



International Conference on
Intellectual Capital Management
ICIM2009




INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
D. A. Sadovnik, 2009
Dushanbe, Tajik

Mathematical models of Interlanguage normalization of some groups of Tajik words

Technological University of Tajikistan
pulatova.shoira@gmail.com




International Conference on
Intellectual Capital Management
ICIM2009




INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
D. A. Sadovnik, 2009
Dushanbe, Tajik

Peculiarities of Tajik language

- The translation of some Tajik words according to a scheme
word of input language ->
word of output language
are not always possible



International Conference on
Intellectual Capital Management
ICIM2009




INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
D. A. Sadovnik, 2009
Dushanbe, Tajik


Group of words with the complex structure

Examples:

- Barodarahoyatonandu**
- Khonangihoyaton**
- Dodarakamon**




International Conference on
Intellectual Capital Management
ICIM2009




INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
P.O. Box 68090
DUBAI, UAE

Simplification of Tajik words with the complex structure

- Therefore it is sometimes reasonable to replace such words by their equivalent expressions, thereby minimizing the number of postfixes attached to the words in the derived fragment of sentence.



International Conference on
Intellectual Capital Management
ICIM2009




INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
P.O. Box 68090
DUBAI, UAE

Simplification of words with the complex structure

- This process is named INTRALINGUISTIC (INTRALANGUAGE) NORMALISATION of word.




International Conference on
Intellectual Capital Management
ICIM2009




INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
P.O. Box 68090
DUBAI, UAE

Admissible group of postfixes for Intralinguistic normalization of words

- In Tajik language there are six groups of postfixes, which allow to perform process of words Intralinguistic normalization




International Conference on
Intellectual Capital Management
IICM2009




INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
P.O. Box 646, 41014
Tehran, Iran

Translation of words

- The process of translation words from one language to another is called INTERLANGUAGE NORMALIZATION




International Conference on
Intellectual Capital Management
IICM2009




INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
P.O. Box 646, 41014
Tehran, Iran

The automatic (computer) translation of words into another language

- The automatic translation of the Tajik words into another language, are preceded by phases, which are executed sequentially:
 - Morphological analysis of a translated word;
 - Intralinguistic normalization of the word;
 - Postediting of the normalized expression;



International Conference on
Intellectual Capital Management
IICM2009





INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
P.O. Box 646, 41014
Tehran, Iran

Mathematical models of Intra/Interlanguage normalization of Tajik words

$$S = O_1 \oplus \delta(q_1) \oplus \delta(m_i) \oplus$$



$$\delta(n) \oplus \delta(q_2) \oplus \delta(m_k) \oplus$$

$$\delta(l_j) \oplus \delta(q_3)$$


Mathematical models of Intra/Interlanguage normalization of Tajik words

$$N(S) = \delta_1(l_j, M_j) \oplus \delta_1(n, \partial ap) \oplus O_1 \oplus \delta(q_1) \oplus \delta_1(m_i, u - M_i) \oplus \delta_1(n, _ \text{by} \partial a z u) \oplus \delta(q_2) \oplus \delta_1(m_k, u - M_k) \oplus \delta_1(l_j, _ \text{me} \partial ou) \oplus l_j \oplus \delta(q_3)$$


Mathematical models of Intra/Interlanguage normalization of Tajik words

$$E(R) = \delta_1(y_j, y_j') \oplus w_{y_j} \oplus \delta_1(d, d') \oplus \delta_1(u, u') \oplus w_{y_j} \oplus \delta_1(v, v') \oplus \delta_1(h_i, h_i') \oplus \text{the most (least)} \oplus O_1' \oplus \delta(q_1') \oplus \delta(q_2') \oplus \delta(q_3') \oplus \delta_1(v, v') \oplus \delta_1(g_i, g_i')$$





Intra/Interlanguage normalization of Tajik words

Examples:
S = Barodardhoyatonandu
M(S) = Barodar|ho|yaton|and|u
N(S) = Onho barodardhoi Shumo meboshand va
E(S) = They are your brothers and




International Conference on
Intellectual Capital Management
ICIM2009




INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
P. O. Box 6800
Dushanbe, TADJIKISTAN

Intra/Interlanguage normalization of Tajik words

Example:
S = honangiaton
M(S) = hona|angi|aton|
N(S) = dar honai Shumo budagi
E(S) = which in your home



International Conference on
Intellectual Capital Management
ICIM2009



INSTITUTE FOR ADVANCED
STUDIES IN BASIC SCIENCES
P. O. Box 6800
Dushanbe, TADJIKISTAN

Thanks for your attention
