FACTA UNIVERSITATIS (NIŠ)
Ser. Math. Inform. Vol. 22, No. 2 (2007), pp. 189–199

INFLUENCE OF NORMS ON THE RESULTS OF QUERY FOR INFORMATION RETRIEVAL MECHANISMS BASED ON THESAURUS

Tanja Sekulić and Petar Hotomski

Abstract. Associated information retrieval mechanisms are defined in order to improve results of information retrieval, and moreover, to enable amplification of search even on the documents which are not in direct connection with conditions imposed by the query. In this paper, influence of different norms on results of the query using thesaurus was especially observed, in a way that T_M norm and supremum appearing in definition of the composition, which obtain the results, were varied with other existing norms and co-norms. Triangular T norms and logical H norm were used. The aim of the research described in this paper, was to show that norms and their co-norms do not have any major influence on the results of the query based on thesaurus and that each of the norms used in the research can be equally successfully applied. Results obtained by the research confirmed the expectations, but nevertheless, minor oscillations in connection with ranking and results arrangement itself were detected, and certain effect on the width of the results was also noted. It can potentially suggest that in this particular case, the width of the results can be affected through various norms as well.

1. Introduction

In information retrieval systems are defined so called associated mechanisms, which enable the system to search throughout additional information, which are not directly connected with terms imposed by the query [?]. Construction of these mechanisms was necessary because it often happened that some of the documents contain information which is, at certain degree, similar, but not identical with terms from the query. Such documents could not be contained in results of the query, until associated mechanisms had appeared. Thesaurus is one example of associated mechanisms which use relationships between terms, and is usually applied in selection of terms which are in a certain way connected with terms from the query. Fuzzy thesaurus is defined in order to express strenght of association between pairs of terms. The usual way of representation of thesaurus is through matrix, so that

Received November 8, 2006.
2000 *Mathematics Subject Classification.* Primary 94D05; Secondary

the element of matrix which is in intersection of one row and one column, expresses the degree of association for two terms, where one term belongs to the named row, and the other to the column.

Current tendencies show that the best results are obtained by constructing the information retrieval mechanisms based on thesaurus, using T_M-norm and supremum. In this research, the T_M-norm and supremum will be varied by other, the most used, T-norms (T_L and T_P, Lukasijevic and a product, respectively), however, the influence of the logical H-norm (especially, phc norm and phd co-norm) will be observed as a particularly interesting case. The H-norm (whose author is P. Hotomski) is different from other T-norms because it is a logical norm, so the results obtained by means of this norm will be compared with the results obtained by means of other T-norms. The T-norms definitions and their dual S co-norms used in the paper are as follows:

1. $T_M(x, y) = \min(x, y)$, $S_M(x, y) = \max(x, y)$,
2. $T_P(x, y) = xy$, $S_P(x, y) = x + y - xy$,
3. $T_L(x, y) = \max(0, x + y - 1)$, $S_L(x, y) = \min(1, x + y)$.

Logical H norm its definition and basic properties will be presented in the next section.

The main goal of this paper was to determine the possible influence of different norms and their dual co-norms on the results of the query, especially on their width and ranking. Expectations are that varying of the norms and their co-norms in mechanisms of information retrieval based on thesaurus is possible to accomplish, and according to that is assumed that the norms will not have major influence on the results of a query.

2. H-Norm

The definition and the characteristics of H-norm appear according to [?].

Definition 2.1. Let $f(z_1, z_2, \ldots, z_k) = 2^{k-1}z_1 + 2^{k-2}z_2 + \cdots + 2^0 z_k$, $z_i \in \{0, 1\} = \mathbb{L}_2$, be one of the operators AND, OR and $x = f(a_1, a_2, \ldots, a_k)$, $y = f(b_1, b_2, \ldots, b_k)$, $x, y \in \mathbb{L}_2^k$,

$$
\begin{aligned}
x * y &\stackrel{\text{def}}{=} f(a_1 * b_1, a_2 * b_2, \ldots, a_k * b_k), \\
\neg x &\stackrel{\text{def}}{=} f(\neg a_1, \neg a_2, \ldots, \neg a_k).
\end{aligned} \tag{2.1}
$$

For these operations $\mathbb{L}_2^k$ is regular Boolean algebra with the first element 0 and the last $(2^k - 1)$. For each propositional formula, $\mathbb{F}(p_1, p_2, \ldots, p_n)$, a term $\mathbb{F}(k_1, k_2, \ldots, k_n)$ is introduced in which $k_i = I_n/(I_{n-i} + 2)$, $I_n = 2^s - 1$, $s = 2^n$, $i = 1, 2, \ldots, n$ and it is shown that the following applies: $\mathbb{F}(p_1, p_2, \ldots, p_n)$ is tautology if and only if $\mathbb{F}(k_1, k_2, \ldots, k_n) = I_n$. Additionally, the following holds for the operations on $\{0, 1, 2, \ldots, I_n\}$:

$$\neg x = I_n - x; \qquad x \vee y = x + y - x \wedge y = x + \neg x \wedge y;$$

$$x \Rightarrow y = \neg x + x \wedge y; \qquad x \Leftrightarrow y = \neg x \wedge \neg y + x \wedge y = \neg(x + y) + 2(x \wedge y),$$

where $+, -$ are ordinary arithmetic operations.

Dividing every element of L_2^s, $s = 2^n$ by I_n leads to the set $\{0, 1/I_n, 2/I_n, \dots, 1\}$ which contains $I_n + 1$ element. By increasing the value of n, sufficiently fine segmentation of $[0, 1]$ may be obtained.

Definition and characteristics of H-norm. Let $\{0, 1/I_n, 2/I_n, \dots, (I_n - 1)/I_n, 1\}$ be a segmentation of the interval $[0, 1]$ where $I_n = 2^s - 1$, $s = 2^n$. Let $x, y \in [0, 1]$ such that $x, y \in \{0, 1/I_n, 2/I_n, \dots, (I_n - 1)/I_n, 1\}$. Then:

$$\mathrm{phc}_n(x, y) \stackrel{\mathrm{def}}{=} (xI_n \wedge yI_n)/I_n \tag{2.2}$$

where $\wedge$ is AND operator on $\{0, 1, 2, \dots, I_n\}$ determined by definition (**??**).

Lemma 2.1. *If there exists n, so that* $x, y \in \{0, 1/I_n, 2/I_n, \dots, (I_n - 1)/I_n, 1\}$, *then* $\mathrm{phc}_n(x, y) = \mathrm{phc}_{n+1}(x, y)$, *i.e., phc is independent of n.*

Lemma 2.2. *There exists a rational number* $x \in [0, 1]$, *for which there is no segmentation, so that* xI_n *is a natural number.*

Consequence. For each segmentation of the interval $[0, 1]$ determined by $\{0, 1/I_n, 2/I_n, \dots, (I_n - 1)/I_n, 1\}$ and $x \in [0, 1]$ it is true that $k/I_n \le x \le (k + 1)/I_n$, i.e., $k \le xI_n \le (k + 1)$, $k \in \{0, 1, 2, \dots, I_n - 1\}$.

The function $\mathrm{round}(xI_n) = k_1$ performs rounding of the arguments respecting the evenness rule, so k_1 is a natural number k or $k + 1$. For $xI_n = k + 0.5$, the value of k_1 is equal either to k (if k is an even number) or to $k + 1$ (if k is an odd number).

Since $x_1 = k_1/I_n = \mathrm{round}(xI_n)/I_n = (xI_n + R)/I_n = x + R/I_n$, $|R| \le 0.5$, $|x_1 - x| \le 1/(2I_n)$, the upper error margin when substituting x by x_1 is $1/(2I_n)$. We name $1/(2I_n)$ *sensitivity coefficient* on an integer k. By proper selection of n, the sensitivity coefficient may be decreased to the required value. For $n = 4$, I_4 is equal to 65535, so that the sensitivity coefficient becomes 0.000008, the value that completely satisfies practical requirements since it provides the accuracy up to five decimals for $x \in [0, 1]$. Now, $\mathrm{phc}(x, y)$ can be defined for any $x, y \in [0, 1]$.

Definition 2.2. H-norm and H-co-norm are defined by the following:

$$\mathrm{phc}(x, y) \stackrel{\mathrm{def}}{=} (\mathrm{round}(xI_n) \wedge \mathrm{round}(yI_n))/I_n,$$

where $\wedge$ is AND operator on the set $\{0, 1, 2, \dots, I_n\}$, $I_n = 2^s - 1$, $s = 2^n$ defined by (**??**), while $\mathrm{round}(xI_n) = x_1 I_n = k_1 \in \{0, 1, 2, \dots, I_n\}$ provides the substitution of argument x by x_1 which is accurate up to $1/(2I_n)$.

$$\mathrm{phd}(x, y) \stackrel{\mathrm{def}}{=} (\mathrm{round}(xI_n) \vee \mathrm{round}(yI_n))/I_n,$$

where $\vee$ is OR operator on the set $0, 1, 2, \dots, I_n$, $I_n = 2^s - 1$, $s = 2^n$ defined by (**??**).

H-norm $\mathrm{phc}(x, y)$ and H co-norm $\mathrm{phd}(x, y)$ define operators AND and OR, whereby all the Boolean properties of these operators are preserved.

The following properties are directly verified for $x \approx x_1$ and $y \approx y_1$ within the limits constituted by the sensitivity coefficient $1/(2I_n)$:

1. $\mathrm{phd}(x, y) = x + y - \mathrm{phc}(x, y)$,
2. $\mathrm{phc}(x, x) = x$,
3. $\mathrm{phc}(0, y) = 0$,
4. $\mathrm{phc}(1, y) = y$,
5. $\mathrm{phc}(x, y)$ is not monotone,
6. $\mathrm{phc}(x, y)$ is extremely sensitive to the change of x, y, i.e., a small modification of arguments significantly changes the value of phc (because of the property of $\wedge$ on the set $\{0, 1, 2, \ldots, I_n\}$).

Other logical operations on $[0, 1]$ can be analogously defined.

Negation is defined as:

$$\neg x \stackrel{\text{def}}{=} \neg \mathrm{round}(xI_n)/I_n = \neg(x_1 I_n)/I_n = (I_n - x_1 I_n)/I_n = 1 - x_1$$

within the limits of the sensitivity coefficient, for $x \approx x_1$ it is true that $\neg x = 1 - x$; $\mathrm{phc}(x, \neg x) = 0$; $\mathrm{phd}(x, \neg x) = 1$. The Boolean H-norm is the natural generalization of the logical conjunction of the set $\{0, 1\}$ to the set $\{0, 1, 2, \ldots, I_n\}$, $I_n = 2^s$, $s = 2^n$, i.e., to $[0, 1]$. The monotonicity is absent here, but the characteristics of AND operator are kept in the regular Boolean algebra.

3. The Basic Properties of Thesaurus and Thesaurus Used Strategies of Search

The definition of thesaurus is based on a concept of relation which connects two similar terms. Such a relation, modeled by thesaurus, can refer to one of the three basic types. Based on [**?**], these types are:

- *The broader term relation* - BT, is used to express that one term has a more general meaning than the other term.
- *The narrower term relation* - NT, is the inverse relation to BT relation.
- *The related term relation* - RT, is defined to exploit synonims or near synonims.

Thesaurus which models these relations is always proposed by a matrix whose elements are values from the set $\{0, 1\}$. According to that, the value contained in intersection of i-th row and j-th column represents the degree in which terms from named row and column are associated with a relation modeled by given thesaurus [**?**]. Especially, when the degrees of association for two terms are from interval $[0, 1]$, then it can be spoken of fuzzy thesaurus.

Assuming that is given a fuzzy thesaurus T, which models relation of association for terms from set $Q = \{q_1, q_2, q_3\}$ with terms from set $P = \{p_1, p_2, p_3, p_4\}$. The fuzzy thesaurus T would be given as follows:

$$T = \begin{array}{c} q_1 \\ q_2 \\ q_3 \\ \\ \end{array} \begin{array}{c} \begin{bmatrix} 0.2 & 1 & 0.6 & 0.1 \\ 0.7 & 0.3 & 0.3 & 0.9 \\ 0.4 & 0.4 & 0.1 & 0.2 \end{bmatrix} \\ \begin{array}{cccc} p_1 & p_2 & p_3 & p_4 \end{array} \end{array} \tag{3.1}$$

From the fuzzy thesaurus represented by matrix (**??**), directly can be established a degree in which each of the pairs (q_i, p_j) are associated with relation modeled by matrix T. For example, it can be read that terms q_2 and p_3 are associated with a relation in degree of 0.3, and that can be directly read from intersection of second row (which belong term q_2) and third column (which belong term p_3). In that way, the fuzzy thesaurus can be defined through a binary fuzzy relation between two finite sets Q and P: $T : Q \times P \to [0, 1]$ and which is represented by matrix T (**??**).

Thesaurus defined as it was shown, can be used as one kind of associated information retrieval mechanism. Namely, these mechanisms are based on composition of conditions imposed by the query, which contains terms relevant for the search, and the fuzzy thesaurus which establish associated relation between terms from the query with terms similar to ones in the documents being searched. Considering that here we have composition of the fuzzy set (query) and a fuzzy relation (fuzzy thesaurus), composition is based on the following definition established by Zadeh [**?**].

Definition 3.1. Let R be fuzzy relation on $X \times Y$ and A fuzzy set on X, then fuzzy subset B of Y can be obtain by composition of R and A, which can be written as follows: $B = A \circ R$. Composition of fuzzy set A and fuzzy relation R is determined by following expression:

$$\mu_B(y) = \sup_x T(\mu_A(x), \mu_R(x, y))$$

where T marks T norm, A the fuzzy set with a membership function $\mu_A(x)$ and R the fuzzy relation with a membership function $\mu_R(x, y)$.

Zadeh proposes sup-min composition, which can be written as follows:

$$\mu_B(y) = \sup_x \min(\mu_A(x), \mu_R(x, y)) \tag{3.2}$$

Other authors concludes that this procedure can be generalized by taking other T norms and T co-norms, but for simpler application for composition of the fuzzy set and fuzzy relation expression (**??**) was finally adopted.

3.1. *Strategy of the search based on fuzzy thesaurus*

Thesaurus represents only one step in procedure of information retrieval. His role is to establish a certain type of relation between the terms, on which base will be evaluated significance, i.e., association of the terms from the query with terms from documents being searched. Strategy of the search based on the fuzzy thesaurus has several steps. First, user places the query, on which base are determined the terms important for conduction of the search. Then the system consults the fuzzy thesaurus, in order to determine a degree of association (or significance) the terms from the query with terms from documents. That is a central place of the searching process where mechanisms defined by norms and co-norms are activated. Based on degree of association obtained as a result of the composition of the query and fuzzy thesaurus, information retrieval system ranks results by degrees of significance, and at the end user gets results of the search in form of the documents ranked by significance in respect to conditions imposed by the query. Figure **??** illustrates one process of the search, as described in the text above.

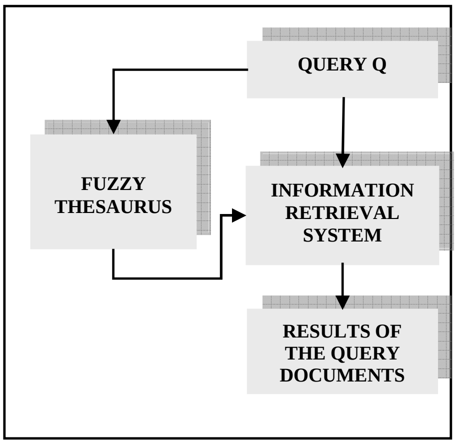

FIG. 3.1: Strategy of the query based on the fuzzy thesaurus

4. Influence of Norms on the Information Retrieval Mechanisms Based on Fuzzy Thesaurus

The following example is entirely acquired from [**?**], and illustrates one mechanism of information retrieval which uses fuzzy thesaurus based on related term relation. In this example, the fuzzy thesaurus is represented by matrix T, which establishes relation of similarity between some musical trends. Let T be following matrix:

$$T = \begin{array}{c} \text{acid} \\ \text{funky} \\ \text{rap} \\ \text{classic} \\ \\ \end{array} \begin{array}{c} \begin{bmatrix} 1 & 0.8 & 0.4 & 0.6 & 0.2 & 0.1 \\ 0.8 & 1 & 0.5 & 0.65 & 0.3 & 0.2 \\ 0.4 & 0.5 & 1 & 0 & 0.2 & 0.1 \\ 0.6 & 0.65 & 0 & 1 & 0.3 & 0.8 \end{bmatrix} \\ \begin{array}{cccccc} \text{acid} & \text{funky} & \text{rap} & \text{classic} & \text{rock} & \text{evergreen} \end{array} \end{array}$$

Now, observe how following query can be solved, using given fuzzy thesaurus:

QUERY: Find musical trends which are similar to acid (degree of matching 1*), funky (degree of matching* 0.5*), rap (degree of matching* 0.3*) and classic (degree of matching* 1*).*

This query, marked with Q, would be presented by the following fuzzy set:

$$Q = \frac{\text{acid}}{1} + \frac{\text{funky}}{0.5} + \frac{\text{rap}}{0.3} + \frac{\text{classic}}{1} = [1 \quad 0.5 \quad 0.3 \quad 1].$$

Results of the query are obtained by applying the composition of the fuzzy set Q and the fuzzy relation imposed by fuzzy thesaurus T. B marks the result, and somewhere is called "built in query", because uses data from thesaurus.

$$\begin{aligned} B = Q \circ T \;&=\; \begin{bmatrix} 1 & 0.5 & 0.3 & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & 0.8 & 0.4 & 0.6 & 0.2 & 0.1 \\ 0.8 & 1 & 0.5 & 0.65 & 0.3 & 0.2 \\ 0.4 & 0.5 & 1 & 0 & 0.2 & 0.1 \\ 0.6 & 0.65 & 0 & 1 & 0.3 & 0.8 \end{bmatrix} \\ &=\; \sup(\min(\mu_Q, \mu_T)) = \begin{array}{c} \begin{bmatrix} 1 & 0.8 & 0.5 & 1 & 0.3 & 0.8 \end{bmatrix} \\ \begin{array}{cccccc} \text{acid} & \text{funky} & \text{rap} & \text{classic} & \text{rock} & \text{evergreen} \end{array} \end{array} \end{aligned}$$

This result can be interpreted in a following way: the degree of similarity for musical trend acid and conditions imposed by the query is 1, for funky 0.8, rap 0.5, classic 1, rock 0.3 and evergreen 0.8. So acid and classic have the largest degree of similarity with conditions imposed by the query, and they are followed by funky, evergreen, rap and rock.

Now, assume that we have another relation R, which is established between musical trends and performers. Let the matrix R represents a relation R:

$$R = \begin{array}{c} \text{acid} \\ \text{funky} \\ \text{rap} \\ \text{classic} \\ \text{rock} \\ \text{evergreen} \\ \end{array} \begin{array}{c} \begin{bmatrix} 0.9 & 0.8 & 1 & 1 & 1 & 0.1 & 0.1 & 0.1 & 0.2 & 0.7 \\ 0.8 & 0.8 & 0.7 & 0.9 & 0.9 & 0.6 & 0.4 & 0.6 & 0.3 & 0.7 \\ 0.9 & 0.2 & 0.7 & 0.9 & 0.9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 & 0 & 0.9 & 0.3 & 0.4 & 1 & 1 \\ 0 & 0.3 & 0 & 0 & 0 & 0.2 & 1 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.7 & 0.3 & 0.3 & 1 & 1 \end{bmatrix} \\ \begin{array}{cccccccccc} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} \end{array} \end{array}$$

where $y_1, y_2, \ldots, y_{10}$ marks the performers.

Using the composition of the relations B and R, fuzzy set D can be obtained, which contains performers connected with wanted musical trends, also considering degrees of association. The result of this composition is:

$$D = B \circ R = \begin{array}{c} \begin{bmatrix} 0.9 & 0.8 & 1 & 1 & 1 & 0.9 & 0.4 & 0.6 & 1 & 1 \end{bmatrix} \\ \begin{array}{cccccccccc} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} \end{array} \end{array}$$

In that way, following results are obtained:

$$D = \frac{0.9}{y_1} + \frac{0.8}{y_2} + \frac{1}{y_3} + \frac{1}{y_4} + \frac{1}{y_5} + \frac{0.9}{y_6} + \frac{0.4}{y_7} + \frac{0.6}{y_8} + \frac{1}{y_9} + \frac{1}{y_{10}},$$

which means that degree of similarity of performer y_1 with the query Q is 0.9, performer y_2 0.8, ...

4.1. *Varying of norms on information retrieval mechanisms based on fuzzy thesaurus*

Varying of norms on information retrieval mechanisms based on fuzzy thesaurus is conducted in a way that every appearance of supremum and minimum in the composition was replaced with other norms and co-norms. More precisely, supremum was replaced with S_L, S_P and phd co-norms, and minimum with T_L, T_P and phc norms. Replacement of the norms was realized on previous presented example.

T_L norm: Result of the query Q was marked with B, and is obtained by following way for T_L norm:

$$B = Q \circ T = S_L(T_L(\mu_Q, \mu_T)) = \begin{array}{c} \begin{bmatrix} 1 & 1 & 0.7 & 1 & 0.5 & 0.9 \end{bmatrix} \\ \begin{array}{cccccc} \text{acid} & \text{funky} & \text{rap} & \text{classic} & \text{rock} & \text{evergreen} \end{array} \end{array}$$

In order to establish connection between musical trend and performer, is necessary to make a composition of vector B and matrix R, which represents connection between performers and musical trends:

$$D = B \circ R = S_L(T_L(\mu_R, \mu_B)) = \begin{array}{c} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \\ \begin{array}{cccccccccc} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} \end{array} \end{array}$$

In that way, the following results are obtained:

$$D = \frac{1}{y_1} + \frac{1}{y_2} + \frac{1}{y_3} + \frac{1}{y_4} + \frac{1}{y_5} + \frac{1}{y_6} + \frac{1}{y_7} + \frac{1}{y_8} + \frac{1}{y_9} + \frac{1}{y_{10}}.$$

T_P norm: Result obtained by application of T_P norm was as follows:

$$\begin{aligned} B = Q \circ T &= S_P(T_P(\mu_Q, \mu_T)) \\ &= \begin{array}{c} \begin{bmatrix} 1 & 0.97 & 0.68 & 1 & 0.55 & 0.84 \end{bmatrix} \\ \begin{array}{cccccc} \text{acid} & \text{funky} & \text{rap} & \text{classic} & \text{rock} & \text{evergreen} \end{array} \end{array} \end{aligned}$$

By composition of vector B and matrix R connection between performers and musical trends is established:

$$\begin{aligned} D = B \circ R &= S_P(T_P(\mu_R, \mu_B)) \\ &= \begin{array}{c} \begin{bmatrix} 0.99 & 0.98 & 1 & 1 & 1 & 0.99 & 0.87 & 0.9 & 1 & 1 \end{bmatrix} \\ \begin{array}{cccccccccc} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} \end{array} \end{array} \end{aligned}$$

Final result:

$$D = \frac{0.99}{y_1} + \frac{0.98}{y_2} + \frac{1}{y_3} + \frac{1}{y_4} + \frac{1}{y_5} + \frac{0.99}{y_6} + \frac{0.87}{y_7} + \frac{0.9}{y_8} + \frac{1}{y_9} + \frac{1}{y_{10}}.$$

H-norm: Result obtained by application of H norm was as follows:

$$\begin{aligned} B = Q \circ T &= \operatorname{phd}(\operatorname{phc}(\mu_Q, \mu_T)) \\ &= \begin{array}{cccccc} [\ 1 & 0.93 & 0.93 & 1 & 0.5 & 0.87\ ] \\ \text{acid} & \text{funky} & \text{rap} & \text{classic} & \text{rock} & \text{evergreen} \end{array} \end{aligned}$$

By composition of vector B and matrix R connection between performers and musical trends is established:

$$\begin{aligned} D = B \circ R &= \operatorname{phd}(\operatorname{phc}(\mu_R, \mu_B)) \\ &= \begin{array}{cccccccccc} [\ 0.93 & 0.93 & 1 & 1 & 1 & 1 & 0.8 & 1 & 1 & 1\ ] \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} \end{array} \end{aligned}$$

Final result:

$$D = \frac{0.93}{y_1} + \frac{0.93}{y_2} + \frac{1}{y_3} + \frac{1}{y_4} + \frac{1}{y_5} + \frac{1}{y_6} + \frac{0.8}{y_7} + \frac{1}{y_8} + \frac{1}{y_9} + \frac{1}{y_{10}}.$$

In order to make the comparation of the results obtained by all norms for queries B and D easier, all results are given in the following tables.

Table 4.1: Results obtained by all norms for query B

	T_M and S_M	T_L and S_L	T_P and S_P	Phc and phd
Q	1	1	1	1
u	0.8	1	0.97	0.93
e	0.5	0.7	0.68	0.93
r	1	1	1	1
y	0.3	0.5	0.55	0.5
B	0.8	0.9	0.84	0.87

When presented in this way, it is obvious that norms have some minor influence on the results of the queries imposed in this example. Moreover, it is possible to determine that some oscilations in width of the results are present. The results obtained by T_L norm were the widest, and they are followed by results obtained by T_P, H and finally, T_M norm. It could be used in the future in order to improve information retrieval techniques and to enable possible control on the width of the results presented to the user.

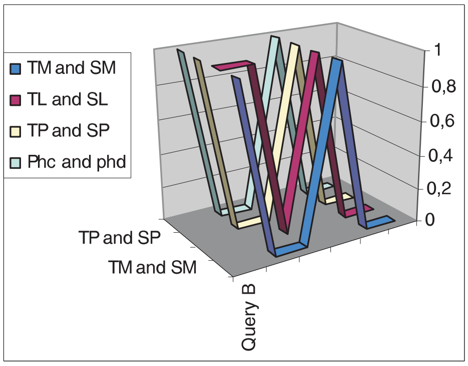

FIG. 4.1: Results obtained by all norms for query B

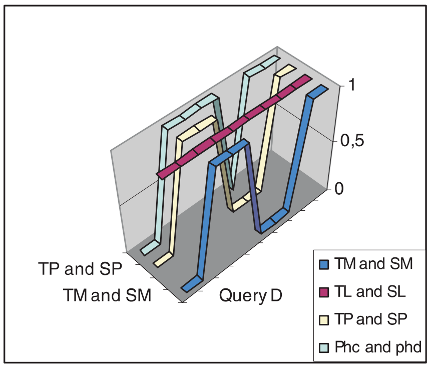

FIG. 4.2: Results obtained by all norms for query D

Table 4.2: Results obtained by all norms for query D

	T_M and S_M	T_L and S_L	T_P and S_P	Phc and phd
	0.9	1	0.99	0.93
Q	0.8	1	0.98	0.93
u	1	1	1	1
e	1	1	1	1
r	1	1	1	1
y	0.9	1	0.99	1
	0.4	1	0.87	0.8
D	0.6	1	0.9	1
	1	1	1	1
	1	1	1	1

5. Conclusion

Observing fuzzy thesaurus as one of information retrieval associated mechanisms, imposes the conclusion that it is possible to replace adopted T_M norm and supremum in the composition which realizes search based on named mechanisms with others triangular norms, and especially, with logical H norm. Even so, the results of the query did not undergo major changes. However, some minor oscilations of membership function values for several results have been established on the presented examples. By comparing results for each norm separately, it has been confirmed that the widest results are given by T_L norm followed by H norm and T_P norm, and at the end came T_M norm which gave the narrowest results. From the aspect of information retrieval mechanisms based on fuzzy thesaurus, appears the question of possible control on scale of query results, using different norms. Thus, the user can select by himself the width of the results which would be presented to him. Future research, which would include even more examples in exploring the influence of norms on information retrieval mechanisms based on fuzzy thesaurus, will give some answers to asked questions, and so it will improve existing information retrieval techniques.

REFERENCES

1. G. Bordogna: *Modeling vagueness in IR, Part 2*, Instituto per le Tecnologie Informatiche Multimediali, Milano; `www.citeseer.ist.psu.edu`
2. F. Crestani and G.Pasi, *Soft information retrieval: Applications of fuzzy set theory and neural networks*; `www.cis.strath.ac.uk/~fabioc/papers/99-fuzzy-book.pdf`
3. B. Nikolić and P. Hotomski: *Comparison of H-logical norm with some T-norms*, Informatica Vol. **14**, No 4, Institute of Mathematics and Informatics Lithuanian Academy of Sciences, Vilnius 2003, 515–528.6.
4. P. Subašić: *Fuzzy Logic and Neural Networks*, Tehnička knjiga, Belgrade, 1997 (in Serbian).
5. T. Sekulić : *Norms of the Comparators in Fuzzy Relational Databases*, Master's Thesis, University of Novi Sad, Technical Faculty "Mihajlo Pupin", Zrenjanin, 2005 (in Serbian).
6. T. Sekulić and P. Hotomski, *Influence of norms on the results of query in fuzzy relational databases*, IMS'2006, 5th Intenational Symposium On Intelligent Manufacturing Systems "Agents and Virtual Worlds", Sakarya, Turkey, May 29–31, 2006.
7. P. Subašić and A. Huettner, *Affect analyzis of text using fuzzy semantic typing*, FUZZ-IEEE 2000, San Antonio, May 2000.
8. D. Tikk, J.D. Yang, P. Baranyi and A. Szakal: *Fuzzy relational Thesauri in information retrieval: Automatic knowledge base expansion by mean of classified textual data*; `www.citeseer.ist.psu.edu`

Tanja Sekulić
Technical College
Djordja Stratimirovića 23
23000 Zrenjanin, Serbia
`tsekulic@sbb.co.yu`

Petar Hotomski
University of Novi Sad
Tehnical Faculty "Mihajlo Pupin"
Djure Djakovića bb
23000 Zrenjanin, Serbia
Phone: ++38123550507
Fax: ++38123550520
`hotomski@ptt.yu`