Barometr Regionalny
Tom 12 nr 1

Ontology of Information about an Enterprise Found in Press Articles

Jan Andreasik
University of Management and Administration in Zamość, Poland

Abstract

In the paper the ontology of information about an enterprise (OIE) is presented. The ontology helps to analyse the information about the examined enterprise. The aim of analysis is to isolate information about an enterprise found in business and financial press articles (news items) in order to define the strategic position of the company. This information is "relative propriety" in the terminology of the Polish philosopher Roman Ingarden's formal ontology. The author proposes the OIE ontology of information about an enterprise (company) based on the theory of individual objects by Roman Ingarden. Analysis of information from press news will provide a method for automatic gathering of data on the current state of an enterprise.

Keywords: ontology of information, OIE, Ingarden, press article

Introduction

Every enterprise quoted on the stock exchange is obliged by adequate provisions to provide current information concerning its financial results, capital structure, market value, revenue from sales, financial liquidity, personnel changes, ownership changes, implemented projects, concluded contracts, implemented investment ventures, etc. This information is published in business newspapers, reports and on websites. Undertaking an investment decision (such as buying or selling shares, obligations or other securities) is connected with the process of obtaining information about a particular enterprise. This information should be synthetic and current, and it should cover those aspects of the enterprise which concern the potential estimation of its market value. An investor should be able to find answers to the following questions:

Is the company viable?
Is the company a leader in its business line?
Is the company operating on a developing market?
Is the company launching the manufacture of a new product?
Is the company implementing investment projects?
Is the company acquiring new orders (contracts)?
Is the company losing clients?
Is the company losing sub-suppliers?
Is the company's own capital increasing?
Is the company settling its liabilities?
Is the company paying dividends?
Is the company hiring workers?
Is the company dismissing workers?
Does the company have financial liquidity?
Does the company have a good prognosis for its financial condition?
What is the value of the market in which the company is operating?

These kinds of questions can be put in three groups, on the basis of which the strategic position of the company can be defined according to a "three dimensional strategy model" (Strategor 1995). In this model, a company's position in three strategic dimensions is specified: in the dimension of security, understood as limitation of competition, in the dimension of the business line value, and in the dimension of competence. Third-generation BI (Business Intelligence) systems attempt to extract information from texts by means of a semantic analysis. Therefore, it is important to create a formal tool for such an analysis. The objective of such an analysis should not only be extracting information about appropriately classified facts, or facts in a framework of domain ontology, but also reading a text in such a way so as to gain new knowledge. This article presents an ontological approach to text analysis, whose aim is to isolate information about an enterprise found in business and financial press articles (news items) to define the strategic position of a company.

Let us consider press information concerning the three above enumerated dimensions of the strategic model.

News item no. 1

> The European Commission strongly supports the development of data processing services using cloud computing. The European Commission adopted a new "strategy to drive European business and government productivity via cloud computing". The European Commission estimates that by 2020 2,5 million new jobs will be created in Europe, which means the creation of a completely new branch of industry, whose value will be 80 billion Euro. Cloud computing is one of the fastest growing segments of the IT market in the world, and in Poland as well. The most recent poll carried out by Ip-sos MORI upon the request of Microsoft showed that the percentage of small and medium enterprises using cloud computing is already about 52%.[1]

This information proves the great value of the cloud computing segment on the information services market. If an enterprise conducts this line of business, it has a high position in the dimension: domain value.

News item no. 2

> Polish Defence Holding (former Bumar), the largest defence manufacturer in Poland, talks about the conditions of a merger with Huta Stalowa Wola. The plans for modernizing the army, which through 2022 will invest 139 billion zlotys in modern arms and artillery, vehicles, ships, helicopters, air-defence missile systems and soldier equipment. PHO Defence Group is struggling for orders and to increase its value, because in two years it is planning to go public.[2]

This information confirms the security of the growth policy of PHO Defence Group as a result of the consolidation of a sector's enterprises.

News item no. 3

> PKS Polonus, a coach carrier, has launched a mobile application thanks to which it is possible to

1. Original text in Polish: "Komisja Europejska mocno stawia na rozwój usług przetwarzania danych w chmurze obliczeniowej. Komisja Europejska przyjęła strategię „ New strategy to drive European business and government productivity via cloud computing". Komisja Europejska szacuje, że do 2020 r. w Europie powstanie 2,5 mln nowych miejsc pracy, co będzie oznaczało stworzenie w Unii Europejskiej całkowicie nowej gałęzi przemysłu o wartości 80 bilionów euro. Cloud computing jest jednym z najszybciej rozwijających się segmentów rynku IT na świecie i w tym w Polsce. Obecnie jak wynika z najnowszego badania Ip-sos MORI przeprowadzonego na zlecenie Microsoftu, odsetek korzystających podmiotów MŚP sięga już ok. 52 procent." See: *Rzeczpospolita*, June 1, 2013, Ekonomia & Rynek, B4.

[In the journal (in both Polish and English texts) European practice of number notation is followed — for example, 36 333,33 (European style) = 36 333.33 (Canadian style) = 36,333.33 (US and British style). Furthermore in the International System of Units (SI units), fixed spaces rather than commas are used to mark off groups of three digits, both to the left and to the right of the decimal point. — Ed.]

2. Original text in Polish: "Polski Holding Obronny, dawny Bumar, największa grupa zbrojeniowa w kraju rozmawia o warunkach połączenia z Hutą Stalowa Wola. Polskiemu Holdingowi Obronnemu i podkarpackiej artylerii sprzyjają plany modernizacji armii która do 2022 r. zainwestuje 139 mld zł na nowoczesną broń pancerną i artyleryjską, pojazdy, okręty, śmigłowce, rakietowe systemy przeciwlotnicze i wyposażenie żołnierskie. Obronna Grupa PHO walczy o zamówienia i podniesienie wartości, bo za dwa lata planuje wejście na giełdę." See: *Rzeczpospolita*, July 10, 2013, Ekonomia & Rynek B8.

> check the timetable, buy tickets and locate the coach via a smartphone or a tablet. This application, which is distributed free of charge, is also available for devices with the Android, Windows Phone and iOS operating systems.[3]

This information shows the increase of the competence of PKS Polonus by its introduction of its application on mobile devices, thanks to which access to the company's services is easier.

In BI systems not only information about facts, which can later be associated with the appropriate goals of an enterprise, is extracted, but also knowledge about the relations between enterprises operating on specific markets. This kind of knowledge is difficult to define. The kind of conceptualization which is required here is one which makes it possible to compare facts to get new knowledge about an enterprise. This kind of knowledge comprises a set of properties attached to a particular enterprise, for example: 'the enterprise is a leader in its industry', 'the enterprise has lost its position of a leader in a particular market segment', 'a product manufactured by the enterprise is mostly preferred by users', etc.

The news item entitled: "Koźmiński among Executive MBA leaders"[4] can only be understood if an analysis of the complete news text is carried out according to a defined format, in which the following data can be distinguished:

- the market: studies for upper management personnel
- the rating: rating of Executive MBA "Financial Times" programmes
- the rating criterion: mean salary of the graduates
- the leader: Kellog/Hong Kong UST Business School
- mean annual salary of UST graduates in China: 416,8 thousand dollars
- 47th place: Leon Koźmiński University
- mean annual salary of the Leon Koźmiński University graduates in Poland: 144,1 thousand dollars

In this paper, the author presents the OIE ontology of information about an enterprise (company) based on the theory of individual objects by Roman Ingarden, a Polish philosopher. OIE ontology helps to identify ontologically interesting elements in a news item and maps the news item onto a semantic network, thanks to which the properties of the object (enterprise) are defined.

1 Extracting information from press news

The research conducted so far, concerning extraction of information from press news, has concentrated on identifying this information according to a particular pattern, allowing for the identification of a fact, which is later associated within a defined domain ontology.

Eivind Bjoraa (2009) worked out a multi-agent system, in which an agent processes information written in a natural language (English) according to a particular pattern from domain ontology (ontology of business news from Reuters) into the RDF language code, and then associates it with appropriate financial data of the XBRL standard.

News item:

The loss before tax is expected to be approximately GBP 39 million.

Diagram 1 shows a graph of the news item structure. In accordance with the procedure of extracting information, the agent first recognizes the currency, then determines the amount, and then sets about assigning it in account terms. Figure 2 shows a graph of an exemplary news item in the RDF language.

3. Original text in Polish: "PKS Polonus, przewoźnik autokarowy, uruchomił mobilną aplikację, która umożliwia sprawdzenie rozkładu jazdy, kupowanie biletów oraz lokalizację autokaru za pomocą smartfona i tabletu. Dystrybuowana za darmo aplikacja dostępna jest na urządzenia z systemem operacyjnym Android, Windows Phone i iOS." See: Rzeczpospolita, June 27, 2013, Ekonomia & Rynek B6.

4. Original Polish title: "Koźmiński wśród liderów Executive MBA." See: Rzeczpospolita, October 22, 2013, Ekonomia & Rynek B5.

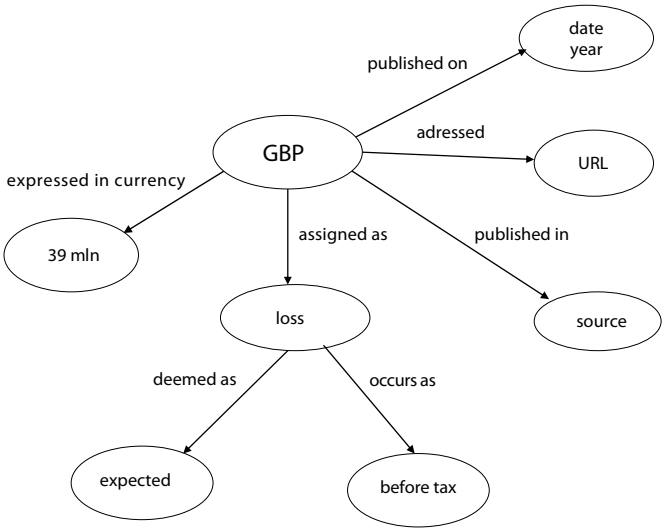

Fig. 1. A business news item graph

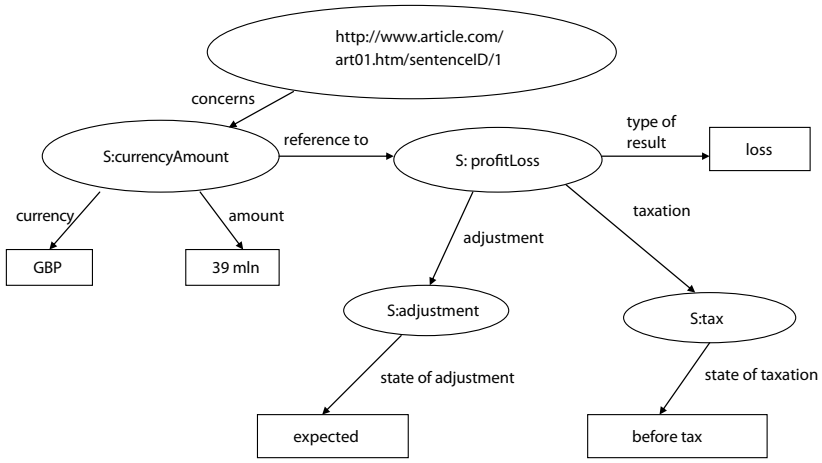

Fig. 2. A fragment of business ontology in the RDF language for a selected example

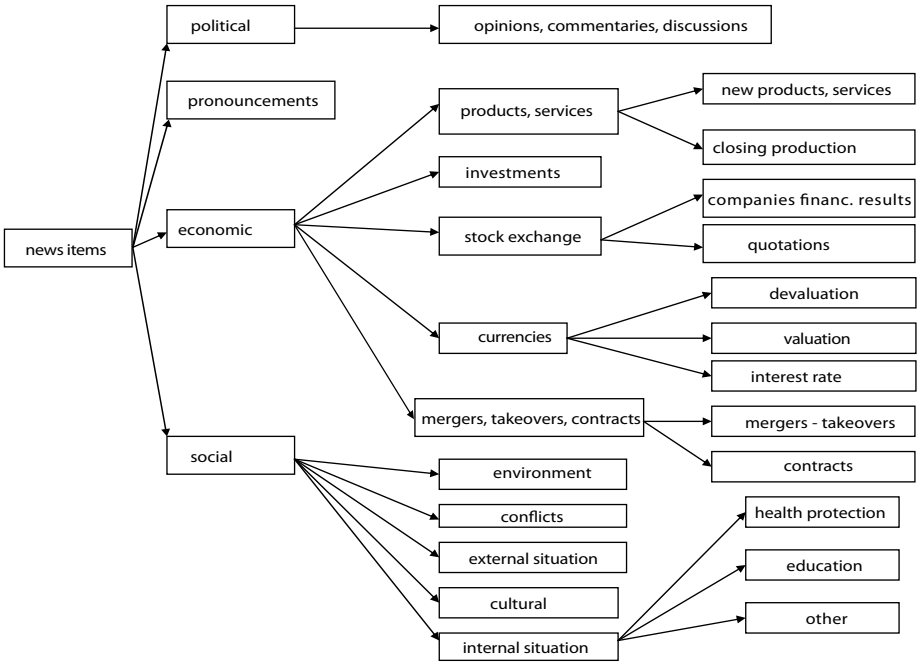

Fig. 3. Ontology of financial news.

Source: (Mellouli, Bouslama, and Akande 2010)

Mullouli et. al. (2010) presented a concept of a system sorting the business news items of Reuters and Bloomberg into categories within the financial news ontology. Each category (concept) has a distinguished set of attributes (properties). The process of text analysis is based on distinguishing attributes in a text and matching them to a set of attributes defining a particular ontological concept. The news item is divided into four basic categories: social, business, informative (announcements) and political. The full financial ontology is shown in figure 3.

For example, the news item:

The stock exchange in Toronto shows a drop in the inflation rate.

This news item is classified into the category: flotation <stock exchange> with the following attributes:

- actor: the stock exchange in Toronto
- determination of activity: there is a drop
- parameter: inflation rate

2. OIE Ontology of news about an enterprise (company)

In the author's opinion information about an enterprise is a pronouncement; a particular judgment is presented together with evidence. For example the news item:

> Chitika Inc., which analyses the mobile advertisement market, claims that 84,3% of tablet web traffic was in June 2013 generated by Apple's iPads. Amazon's Kindle Fire took second place with 5,7% and Samsung Galaxy claimed third place with 4,2%. IPad usage share increased from June to May, Chitika claims, whereas the share of other companies dropped. Regardless of what you might hear or read about how many are bought or sold or activated, iPad is used more than any of the rest, says Cook, Apple's chief executive, adding that iPad is used 4 times more than all the other tablets put together.[5]

This information may be read using the relation form categories as defined by Roman Ingarden (1987) in the following way:

Relation bond: The comparison relation between Apple and Amazon is specified by Chitika Inc.'s report, which analyses the mobile advertisement market. The report presents a rating of tablets according to the criterion of web share usage.

Fundamentum relationis: From what Tim Cook, Apple's chief executive, says, we find that the constitutive nature of iPads is its Web usage share. It is a value which is more important than the number of sold or activated iPads. Tablet web traffic constitutes fundamentum relationis.

Carrier A (Apple): The carrier is the process of Web traffic generated by iPads. It is the result of using iPads. Apple's iPad Web traffic share is 84,3%.

Carrier B (Amazon): The carrier is the process of Web traffic generated by Kindle Fire. Amazon's Kindle Fire Web traffic share is 5,7%.

A relation segment: Apple's iPads.

B relation segment: Amazon's Kindle Fire tablets.

A relation exponent: iPad is used more often than all the rest.

B relation exponent: Kindle Fire takes second position.

The above enumerated details are obtained from the analysis of the news text, which can be presented using the diagram in figure 4.

5. Original text in Polish: "Firma Chitika Inc., która analizuje rynek reklamy mobilnej, twierdzi, że 84,3 procent ruchu w Internecie generowanego przez tablety pochodziło w czerwcu 2013 z iPadów Apple'a. Na drugim miejscu znalazł się Kindle Fire Amazona z 5,7 procent a na trzecim tablety Samsung Galaxy z 4,2 procent. Wykorzystanie iPadów wzrosło w okresie od czerwca do maja, twierdzi Chitika, podczas gdy udział innych firm spadł. Bez względu na to, co czytacie lub słyszycie o wynikach sprzedaży lub aktywacji, iPad jest wykorzystywany częściej niż cała reszta – powiedział Cook (prezes Apple'a) – dodając, że iPad jest wykorzystywany czterokrotnie częściej niż konkurencja." *Rzeczpospolita*, October 24, 2013, Ekonomia & Rynek, B6.

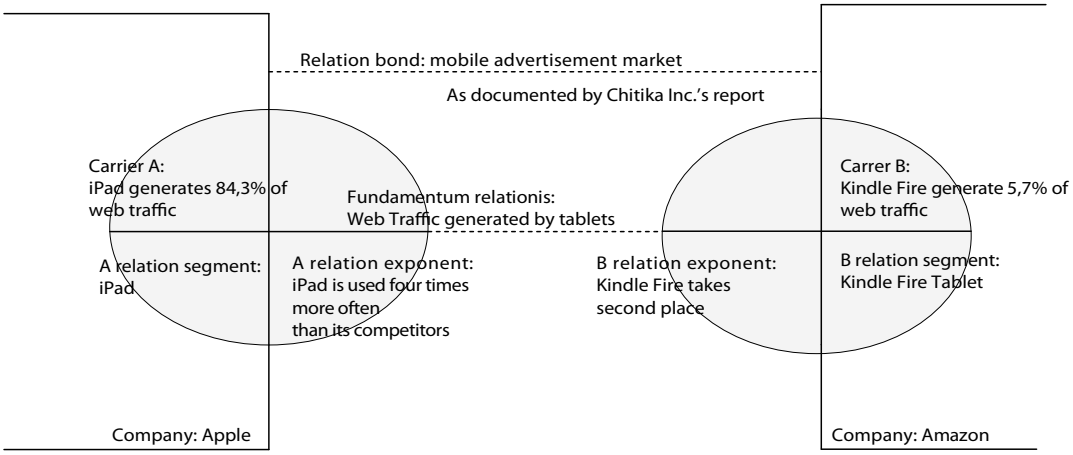

Fig. 4. The form of a relation between two companies producing tablets according to the theory of objects by Roman Ingarden

In order to assign a relative state of matters to object A (Apple): iPad is used four times more often than its competitors, a rule concerning the definition of all the relation categories enumerated in figure 5, which provide material and formal evidence for the truth of the judgment, should be fulfilled.

For further analysis the author adopts the following terminology in order to define classes in the ontology of news about an enterprise:

RC — relationship criterion — in Ingarden's definition of relation it corresponds to the category: 'Fundamentum relationis'. The relation criterion constitutes the main criterion for the comparison of objects A and B within the rating. In the said example, the criterion is the tablet Web traffic.

RD — relation documentation — in Ingarden's definition of relation it corresponds to the category: 'relation bond'. The documentation comprises a report, expert opinion, scientific paper or business documentation in which the rating or comparison of A and B is presented. In the said example it is Chitika Inc. report — Mobile Advertisement Market Analysis.

(A) RS — Company A's share in the relation — in Ingarden's definition of relation it corresponds to the category: 'Carrier A'. The object share in the relation is defined by giving a parameter number in the process of using Company A's product or device, which represents Company A in the said relation. In the said example it is the share of iPad in tablet Web traffic.

(B) RS — Company B's share in the relation — in Ingarden's definition of relation it corresponds to the category 'Carrier B'. The object share in the relation is defined by giving a parameter number in the process of using Company B's product or device, which represents Company B in the said relation. In the said example it is the share of Kindle Fire in tablet Web traffic.

(A) RR — Company A representative in the relation — in Ingarden's definition of relation, it corresponds to the category: 'relation A segment'. Company A representative comprises the products, services and processes which are compared to the representatives of Company B. In the said example, the representative of Apple is iPad.

(B) RR — Company B representative in the relation — in Ingarden's definition of relation, it corresponds to the category: 'relation B segment'. Company B representative comprises the products, services and process which are compared to the representatives of Company A. In the said example, the representative of Amazon is Kindle Fire.

AR (A/B) — assessment of Company A in relation to Company B — in Ingarden's definition of relation it corresponds to the category: 'left relation exponent', which is the so called 'relative property'. The assessment of Company A in relation to Company B may be either positive or negative. In the said example A/B assessment is the following: iPad is used more often than all the rest. It is a positive assessment which designates a positive state of matters for company A.

AR (B/A) — assessment of Company B in relation to Company A — in Ingarden's definition of relation it corresponds to the category: 'right relation exponent', it is the so called 'relative property'. In the said example the B/A assessment is the following: Amazon's Kindle Fire took second place.

PC — companies with a positive state of matters — it is the class of companies for which a positive state of matters is defined. In this example Apple is an instance of this class.

NC — companies with a negative state of matters — it is the class of companies for which a negative state of matters is defined. In this example Amazon is an instance of this class.

Figure 5 shows the ontology of information about an enterprise taking into account the above enumerated concepts (classes) and object properties between them. The following object properties are distinguished:

```
CA < is A represented> (A) RR
CB < is B represented> (B) RR
(A) RR <has an A share in the relation> (A) RS
(B) RR <has a B share in the relation> (B) RS
(A) RS <is A assessed> (A/B) AS
(B) RS <is B assessed> (B/A) AS
RD <comprises assessment of A> (A/B) AR
RD <comprises assessment of B> (B/A) AR
RC <is documented> RD
RC <has an A share according to the criterion> (A) RS
RC <has a B share according to the criterion> (B) RS
(A, B) N <is quoted> RD
(A, B) N <is in a relation to A> CA
(A, B) N <is in a relation to B> CB
(A/B) AR <has a positive state assigned> CA
(B/A) AR <has a negative state assigned> CB
```

Each OIE ontology concept constitutes a class with distinguished attributes (data types). Below there is a list of attributes for each OIE ontology class with appropriate instances from the presented example.

```
Class CA          <Company A>
Attributes:       <company name>: Apple
                  <CEO>: Tim Cook
                  <activity>: computer production
                  <stock exchange>: New York
Class: (A) RR     <representative of A Company in the relation>
Attributes:       <product type>: tablet
                  <commercial name>: iPad
                  <price>: $499
Class: RC         <relation criterion>
Attributes:       <market name>: mobile advertisement market
                  <market parameter>: tablet Web traffic
Class: (A) RS     <Company A share in the relation>
Attributes:       <commercial name>: iPad
                  <A market parameter>: tablet A Web traffic
                  <share value>: 84,3%
Class: RD         <relation documentation>
Attributes:       <documentation type>: report
                  <author of the report>: Chitika Inc.
                  <date of issue>: 2013
                  <period>: May-June 2013
                  <issuer>
                  <www address>
Class: (A, B) N   <news about A, B company>
Attributes:       <place of publication>: Rzeczpospolita, Ekonomia & Rynek
                  <date>: 24 Oct 2013
                  <index>: B9
                  <Company A name>: Apple
                  <Company B name>: Amazon
```

```
Class: (A/B) AR <assessment of Company A in relation to Company B>
Attributes:     <assessment content>: "iPad is used four times more often than its
                                      competitors"
                <linguistic variable>: four times
                <number parameter of the assessment>: 1st place in the rating
                <state>: positive
```

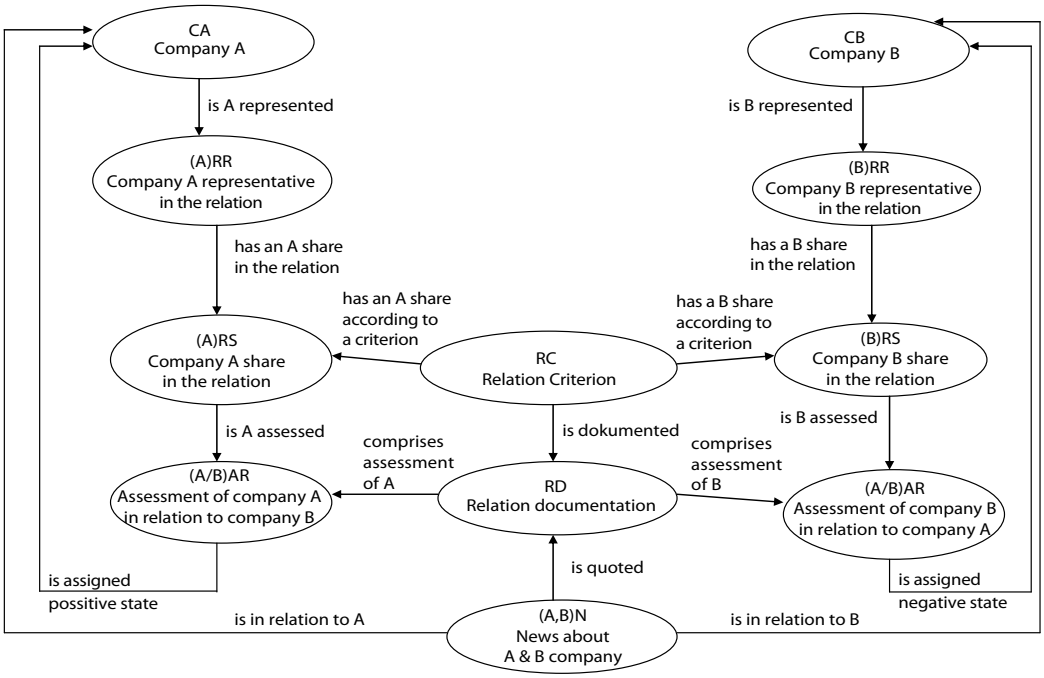

Fig. 5. Ontology of information about an enterprise (company)

According to the diagram above, instance B, Amazon, can be presented.

The author of this study has created a database from press news about companies quoted on the Warsaw Stock Exchange according to OIE ontology.

Conclusion

Extracting information from press news is a matter which does not only mean distinguishing the grammatical structure of a sentence written in a natural language. The process of recognition cannot only be limited to finding data about facts in accordance with a particular structure. Such a news item should be approached according to epistemological categories (Woleński 2005). Information about an enterprise should be given in the form of true judgments. To determine that a particular company A is the leader in its line of business, the company should be compared to a different enterprise according to a particular criterion. This comparison is a kind of rating, so the rating which should be quoted here is a definite rating carried out by an appropriate agency, which has a definite level of credibility.

In this study the author has presented the OIE ontology of information about an enterprise through an analogy to Roman Ingarden's concept of individual objects. The study comprises the thesis that it is possible to map the concepts of ontology defined on the grounds of philosophy (ontology for the first time according to Garbacz and Trypuz) onto domain ontology defined on the grounds of information studies (ontology for the second time) (Garbacz and Trypuz 2012). The author has created a database about companies quoted on the Warsaw Stock Exchange on the basis of information obtained from press news from the supplement "Ekonomia & Rynek" to *Rzeczpospolita* daily.

References

Bjoraa, E. 2009. *Ontology Guided Financial Knowledge Extraction. From Semi-Structured Information Sources*: VDM Verlag.

Garbacz, P., and R. Trypuz. 2012. *Ontologie poza ontologią. Studium metateoretyczne u podstaw informatyki*. Lublin: Wydawnictwo KUL.

Hobbs, T.L., U. Karmarkar, and J.R. Hobbs. n.d. Open-Domain Information Extraction from Business News. http://www.isi.edu/~hobbs/open-domain.pdf.

Ingarden, R. 1987. *Spór o istnienie świata, t. 2, Ontologia formalna, cz. 1, Forma i istota*. 3rd ed, Dzieła filozoficzne/R. Ingarden. Warszawa: Państwowe Wydawnictwo Naukowe.

Lupiani-Ruiz, E., I. Garcia-Manotas, R. Valencia-Garcia, F. Garcia-Sanchez, D. Castellanos-Nieves, J.T. Fernandez-Breis, and J.B. Camon-Herrero. 2011. "Financial News Semantic Search Engine." *Expert Systems with Applications* no. 38 (12): 15565–15572. doi: 10.1016/j.eswa.2011.06.003.

Mellouli, S., F. Bouslama, and A. Akande. 2010. "An Ontology for Representing Financial Headline News." *Journal of Web Semantics* no. 8 (2–3): 203–208. doi: 10.1016/j.websem.2010.02.001.

Nowak, A.J., and L. Sosnowski. 2001. *Słownik pojęć filozoficznych Romana Ingardena*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych "Universitas".

Saggion, H., A. Funk, D. Maynard, and K. Bontcheva. 2007. "Ontology-Based Information Extraction for Business Intelligence." *Semantic Web, Proceedings* no. 4825:843–856.

Strategor. 1995. *Zarządzanie firmą. Strategie, struktury, decyzje, tożsamość*. Translated by K. Bolesta-Kukułka. Warszawa: Państwowe Wydawnictwo Ekonomiczne.

Woleński, J. 2005. *Epistemologia. Poznanie, prawda, wiedza, realizm*. Warszawa: Wydawnictwo Naukowe PWN.