The Paradox of the Fuzzy Disambiguation in the Information Retrieval

Anna Bryniarska
Department of Electrical and Computer Engineering
Opole University of Technology
Opole, Poland
Email: a.bryniarska@doktorant.po.edu.pl

Abstract—**Current methods of data mining, word sense disambiguation in the information retrieval, semantic relation, fuzzy sets theory, fuzzy description logic, fuzzy ontology and their implementation, omit the existence of paradox called here the paradox of the fuzzy disambiguation. The paradox lies in the fact that due to fuzzy data and the experts knowledge it can be obtained precise knowledge. In this paper to describe this paradox, is introduced a conceptual apparatus. Moreover, there is formulated an information retrieval logic. There are suggested certain applications of this logic to search information on the Web.**

Index Terms—**fuzzy disambiguation paradox, Description Logic, FuzzyDL, Information Retrieval Logic, Semantic Web.**

I. Introduction

Recently information retrieval (*IR*) on the Semantic Web usually means searching a reliable source of information. So far systems of information retrieval and systems of semantic relation indicated only for the most semantically similar source of searched information [13], [14].

To define the semantic relationships is typically used measurement of the keywords incidence. However, meaning of these words is exactly identified and represents certain knowledge. Therefore, these *IR* methods cannot always be used.

Sometimes, the information retrieval about an object can lead to uncertain knowledge described in the appropriate, ontology language. In spite of this uncertainty, it can be found a disambiguated source of information about this object. In this way, the compliance with the description of the object model (compliance with the thesaurus) is obtained. The situation described above is called here **the paradox of the fuzzy disambiguation in the information retrieval**. Methods of data mining [11], word sense disambiguation in information retrieval [2], [7], [8], [13], [16], semantic relation [7], [8], fuzzy sets theory and fuzzy logic [10], fuzzy description logic [3], [4], [17], [18], and also their implementation (i.e. in OWL language), do not concern about this paradox. The following are two examples to illustrate this paradox.

A. Example 1

Data from user of X iron: the fabric *Y* shrunk and waved after ironing.

Data from expert: some threads of the fabric *Y* shrink, but only in the water steam at 100°C. Only program 1 uses the water steam.

Data from user of the *X* iron are uncertain, if we want to find out which program was used for ironing. However, from these uncertain data indicate a precise data: during ironing was turned on the program 1 with the steam. Reasoning in this case uses particular data from the expert. The situation that when ironing the fabric was only heated by the iron is excluded with the experts knowledge.

B. Example 2

Whether "Ralf Möller", a German bodybuilder and TV actor and "Ralf Möller", a professor at the Hamburg University of Technology, is the same person?

Data from Web resources: Ralf Möller, the actor, born in 1959 r.; Ralf Möller, the professor, born when the German Chancellor was Ludwig Erhard (from history it was between 1963 and 1966).

Data from expert: Any attribute for a single person has only one value.

Data on the Web resources are ambiguous and uncertain. Names of people are: *name(person1)="Ralf Möller"* for the actor, *name(person2) = "Ralf Möller"* for the professor. Whether *person1 = person2*? Complementing the uncertain knowledge of the attributes values of expert knowledge: *birthDate(person1) = 1959, birthDate(person2) > 1963*, it can be concluded that *person1 ≠ person2*. The result of reasoning is accurate information.

II. Residuum rule in *IR*

The information retrieval on the Semantic Web is based on finding a copy of data which are:

1) values of single attribute arguments, i.e. **concepts** – data representing knowledge of certain properties or object types,
2) values of multi attribute arguments, i.e. **roles** – data representing knowledge about relationships between objects.

Firstly, concepts and roles are described by the language of the *Description Logic* (*DL*) [1]. Secondly, the *DL* logic, describing concepts and roles, is extended for some first-order logic formulas. Then, in this extended logic, is created the **thesaurus** – language describing the reference concepts and roles. While the **ontology** describes the real, found on Web pages, concepts and roles which are searched. If the interpretation of

concepts and roles from the ontology, accordingly to experts knowledge and criteria, will result in the interpretation of concepts and roles of thesaurus, then this relationship is called **the residuum**. This interpretation determines the membership degree of the data copies (set of the Web addresses *X*). This degree also includes the semantic structure of the resources, determined by the Semantic Web.

Due to fuzzy degree of knowledge, concepts and roles can be interpreted as fuzzy sets in the space $X \cup X \times X$ of the Web addresses and their pairs. Then can be made **the fuzzification** of knowledge [5]. Whereas setting residuum is necessary to make the knowledge **defuzzification** [5]. Then for a given query, it can be indicate, a reliable-for-experts set of Web addresses representing this knowledge. Interpretation sets forming residuum will be further treated as an information search result. Then, the following search rule is applied.

Firstly, the question (search query) is compared to the thesaurus. Secondly, if for all interpretation the set of searched addresses is empty, then this question is compared to ontology, so that the compared entry has the most similar meaning to the thesaurus. Found, for this entry, the set of Web addresses represents knowledge which is identical or the most similar to the searched one. This rule of *IR* is called **the residuum rule**. Below is introduced **the information retrieval logic** (*IRL*) using this rule. Applying this logic to the information search is an attempt to develop a new, more universal *IR* method, using the artificial intelligence.

III. LANGUAGE OF *IRL*

In the context of the Semantic Web research [1], [3]–[5], the representation of knowledge in the Semantic Web can be defined by the attribute language (*AL*) of the Description Logic (*DL*) [1]. Then knowledge is represented by concepts **TBox**, roles **RBox** and assertions **ABox**. The Semantic Web might be extended by edges representing relationships between concepts and roles. Descriptions of these edges are called **axioms**. Then knowledge is represented by two systems: the terminology called **TBox** and the set of assertions called **ABox**. Where the assertion is the relationship between the concept or the role and their instances.

Further is presented the syntax of the language *AL* for the *IRL*, analogously to the fuzzy Description Logic (*fuzzyDL*) [3]. Articles [10], [11] show the semantic and the interpretations of this language which are called the fuzzification and the defuzzification.

A. Syntax of TBox

The following names are included to the set of concepts and roles names:

The universal concept $\top$ (Top) and *the empty concept* $\perp$ (Bottom).

The universal concept includes all instances of concepts and the empty concept informs about no instance of a concept.

Let C, D be the names of the concepts, R be the name of a role, and m be the modifier. Then complex concepts are:

$\neg C$ – *the concept negation*; it means all instances of concepts which are not an instance of the concept C;

$C \sqcap D$ – *the intersection of concepts* C and D; it means all instances of both concepts C and D;

$C \sqcup D$ – *the union of concepts* C and D; it means all instances either of the concept C or the concept D;

$\exists R.C$ – *the existential quantification*; it means all instances of the concept C which are in role R with at least once occurrence of the concept C;

$\forall R.C$ – *the universal quantification*; it means all occurrence of the concept C which is in role R with some occurrence of the concept C;

$m(C)$ – *the modification* m *of the concept* C; it means the concept C which is modified by the word m. For example m can occur as a word: very, more, the most, high, higher or the highest.

Concepts which are not complex are called *atomic*.

B. Syntax of ABox

For any concepts instances t_1, t_2, the concept name C and the role name R, **the assertions** are "$t_1 : C$", "$(t_1, t_2) : R$". We read them: t_1 is an instance of the concept C, the pair (t_1, t_2) is an instance of the role R.

For any concepts instances t_1, t_2, the concept name C and the role name R, **the assertions with membership degree** α are "$< t_1 : C, \alpha >$", "$< (t_1, t_2) : R, \alpha >$". We read them: t_1 is an instance with membership degree α of the concept C, the pair (t_1, t_2) is an instance with membership degree α of the role R.

C. Syntax of axioms TBox

For any concepts names C, D and any number $\alpha \in [0, 1]$, the axioms are:

$C \sqsubseteq D$ – the concept C is the concept D,

$C = D$ – the concept C is identical with the concept D,

$< C \sqsubseteq D, \alpha >$ – the concept C is the concept D in the α degree,

$< C = D, \alpha >$ – the concept C is identical with the concept D in the α degree,

D. Syntax of axioms RBox

For any roles R_1, R_2 and any number $\alpha \in [0, 1]$, the axioms are:

$R_1 \sqsubseteq R_2$ – the role R_1 is the role R_2,

$R_1 = R_2$ – the role R_1 is identical with the role R_2.

$< R_1 \sqsubseteq R_2, \alpha >$ – the role R_1 is the role R_2 in the α degree,

$< R_1 = R_2, \alpha >$ – the role R_1 is the role R_2 in the α degree,

E. Syntax of formula

Any assertion or axiom is a formula. For any formula φ, ϕ, a variable x and a number $\alpha \in [0, 1]$, formulas are:

$\neg \varphi$ – the negation;

$< \varphi, \alpha >$ – the formula φ true in the degree α;

$\forall_x \varphi$ – the existential quantification of the formula φ for variable x;

$\exists_x \varphi$ – the universal quantification of the formula φ for variable x;

$\varphi \Rightarrow \phi$ – the implication;

$\varphi \wedge \phi$ – the conjunction;

$\varphi \vee \phi$ – the alternative;

$\varphi \Leftrightarrow \phi$ – the equivalence formulas.

IV. POSTULATES OF FUZZYDL AND FUZZY DISAMBIGUATION IN *IR*

The occurrence of the paradox of the fuzzy disambiguation in *IR* determine the following postulates (**P1** – **P9**):

P1. There is a thesaurus which is a set of certain reference terms and formulas of the *IRL* language. Thesaurus terms represent knowledge in the same area as searched information and can be found in a text document (from thesaurus).

P2. An ontology includes all terms and formulas of the *IRL* language, which are semantically related to the searched information. The ontology includes the thesaurus. All thesaurus formulas are constructed of certain terms and assertions from the base set *Tez*. Likewise, all ontology formulas are constructed of certain terms and assertions belonging to the base set *Ont*. The degree of semantic similarity of ontology expressions to the thesaurus expressions is determined by an expert system based on the experts knowledge.

P3. The space *IR* is a group of addresses of knowledge resources on the Web, semantically related to the terms and formulas representing the searched information. Knowledge resources are text documents available at these addresses.

P4. Finding information is to search the text document, which semantic structure (terms from the Semantic Web created by the ontology) is the most similar to the structure of the thesaurus document. If both documents contains expressions that are equally used by agents in the communication process, then these expressions represent the same knowledge. Furthermore information retrieval is searching for the text document that represent the same knowledge or most similar knowledge to the one from the thesaurus [6]. Therefore, the residuum rule of information retrieval is applied.

P5. The information retrieval of the intersection of concepts represents collective knowledge of these concepts. *Complementary formulas* φ, ϕ are formulas $\varphi \& \phi$ represent the collective knowledge represented by the data set of these formulas.

P6. According to the intuition and practice of *IR*, if $x \in [0, 1]$ is the degree of the semantic similarity of the instance t_1 of a concept C_1 or formula φ to the concept instance or formula belonging to the thesaurus and $y \in [0, 1]$ is the degree of semantic similarity of the instance t_2 of a concept C_2 or formula ϕ to the concept instance or formula belonging to the thesaurus, then the degree of the concepts intersection C_1, C_2 or complementary formulas $\varphi \& \phi$ are a number $x \bullet y$. The operation $\bullet : [0, 1] \times [0, 1] \to [0, 1]$ is some t-norm. This t-norm

has the following properties. For all $x, y, z, x_0, y_0 \in [0, 1]$:

$$x \bullet y = y \bullet x \tag{1}$$

$$(x \bullet y) \bullet z = x \bullet (y \bullet z) \tag{2}$$

$$x \leqslant x_0 \text{ and } y \leqslant y_0 \text{ implies } x \bullet y \leqslant x_0 \bullet y_0 \tag{3}$$

$$1 \bullet x = x; 0 \bullet x = 0 \tag{4}$$

Each t-norm determines uniquely its corresponding implication $\to$ (the residuum), defining a similarity degree of the formulas implications, satisfying for all $x, y, z \in [0, 1]$:

$$z \leqslant (x \to y) \text{ iff } x \bullet z \leqslant y \tag{5}$$

or

$$(x \to y) = \sup\{t \in [0, 1] : x \bullet t \leqslant y\}. \tag{6}$$

The implication of these formulas is semantically similar to the thesaurus formulas, if the similarity degree of the implication predecessor is the closest to its successor. Furthermore, this means that in the found text document is the formula φ, which implies a certain formula ϕ contained in this document, or represents the searched information. When ϕ represents the searched information and is not contained in this document, then the formula φ is supplemented with the formula θ representing the experts knowledge, so that the complementary formulas $\varphi \& \phi$ have the same similarity degree as search formula ϕ (the degree equal 1). Thus, in the case of imprecise implications predecessor φ, the successor is a sharp expression with the semantic similarity degree equal 1 (Example 1).

P7. If in the text document is a conjunction $\varphi \wedge \phi$, according to the classical propositional calculus, then $\varphi \Rightarrow \phi$. Thus, the conjunction is recognized by firstly recognizing the formula φ and secondly by recognizing the implication $\varphi \Rightarrow \phi$. Therefore, the conjunction $\varphi \wedge \phi$ is recognized as $\varphi \& (\varphi \Rightarrow \phi)$. Further is assumed that:

$$\varphi \wedge \phi := \varphi \& (\varphi \Rightarrow \phi) \tag{7}$$

Hence, the similarity degree of the conjunction $\varphi \wedge \phi$ to the thesaurus formulas is defined:

$$x \otimes y := x \bullet (x \to y), \text{ for } x, y \in [0, 1] \tag{8}$$

where x, y are the similarity degrees of formulas φ, ϕ to the thesaurus formulas.

P8. If in the text document is an alternative $\varphi \vee \phi$, then based on propositional logic, it can be assumed that this alternative is recognized based on the following assignment:

$$\varphi \vee \phi := ((\varphi \Rightarrow \phi) \Rightarrow \phi) \wedge ((\phi \Rightarrow \varphi) \Rightarrow \varphi) \tag{9}$$

Hence, the similarity degree of the alternative $\varphi \vee \phi$ to the thesaurus formulas is defined:

$$x \otimes y := ((x \to y) \to y) \otimes ((y \to x) \to x) \tag{10}$$

where $x, y \in [0, 1]$ are the similarity degrees of formulas φ, ϕ to the thesaurus formulas.

P9. The algebra $BL = < L, \otimes, \oplus, \bullet, \to, 0, 1 >$ is a regular residuated lattice (or a *BL-algebra*). It is the algebra such that:

1) $< L, \otimes, \oplus, 0, 1 >$ is a complete lattice with the largest element 1 and the least element 0;

2) $< L, \bullet, 1 >$ is a commutative semigroup with the unit element 1, i.e. $\bullet$ is commutative, associative, and $1 \bullet x = x$ for all x;

3) the following conditions hold (for all $x, y, z \in [0,1]$):

$$z \leqslant (x \to y) \text{ iff } x \bullet z \leqslant y \qquad (11)$$

$$x \otimes y = x \bullet (x \to y) \qquad (12)$$

$$x \oplus y = ((x \to y) \to y) \otimes ((y \to x) \to x) \qquad (13)$$

$$(x \to y) \oplus (y \to x) = 1 \qquad (14)$$

In the BL-algebra can be defined operations of the completeness $'$ and the equivalence $\leftrightarrow$:

$$x' := (x \to 0) \qquad (15)$$

$$x \leftrightarrow y := (x \to y) \bullet (y \to x) \qquad (16)$$

V. FUZZIFICATION

The expressions of the *IRL* logic are interpreted in the regular residuated lattice $BL =< L, \otimes, \oplus, \bullet, \to, 0, 1 >$ and in the chosen, ordered algebra of fuzzy sets:

$$\boldsymbol{F} =< F, \wedge^F, \vee^F, \neg^F, c^F, e^F, 0^F, 1^F, M, F_0 > \qquad (17)$$

Where for the space $X \cup X \times X$, F is a family of fuzzy sets, $\mu : X \cup X \times X \in [0,1]$, described as follow. For any fuzzy set μ there are exactly two fuzzy sets $\mu_1 : X \to [0,1]$ and $\mu_2 : X \times X \to [0,1]$:

$$\mu(x) = \begin{cases} \mu_1(x), & \text{for } x \in X \\ \mu_2(x), & \text{for } x \in X \times X \end{cases} \qquad (18)$$

F is a set of all fuzzy sets in the $\boldsymbol{F}$ algerba, which only apply to mentioned bellow operations and relation, described by t-norm $\bullet$ [8], conclusion, equality and modification norm.

The operation:

1) $\wedge^F$ is intersection of fuzzy sets;
2) $\vee^F$ is a sum;
3) $\neg^F$ is a complement operation;
4) c^F is a function $c^F : F \times F \to [0,1]$ called the degree of containment of fuzzy sets [8];
5) e^F is a function $e^F : F \times F \to [0,1]$ called the degree of equality of fuzzy sets [8].

These operations are defined in the regular residuated lattice $BL =< L, \otimes, \oplus, \bullet, \to, 0, 1 >$ defined as follows. For any fuzzy sets $\mu_A, \mu_B \in F$ and $x \in [0,1]$:

$$(\mu_A \wedge^F \mu_B)(x) = \mu_A(x) \otimes \mu_B(x) \qquad (19)$$

$$(\mu_A \vee^F \mu_B)(x) = \mu_A(x) \otimes \mu_B(x) \qquad (20)$$

$$c^F(\mu_A, \mu_B)(x) = \mu_A(x) \to \mu_B(x) \qquad (21)$$

$$e^F(\mu_A, \mu_B)(x) = \mu_A(x) \leftrightarrow \mu_B(x) \qquad (22)$$

$$(\neg^F \mu_A)(x) = \mu_A(x) \to 0 \qquad (23)$$

The symbol 0^F is any fuzzy set only with values 0, the symbol 1^F is any fuzzy set only with values 1, M is a set of one-argument operations $f : [0,1] \to [0,1]$ called **the modification functions**; F_0 is a subset of F.

Let X is a set of all objects (data copies), which are part of the Semantic Web and $X \times X$ is a set of all ordered pairs of the set X. Then there can be described the interpretation $I = (F, ^I)$ which:

F1. For the concept instances t assigns certain values $t^I \in X$ and for the pair of instances (t_1^I, t_2^I) assigns pairs $(t_1^I, t_2^I) \in X \times X$. Most frequently concept instances are associated with data copies. These copies are considered by IT specialists as objects. Thus, the space $X \cup X \times X$ is a set of Web resources which include documents with considered data. For example specific word in a computer screen is an instance of data copy indicated by the specific Web resource address. Also the relationship between this word and other word is indicated by a pair of Web resources addresses.

F2. For the concept name C assigns fuzzy set $C^I : X \cup X \times X \to [0,1]$, that for any $x, y \in X, C^I(x, y) = C^I((x,y)) = C^I(y)$.

F3. For the role name R assigns a fuzzy set $R^I : X \cup X \times X \to [0,1]$, equals 0 for arguments from X,

F4. For the modifier m assigns a function $m^I : [0,1] \to [0,1]$, where $m^I \in M$,

F5. For formulas φ, including assertions and axioms, assigns some number $\varphi^I \in [0,1]$.

A. Semantic of concepts Tbox

For any $x \in X$, concept names C, D, the role name R and the modifier m:

$$\top^I(x) = 1 \qquad (24)$$

$$\bot^I(x) = 0 \qquad (25)$$

$$(\neg C)^I(x) = (\neg^F C^I)(x) \qquad (26)$$

$$(C \wedge D)^I(x) = (C^I \wedge^F D^I)(x) \qquad (27)$$

$$(C \vee D)^I(x) = (C^I \vee^F D^I)(x) \qquad (28)$$

$$(\exists R.C)^I(x) = \sup_{y \in X} (R^I \wedge^F C^I)(x, y) \qquad (29)$$

$$(\forall R.C)^I(x) = \inf_{y \in X} (\neg^F R^I \vee^F C^I)(x, y) \qquad (30)$$

$$(m(C))^I(x) = m^I(C^I(x)) \qquad (31)$$

B. Semantic of assertions ABox

For any instance t of the concept C and any instances t_1, t_2 of the role R:

$$(t : C)^I = C^I(t^I) \qquad (32)$$

$$((t_1, t_2) : R)^I = R^I(t_1^I, t_2^I) \qquad (33)$$

C. Semantic of axioms

For any concept names C, D and roles R_1, R_2:

$$(C \sqsubseteq D)^I = c^F(C^I, D^I) \qquad (34)$$

$$(R_1 \sqsubseteq R_2)^I = c^F(R_1^I, R_2^I) \qquad (35)$$

$$(C = D)^I = e^F(C^I, D^I) \qquad (36)$$

$$(R_1 = R_2)^I = e^F(R_1^I, R_2^I) \qquad (37)$$

D. Semantic of formulas

For any formulas (assertions and axioms) φ, ϕ and degree $\alpha \in [0,1]$:

$$(\neg\varphi)^I = \varphi^I \to 0 \tag{38}$$

$$(\varphi \wedge \phi)^I = \varphi^I \otimes \phi^I \tag{39}$$

$$(\varphi \vee \phi)^I = \varphi^I \oplus \phi^I \tag{40}$$

$$(\varphi \Rightarrow \phi)^I = \varphi^I \to \phi^I \tag{41}$$

$$(\varphi \Leftrightarrow \phi)^I = \varphi^I \leftrightarrow \phi^I \tag{42}$$

$$(\exists x \varphi(x))^I = \sup\{y \in [0,1] : \text{ exists the instance } t \\ \text{of the concept } T \text{ that } y = (\varphi(t))^I\} \tag{43}$$

$$(\forall x \varphi(x))^I = \inf\{y \in [0,1] : \text{ exists the instance } t \\ \text{of the concept } T \text{ that } y = (\varphi(t))^I\} \tag{44}$$

$$< \varphi, \alpha >^I = \begin{cases} 1, & \varphi^I \geqslant \alpha \\ \varphi^I, & \varphi^I < \alpha \end{cases} \tag{45}$$

When the interpretation function I satisfies the conditions **F1 – F5** and (24)–(45), then it is called the fuzzification of the *IRL* logic language. If after the fuzzification as the result there are only characteristic functions, then this interpretation is called an exact. Then it is equivalent to the standard interpretation of description logic DL [2] and it satisfies the conditions **F1 – F5** and (24)–(37) are satisfied.

VI. BASIC INFORMATION RETRIEVAL LOGIC

The Information Retrieval System can be extended for searching reliable for experts subsets of $X \cup X \times X$, where X is a set of Web resources relating to a chosen field of knowledge. These subsets for a given query indicate reliable for experts Web addresses, representing the searched knowledge. Therefore, these sets can be used to reliable interpret the *IRL* expression.

For this purpose, as in the statistics, is used the confidence range V. It is considered that the most important is that all experts, on the basis of the confidence range V, accept the set of membership degrees of an object to the fuzzy set. This set represent fuzzification of the concept or the role, defined by the knowledge base $K =< Tez, Fuz, V, Ont >$. Tez is a set of concepts and roles from the thesaurus and Ont is a set of concepts and roles from the ontology defined due to the postulate **P2**. Fuz is a set of possible-to-use interpretations (the fuzzification) defined due to postulates **P1** – **P9**, conditions **F1 – F5** and (24) – (45). All *IRL* formulas consist of the set $Tez \cup Ont$ and are interpreted in the **F** algebra by means of the fuzzification set Fuz. The formula φ is true in the knowledge base K, when for any fuzzification $I \in Fuz, \varphi^I = 1$. Since the formulas interpretations are at once the *BL-algebra* interpretations, algebra which is uniquely defined by these interpretations (the **F** algebra), we write $val_{BL}(\varphi) = 1$. Thus, the set of all *IRL* formulas belongs to the class of formulas sets of the fuzzy logic $BL\forall$ interpreted in the *BL-algebra*. These logic were studied by Hàjek [9]. In the $BL\forall$ there are the following inference rules:

1) *Modus ponens*: from φ and $\varphi \to \phi$ infer ϕ;
2) *Substitution rule*: we can substitute any formulas for the propositional variables;
3) *Generalization*: from φ infer $\forall x \varphi(x)$.

Theorem (Soundness and Completeness). Let φ be a formula of the $BL\forall$, T be a set of all formulas from the $BL\forall$-*theory*. Then the following conditions are equivalent (Proof. see [9]):

1) the formula φ is derived from T-theory with use of the inference rules;
2) $val_{BL}(\varphi) = 1$ for each *BL-algebra* (with infinite intersection and infinite union) that is model for T.

The *IRL* is the two-variable fragment of the second-order logic. The values of all predicated variables are concepts and roles. However, the validities of a monadic predicate calculus with identity are decidable. When the formula with predicate variables and roles which are predicates would be removed, then we obtain the monadic predicate calculus. This fragment of the *IRL* is decidable. There are some fragments of the *IRL* of roles, which are decidable and some are known for their use [9], [12].

VII. DEFUZZIFICATION IN *IRL*

In this paper, the defuzzification is identified with the interpretation $< K,^{Def} >$ in the *IRL* logic. This interpretation, for a given query, indicates a reliable-for-experts set of the Web addresses representing knowledge from the knowledge base K.

In this purpose, for the knowledge base $K =< Tez, Fuz, V, Ont >$ and for some fuzzification $I \in Fuz$, any concept C or any role R, have interpretation belonging to the set of the fuzzy confidence range V. It is accepted by all experts and is defined by:

$$V(C) \subseteq \{\alpha : \text{for some instance } t \text{ of the concept } C \\ \text{or } I \in Fuz, \alpha = (t : C)^I\} \tag{46}$$

$$V(R) \subseteq \{\alpha : \text{for some instances } (t_1, t_2) \text{ of the role } R \\ \text{or } I \in Fuz, \alpha = ((t_1, t_2) : R)^I\}. \tag{47}$$

Experts consider knowledge, which is in the fuzzy confidence range, as the exact knowledge and as part of the possible interpretations. The designation of such subsets will be identified with the knowledge defuzzification of the objects, belonging to the X or $X \times X$ [5].

The function $(.)^{Def}$ is called **the defuzzification interpretation** or **the defuzzification of the knowledge base** $K =< Tez, Fuz, V, Ont >$, if following formulas are true. For any concepts C, D, roles R, R_1, R_2, instances of concepts t, t_1, t_2:

$$\perp^{Def} = \oslash, \top^{Def} = X \tag{48}$$

$C^{Def} = X_C, \text{ where}$

   1) $X_C \subseteq X$

   2) $x \in X_C$ iff exists fuzzification $I \in Fuz$, $\quad$ (49)

     such that $(t : C)^I \in V(C)$ and $x = t^{Def}$

$R^{Def} = (X \times X)_R, \text{ where}$

   1) $(X \times X)_R \subseteq X \times X$

   2) $(x,y) \in (X \times X)_R$ iff exists fuzzification $\quad$ (50)

     $I \in Fuz$, such that $((t_1, t_2) : R)^I \in V(R)$

     and $x = t_1^{Def}, y = t_2^{Def}$

$$(\neg C)^{Def} = X \setminus C^{Def} \tag{51}$$

$$(C \vee D)^{Def} = C^{Def} \sqcup D^{Def} \tag{52}$$

$$(C \wedge D)^{Def} = C^{Def} \sqcap D^{Def} \tag{53}$$

$(\exists R.C)^{Def} = \{x \in X : \text{ exists } y \text{ such that}$

$\quad\quad (x,y) \in R^{Def} \text{ and } y \in C^{Def}\}$ $\quad$ (54)

$(\forall R.C)^{Def} = \{x \in X : \text{ exists } y,$

$\quad\quad \text{if } (x,y) \in R^{Def}, \text{ then } y \in C^{Def}\}$ $\quad$ (55)

$$(t : C)^{Def} \text{ iff } t^{Def} \in C^{Def} \tag{56}$$

$$((t_1, t_2) : R)^{Def} \text{ iff } (t_1^{Def}, t_2^{Def}) \in R^{Def} \tag{57}$$

$(C \sqsubseteq D)^{Def} \text{ iff } C^{Def} \sqsubseteq D^{Def} \text{ and } (C = D)^{Def}$

$\quad\quad \text{iff } C^{Def} = D^{Def}$ $\quad$ (58)

$(R_1 \sqsubseteq R_2)^{Def} \text{ iff } R_1^{Def} \sqsubseteq R_2^{Def} \text{ and } (R_1 = R_2)^{Def}$

$\quad\quad \text{iff } R_1^{Def} = R_2^{Def}$ $\quad$ (59)

For any formulas ϕ, φ and degrees $\alpha \in [0,1]$:

$$< \phi, \varphi >^{Def} \text{ iff } \phi^I \geq \alpha \text{ for any } I \in Fuz \tag{60}$$

$$(\neg \phi)^{Def} \text{ iff } (\text{not } \phi^{Def}) \tag{61}$$

$$(\phi \wedge \varphi)^{Def} \text{ iff } (\phi^{Def} \text{ and } \varphi^{Def}) \tag{62}$$

$$(\phi \vee \varphi)^{Def} \text{ iff } (\phi^{Def} \text{ or } \varphi^{Def}) \tag{63}$$

$$(\phi \Rightarrow \varphi)^{Def} \text{ iff } (\text{ if } \phi^{Def} \text{ then } \varphi^{Def}) \tag{64}$$

$$(\phi \Leftrightarrow \varphi)^{Def} \text{ iff } (\phi^{Def} \text{ iff } \varphi^{Def}) \tag{65}$$

$(\exists x \phi(x))^{Def} \text{ iff exists the instance } t$

$\quad\quad \text{of the concept } T \text{ such that } (\phi(t))^{Def}$ $\quad$ (66)

$(\forall x \phi(x))^{Def} \text{ iff for any instance } t$

$\quad\quad \text{of the concept } T : (\phi(t))^{Def}$ $\quad$ (67)

It can be noticed that as a result of the defuzzification of the knowledge base K is some theory *Theory(K)* of sets from the space $X \cup X \times X$. In the process of information retrieval, the knowledge base K should be define in such way, that the defuzzification process of true formula in the base K, is thesis from *Theory(K)*: if for any fuzzification $I \in Fuz$, $\phi^I = 1$, then there is ϕ^{Def}. This knowledge base is called the **adequate knowledge base**. Furthermore, definition of this knowledge base allows to disambiguate fuzzy knowledge which is searched.

VIII. CONCLUSION

The information retrieval system is a kind of **the information retrieval agent** on the Semantic Web, if it meets the postulates **P1 – P9**. Commonly, the information retrieval systems and semantic relation systems have indicated only semantically closest resources. This not is always the case. When searching information about any object you can get the fuzzy knowledge described in the ontology and this uncertainty can lead to the uniquely determined knowledge resource. In this way we obtain the knowledge disambiguation and compliance with the description of the object by the expression used in the thesaurus.

Introduced conceptual apparatus will be used to develop the information retrieval agent, supporting effective and intelligent use of search engines. Likewise, this research can be used to extend the software used in search engines or to create a new type of search engine.

REFERENCES

[1] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi and P. .F. Patel-Schneider eds., *The description logic handbook: theory, implementation, and applications*, Cambridge University Press, 2003.

[2] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison-Wesley, 1999.

[3] F. Bobillo and U. Straccia, *A fuzzy description logic with product t-norm.*, the IEEE International Conference on Fuzzy Systems (Fuzz IEEE-07), pp. 652–657. IEEE Computer Society, 2007.

[4] F. Bobillo and U. Straccia, *Fuzzy description logics with general t-norms and datatypes*, Fuzzy Sets and Systems 160 (23), pp. 3382–3402, 2009.

[5] A. Bryniarska, *The algorithm of knowledge defuzzification in semantic network*, Information Systems Architecture and Technology, Networks Design and Analysis, pp. 23–32, Wrocaw, Poland, 2011.

[6] A. Bryniarska, *An Information Retrieval Agent in Web resources.*, IADIS International Conference Intelligent Systems and Agents 2013, Proceedings, pp. 121–125, Prague, Czech Republic, 2013.

[7] R. Ceglarek and W. Rutkowski, *Automated Acquisition of Semantic Knowledge to Improve Efficiency of Information Retrieval Systems*, In: Business Information Systems 10th International Conference, Proceedings, LNCS, Springer, pp. 329–341, 2006.

[8] R. Ceglarek, K. Haniewicz and W. Rutkowski, *Semantic compression for specialized Information Retrieval systems*, In: Find out how to access preview-only content Advances in Intelligent Information and Database Systems Studies in Computational Intelligence, vol. 283, pp. 111–121, 2010.

[9] M. Cerami, F. Francesc Esteva and F. Bou, *Decidability of a Description Logic over Infinite-Valued Product Logic*, Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010), pp. 203–213, 2010.

[10] P. Hàjek, *Metamathematics of fuzzy logic*, vol. 4 of Trends in LogicStudia Logica Library, Dordrecht: Kluwer Academic Publishers, 1998.

[11] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.

[12] U. Hustadt, R. A. Schmidt and L. Georgieva, *A Survey of Decidable First-Order Fragments and Description Logics*, Journal on Relational Methods in Computer Science, Vol. 1, pp. 251–276, 2004.

[13] S. K. Jayanthi and S. Prema, *Word Sense Disambiguation in Web Content Mining Using Brills Tagger Technique*, In: International Journal of Computer and Electrical Engineering, vol. 3, No. 3, pp. 358–362, 2011.

[14] Ch. D. Manning, P. Raghavan and H. Schtze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[15] M. Sanderson, *Word Sense Disambiguation and Information Retrieval*, In: Proceedings of the 17th International ACM SIGIR, Dublin 1994.

[16] Ch. Stokoe, M. P. Oakesand J. Tait, *Word Sense Disambiguation in Information Retrieval Revisited*, SIGIR 2003, 2003.

[17] U. Straccia, *A fuzzy description logic*, In Proc. of the 15th Nat. Conf. on Artificial Intelligence (AAAI-98), Madison, USA, pp. 594–599, 1998.

[18] U. Straccia, *Reasoning with Fuzzy Description Logics*, In: Journal of Artificial Intelligence Research 14, pp. 137–166, 2001.