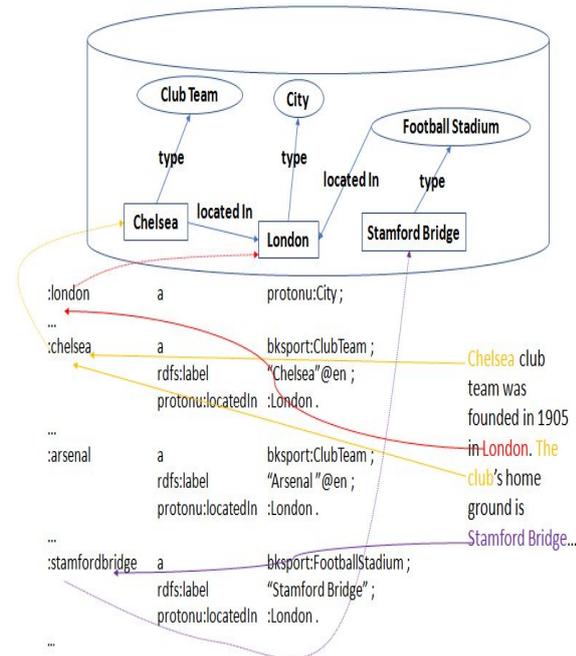


Fig. 1. Semantic annotation example

For example, a semantic annotation might involve "Chelsea" in a text to an ontology which both identifies it as the concept "Club Team" and associates it to the instance "London" of the concept "City", as illustrated in Figure 1. Thus, the meaning about "Chelsea" described in semantic annotation is unambiguous.

The most known benefits of semantic annotation are improved information retrieval and enhanced interoperability. The improvement of information retrieval comes from the ability to do a search with the inferences about data using ontology. News items are published from heterogeneous sources can be integrated in their annotations share a common ontology.

3. Automatic semantic annotation of sport news items

The improvement of information searching, classifying or filtering cannot be achieved without the availability of semantic annotations in BKSport. However, manual annotation of web news is a tedious and time-consuming task and is evidently impractical and unscalable for the high number of daily news in an aggregation system. In this section, we present progressive results on a method of automatic semantic annotation for sport news items. Our proposal relies on a mapping of named entities in the news with instances in the knowledge base on sport. These instances participate in the semantic extraction task to generate annotation in the form of triples. Figure 2 depicts the main step of the annotation generating process.

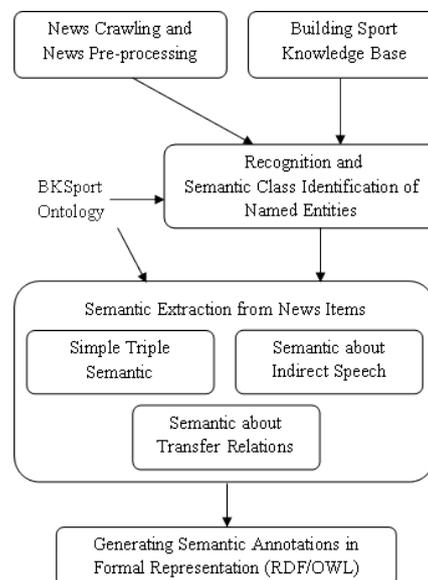


Fig. 2. Process of automatic sport news annotation

3.1 Ontology and knowledge base construction

The Semantic Web proposes annotating document content using concepts and properties from domain ontologies [2]. Thus, building a proper ontology which provides a domain specific vocabulary for semantic annotations is the first step of the annotation process. The development of BKSport ontology is guided by Gruber's principles to assure the clarity, consistency. The most important requirement is the sufficiency of vocabulary. We decided to reuse useful parts of BBC sport ontology and add new concepts and properties, focusing on the representation of important sport figures, events and activities. BKSport must also be compatible with the PROTON ontology in order to reuse the KIM [10] platform for the task of named entity detection.

Figure 3 depicts some concepts and properties of BKSport ontology

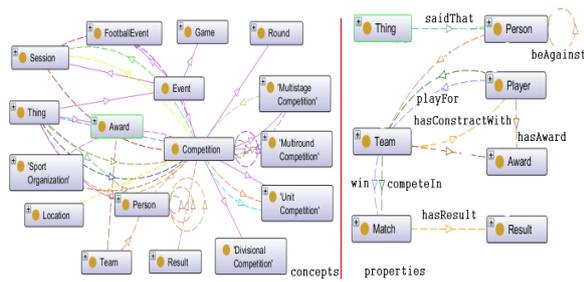


Fig. 3. A part of BKSport ontology

The performance of entity detection depends on the quality and the completeness of the knowledge base. It is expected to cover sufficiently the sport domain with knowledge about players, coach, clubs, awards, stadiums, etc. The construction process consists of the following steps:

- Collect and extract data from large and prestigious sources such as UEFA, ESPN, ATP World Tour using web crawlers and wrappers, then store them in XML format.
- Design mapping rules between relations in XML schema and properties in the ontology, formalize them using XSLT language.
- Transform data from XML to RDF.

3.2 Identifying named entity as a class instance in knowledge base

Appearing frequently in news articles as the name of players, coaches, managers, clubs, stadiums or sport events, etc., named entities are important for capturing certain semantics from news content. Named entity recognition (NER) involves identifying boundaries of named entities in text and classifying them into a predefined set such as people, organizations and locations. Using GATE, KIM [Popov] is a platform providing NER task for general domain. However, the objective of this step is to detect these entities and map them to the corresponding instances in the knowledge base on sport. For example, in the text "Liverpool has completed the 9 million signing of Egyptian striker Mohamed Salah", Liverpool should be identified as a football club and the BKSport needs to understand Mohamed Salah is a name of a football player.

We address this problem by extending KIM Proton ontology with the vocabulary and semantic data from our sport ontology and knowledge base. The mapping between these ontologies was carried out in the sense that more specialized concepts in the BKSport ontology will replace the abstract concept in Proton in recognition process using subclassOf

property. For example, classes of BKSport ontology such as Coach, Winger, Forward, Defender, are understood as sub-classes of the Person class. Figure 4 illustrates some classes mapped from BKSport ontology to PROTON ontology.

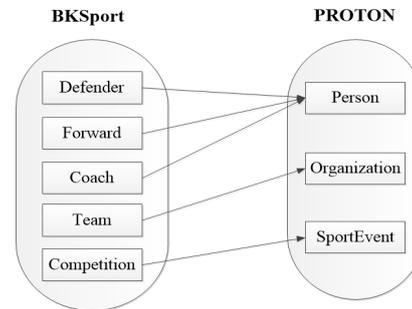


Fig. 4. Mapping from BKSport to PROTON

As depicted in Figure 5, Steven Caulker is not only understood as Person but also an instance of the class Defender.

oph Kramer has confirmed to Bild that Napoli have made an offer for him. The 23-year-old is ussia Monchengladbach but it appears a bid may well take him to Serie A. and their offer is interesting, especially because it comes from a great coach like (Rafa)

out for the signature of Cardiff's England centre-back Steven Caulker, but it appears Harry ace for the 22-year-old.

| Type: SportPerson Instance | |
|----------------------------|------------------------------------|
| Feature | Value |
| class | http://bk.sport.owl#Defender |
| originalName | Steven Caulker |
| inst | http://bk.sport.owl#steven-caulker |
| rule | SportPersonGen |

Close

Fig. 5. Named entity recognition using sport knowledge base

In addition, certain improvements have been realized to enhance the recognition effectiveness, as presented below.

Entity recognition by nickname. In many news items, the nickname of the sports figure appears quite popular. For example, readers often meet the words as Leo, El Pulga in articles about Messi or Fergie is widely understood as a nickname of Sir Alex Ferguson. By enriching the knowledge base with aliases and synonyms, our proposed method can identify entities relying on the appearance of their nickname.

Entity recognition at more detailed conceptual level. A named entity may be identified as an instance of a general ontology concept such as Person or Player instead of Forward or Player, if its description is missing from sport knowledge base or is not at a

detailed enough conceptual level. Noticing that some entities are represented as "occupation" followed by "private name" (e.g. Goalkeeper van der Sar, Striker Messi, etc.), while occupation may correspond to label of a concept, proper rules were built to recognize the correct type of them.

Recognition of shortened name entity. In sports news, sometimes a shortened name of an entity is used instead of the full names, especially when the full name was used previously. For example, "Boca Juniors striker Carlos Tevez has said Lionel Messi is "a natural at being the best in the world"... Tevez said: "Cristiano is totally different to Messi.". To recognize a shortened name, we compare it with the label of instances corresponding to full named entities identified before.

Disambiguation of entities having the same name but belonging to different types.

As the knowledge base is built by collecting data from various sources such as Premier League, La Liga, Champions League, ATP, there are instances belonging to different types but have the same name. For example, Giuseppe Meazza is the name of a player, but also the name of a stadium. We addressed this problem by matching the word standing right after an entity with the concepts in our ontology.

All instances and concepts recognized in previous step are stored in specific structure called *annotations*. They are evidently used in the semantic information extraction algorithms.

3.3 Extracting semantics from sport news

The heart of our method for generating semantic annotations is the semantic extraction step. In this study, we have no ambition to fully detect the meaning of the text. Instead, we focus on a number of important semantics that readers are most interested in sports news.

- Simple events in the form of triple
- Important entities and
- Indirect speech.
- Football transfer events.

Semantic about simple event or activity. On the very first period, we managed to recognize the popular information having a simple representation structure in sport news. They may involve the result of a sport event such as "Adebayor double help Spurs beat Swans", the interaction between sport persons, for example "Fergie defends Rooney Temperament" or the attitude of a player (or a coach or a referee) to a club or a league. To address this problem, we define

extraction patterns representing these semantics in natural language as follows:

- $\langle Person \rangle \langle relation \rangle \langle Person \rangle$, e.g. $\langle Paul Pogba \rangle \langle argue \rangle \langle Jose Mourinho \rangle$
- $\langle Organization \rangle \langle relation \rangle \langle Organization \rangle$, e.g. $\langle Manchester United \rangle \langle defeat \rangle \langle Chelsea \rangle$
- $\langle Person \rangle \langle relation \rangle \langle Organization \rangle$, e.g. $\langle Cristiano Ronaldo \rangle \langle transfer to \rangle \langle Real Madrid \rangle$

where $\langle Person \rangle$ stands for the occurrence of any instance of the Person concept or its subclass in ontology and it is similar for $\langle Organization \rangle$. From above pattern, we create extraction rules using JAPE grammar to match entities and token in the text with ontology vocabulary and extract successfully semantic triples from news items. For example, relation $\langle be\ against \rangle$ is represented by the rule follows:

```
Annotation.type=="SportPerson"{{Token.string=="is"}}{Token.string=="against"}{Annotation.type=="SportPerson"}
```

and the following rule detects result of a match, e.g. "Barcelona 3-2 Getafe":

```
"Annotation.type=="SportTeam"{{Annotation.type=="Number"}}{Token.string=="-"}{Annotation.type=="Number"}{Annotation.type=="SportTeam"}.
```

Each relationship can be represented by a set of tokens when appearing in the text; hence they are used in the extraction rules to enhance the detection ability.

Semantic about important entities. This task involves identifying the key entities that the news refers to, besides generating basic metadata such as titles. We define a weight for an instance to determine whether it is important in a news item or not. The calculation of this weight is not only based on the occurrence number of an instance, but also the position of the occurrence in the text as well as the relation between the type of instance with other concepts in the ontology. In addition, when an extraction rule is applied, the dependence weight between the class of the instance being matched with the rule itself would also be taken into account. The algorithm for extracting simple events and the important entities are presented as follows.

Algorithm for simple triples and important entities extraction

Input: wc_c - weight of concept c for the news content

$w_t c$ - weight of concept c for the news title

w_{dc} - distance weight of concept c with other concepts w_{rc} - weight of concept c with extraction rule r .

R - set of extraction rules, $W_{total} = 0$

Extract triple: $\langle \text{webpage.uri} \text{ bk:hasTitle} \text{ webpage.title} \rangle$

for each named entity i recognized as instance of concept c

m = number of occurrences of i in title.

$W_{\text{title-}i} = m * w_{tc}$

k = number of occurrences of i in content.

$W_{\text{content-}i} = k * (w_{cc} + w_{dc})$, $W_{\text{semantic-}i} = 0$

foreach sen in {news sentences} do

foreach rule r in R do

compare r with *annotations* in sen

if r matches instance i {

Extract triple corresponding r

$W_{\text{semantic-}i} = W_{\text{semantic-}i} + w_{rc}$

endfor

endfor

$W_i = W_{\text{title-}i} + W_{\text{content-}i} + W_{\text{semantic-}i}$

$W_{\text{total}} = W_{\text{total}} + W_i$

endfor

$\text{mean}W = W_{\text{total}} / \text{number of entities}$

for each named entity i recognized in news

if $W_i > \text{mean}W$

Extract triple $\langle \text{webpage.uri} \text{ bk:about} \text{ element.uri} \rangle$

else Extract triple

$\langle \text{webpage.uri} \text{ bk:contain} \text{ element.uri} \rangle$

endfor

Semantic about indirect speech. Indirect statements are frequently given in the sport article, for example "*And Chelsea beat Tottenham in a very important game*" Shevchenko told the Sky Sports News in Kiev".

To generate semantic annotation about this kind of information, a table defining keyword for an indirect statement such as "said that, told, statement, speech, announce, added, ..." is built. We then analyze the indirect clauses followed vocabularies defined using JAPE rules. The processing is conducted as follows:

// P is a set of reification pattern (e.g. A "said that" B)

$P = \{A \text{ "said That"/"announce" } B\}$;

foreach (Annotation p in P) do {

statement = $p.get("B")$;

annotationSet = BKSport.annotate(statement);

foreach(Annotation a in annotationSet)

if ($a.contains("semantic")$){

subject= annotation.get("subject");

predicate= annotation.get("predicate");

object= annotation.get("object");

Generate triples:

$\langle A \rangle \langle \text{bksport:saidThat} \rangle \langle \text{statement} \rangle$.

$\langle \text{statement} \rangle \langle \text{rdf:subject} \rangle$ subject.

$\langle \text{statement} \rangle \langle \text{rdf:predicate} \rangle$ predicate.

$\langle \text{statement} \rangle \langle \text{rdf:object} \rangle$ object.

endif

endfor

endfor

Semantic about football transfer events. Transfer information is one of the attractive news categories in many sport newspapers. Comparing with simple events, the semantics about a player moving from a soccer club to another or a contract signing have a more complex form of representation. The extraction patterns for simple sport events is extended to recognize these semantics, as depicted in Figure 6.



Luis Suarez transferred to Barcelona.



MILLWALL have completed the signing of Plymouth Argyle midfielder Nadjim Abdou.



Barcelona forward Lionel Messi signed a new contract.

Fig. 6. Extended extraction pattern for transfer relations

In this context, named entity represents often a football player or a soccer club and the *Phrasal verb* is modeled as follows:

$\langle \text{ExtraVerb} \rangle \langle \text{Main Verb} \rangle \langle \text{Adverb/Preposition} \rangle$, where "Extra Verb" includes tokens standing right

before the main verb. It helps determining whether the transfer took place, can happen in the near future or the transfer was unsuccessful. For example, in sentences such as "Former Aletico goalkeeper De Gea has signed a four-year deal at MU" or "Barcelona forward Messi will make a new contract.", recognize the extra verb as "signed" or "will" in the pattern lead to difference extracted temporal semantics. Thanks to JAPE grammar, a complex extraction rule can be represented as the combination of sub rules which are designed to identify elements described above.

Once a named entity is mentioned in a news item, it may be replaced by pronouns in subsequent sentences. Thus, detecting pronouns corresponding to named entities help enhancing the performance of entity recognition, then contribute to the transfer semantic extraction effectiveness. Our proposal is to construct pronoun recognition rules based on the following principles:

- Pronouns such as 'he', 'him', 'i', 'me' represent SportPerson while 'they', 'them', 'we', 'us' represent SportTeam.
- Pronouns such as 'i', 'me', 'we', 'us' appearing in indirect statements, represent agents (SportPerson or SportTeam) which make that statement. There are two forms of indirect statement:
 - o Agent standing in front of indirect statement.
 - o Agent standing behind indirect statement.
- The pronouns representing named entities (SportPerson or SportTeam) appear in front of or near such pronoun while in case of indirect statement, the pronoun may represent entities behind it.

Finally, to improve the recall score of semantic extraction for football transfer news, we pre-process sentences to transform the possessive case to the standard form, for example <Named Entity>'s signature is transformed to the signature of <Named Entity>.

4. Experimental Results

As there are not standard datasets for the automatic news semantic annotation on sport domain, the evaluation is based on our dataset consisting of 387 news items crawled from different sources including SkySport, ESPN, PremierLeague.com, BBC Sport. The dataset comprises actually 130 news items of Premier League and UEFA Champion League and 237 items of football transfer category.

Fergie defends Rooney temperament

Manchester United manager Sir Alex Ferguson has backed under-fire England striker Wayne Rooney following his red card against Montenegro last Friday. Rooney, 25, was sent off in the second half of England's 2-2 draw in Montenegro that booked the Three Lions' place at Euro 2012 for kicking out at Miodrag Dzulovic.



UEFA's Control and Disciplinary Body is to meet on Thursday in Nyon to make a decision over the length of the ban Rooney will serve at next summer's tournament.

Ferguson has yet to see the incident but has defended the forward, expressing his belief that Rooney's temperament has improved over recent years. "I texted him (Rooney) but he's not got back to me yet," Ferguson said on The Football Show. "Obviously he will be disappointed."

| Entity recognized by algorithm | Value |
|--------------------------------|---|
| SportPerson | Wayne Rooney |
| SportPersonReasoning | Wayne Rooney's behaviour has shown signs of improvement |
| SportTeam | Manchester United |
| Forward | Wayne Rooney |
| OriginalName | Wayne Rooney |
| Inst | http://bk.sport.owl#wayne_rooney |

```

<owl:NamedIndividual rdf:resource="http://bk.sport.owl#news_00000">
  <rdf:type rdf:resource="http://bk.sport.owl#News"/>
  <bk:hasUrl rdf:resource="http://www.bbc.com/sport/0/football/2012/02/120225_fm_uefa_disciplinary_rooney">
  <bk:hasTitle xml:lang="en"> Fergie defends Wayne Rooney temperament</bk:hasTitle>
  <bk:about rdf:resource="http://bk.sport.owl#wayne_rooney"/>
  <bk:about rdf:resource="http://bk.sport.owl#alex_ferguson"/>
  <bk:contain rdf:resource="http://bk.sport.owl#manchester_united_fc"/>
  <bk:hasRelation rdf:resource="http://bk.sport.owl#defend_00001"/>
</owl:NamedIndividual>
<owl:NamedIndividual rdf:about="http://bk.sport.owl#defend_00001">
  <rdf:type rdf:resource="http://bk.sport.owl#Defend"/>
  <bk:hasAgent rdf:resource="http://bk.sport.owl#alex_ferguson"/>
  <bk:hasObject rdf:resource="http://bk.sport.owl#wayne_rooney"/>
</owl:NamedIndividual>
    
```

Fig. 7. Named entity identification and generated semantic annotation about simple event

```

<owl:Thing rdf:about="http://bk.sport.owl#andriy-shevchenko">
  <saidThat rdf:resource="http://bk.sport.owl#statement9"/>
</owl:Thing>
<owl:Thing rdf:about="http://bk.sport.owl#statement9">
  <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Statement"/>
  <rdf:subject rdf:resource="http://bk.sport.owl#chelsea-fc"/>
  <rdf:predicate rdf:resource="http://bk.sport.owl#defeat"/>
  <rdf:object rdf:resource="http://bk.sport.owl#tottenham-hotspur-fc"/>
</owl:Thing>
    
```

| Feature | Value |
|--------------|--|
| subject | http://bk.sport.owl#luis-suarez |
| rule | Transfer |
| predicate | http://bk.sport.owl#transferTo |
| class | null |
| originalName | Luis Suarez has arrived in England to complete his move to Manchester City |
| object | http://bk.sport.owl#manchester-city-fc |
| info | relation |

Fig. 8. Recognized semantics about indirect speech and football transfer

We assess the quality of two tasks in our approach: named entity recognition as an instance in

sport knowledge base and semantic annotation extraction. Figure 7 shows an example demonstrating that named entities are identified as instances of classes in the ontology and certain semantic annotations about sport event are created. Figure 8 demonstrates a case in which semantic annotations about indirect speech and football transfer are generated correctly.

Each task is evaluated w.r.t precision and recall defined as follows:

$$P = \frac{\text{Relevant recognized instances (triples) (RR)}}{\text{Total recognized instances (triples) (TR)}} 100 (\%)$$

$$R = \frac{\text{Relevant recognized instances (triples) (RR)}}{\text{Total relevant instances (triples) (TRE)}} 100 (\%)$$

Table 1 demonstrates the positive performance in instance detection and generating semantic annotation of proposed method when testing with general football news items sub dataset.

Table 1. Precision and Recall score for instance detection and triples generation on general football news

| Task | TR | RR | TRE | P% | R% |
|----------------------------|------|------|------|-------|-------|
| Named Entities Recognition | 2699 | 2692 | 4415 | 99,74 | 60,97 |
| Triples Extraction | 1002 | 890 | 1663 | 88,82 | 53,52 |

Table 2 shows the experimental results on generating semantic annotations for sub dataset about football transfer in two scenarios: using pronoun annotation and not. We can see that this technique improves the detection of instances in sentences, thus brings better recall score of semantic triples extraction.

Table 2. Performance on football transfer dataset

| Triples extraction | TR | RR | TRE | P% | R% |
|----------------------------|-----|-----|-----|------|------|
| Without pronoun annotation | 180 | 145 | 264 | 80.5 | 54.9 |
| With pronoun annotation | 213 | 173 | 264 | 81.2 | 65.5 |

5. Related work

Certain works have been carried out on the development of the semantic annotation framework. Some among them provide only manual annotation such as [7] while others aim at addressing this problem in the general domain [4].

The KIM provides itself a solution for automatic semantic annotation. However, as other platforms

Ontea [8], C-Pankow[3] semantic annotation is limited at assigning entities in the text to their semantic descriptions defined by an ontology. Our work is among the first attempt in developing an automatic method for this problem on the sport domain. Our contribution is not only more effective instance detection, but also the capacity of generating annotations in the form of triples which represent certain important semantics in a news item.

6. Conclusion

In this paper, we have studied the problem of generating semantic annotation for news items in a sport aggregation system. The novelty of our system lies in the combination of an effective instance detection using knowledge base on sport and a deep analysis of language patterns representing certain semantics of the text. One of the strengths of this method is that the generated annotations are formalized in the form of triples indicating important entities, simple events, indirect speech and football transfer information which are not considered in related works.

Thanks to many improvements, the proposed method proves effective in our experimental study with positive precision and recall scores on both two tasks: named entity detection and semantic extraction. As future work we will focus on the problem of learning extraction rules to enhance the scalability of the approach. Also, we intend to extract more complex semantics from news articles and represent them in a proper model such as quadruple.

References

- [1] Bechhofer, S., Carr, L., Goble, C., Kampa, S. and Miles-Board, T., The Semantics of Semantic Annotation. In Proceedings of the 1st International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems. 1151-1167
- [2] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. Scientific American, 284(5), pp. 34-43.
- [3] Cimiano, P., Ladwig, G., Staab, S., Gimme' the context: context-driven automatic semantic annotation with C-PANKOW, in: Proceedings of the 14th International World Wide Web Conference, Tokyo, Japan, 2005.
- [4] Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., McCurley, K.S., Rajagopalan, S., Tomkins, A., Tomlin, J.A., Zienberer, J.Y., A case for automated large scale semantic annotation, J.Web Semantics 1 (1) (December 2003).
- [5] Ding, Y., Sun, Y., Chen, B., Börner, K., Ding, L., Wild, D., Wu, M., DiFranzo, D., Fuenzalida, A.G.,

- Li, D., Milojević, S., Chen, S., Sankarangarayanan, M., Toma, I., Semantic Web Portal: A Platform for Better Browsing and Visualizing Semantic Data. Proceedings of the 2010 International Conference on Active Media Technology, Toronto, Canada
- [6] Fernández, N. Semantic Annotation Introduction, (2010). Available at <<http://www.it.uc3m.es/labgimi/teoria/Module2/SA-Intro.pdf>>
- [7] Handschuh, Staab, S., Studer, R., Leveraging metadata creation for the Semantic Web with CREAM, in Proceedings of the Annual German Conference on AI, September 2003
- [8] Laclavik, M., Ciglan, M., Šeleng, M., Krajčí, S., Ontea: Semi-automatic Pattern based Text Annotation empowered with Information Retrieval Methods, Tools for Acquisition, Organisation and Presenting of Information and Knowledge (2007), 119-129.
- [9] Nguyen, Q-M., Cao, T-D.; A novel approach for automatic extraction of semantic data about football transfer in sport news. International Journal Pervasive Computing and Communications, Vol. 11 Iss: 2, pp. 233-252, ISSN: 1742-7371 (2015)
- [10] Popov, B., Kirayakov, A., Ognyanoff, D., Manov, D., Kirilov, A., KIM—a semantic platform for information extraction and retrieval, Nat. Lang. Eng. 10 (3/4) (2004) 375–392
- [11] Talantikite, H.N., Aïssani, D., Boudjlida, N. Semantic annotations for web services discovery and composition. Computer Standards & Interfaces. Vol. 31, N°6. 1108-1117(2009)
- [12] Rayfield, J., Wilton, P., Oliver, S., “Sport ontology”.<http://www.bbc.co.uk/ontologies/sports>