

*InfoStream Translate.*

**Abstract.** *In article modern methods of creation of statistical translators are described. As an example in article stages of creation of system of statistical machine translation InfoStream Translate are presented. Problems which authors have faced at creation of system, and also methods of their decision are described.*

**Keywords:** *SMT, Stream News Translator, Russian-Ukrainian Translator, Ukrainian-Russian Translator.*

Google (http://translate.google.com) 57-  
 Google Bing Translator  
 (http://www.microsofttranslator.com/), Microsoft, 20-  
 (Hutchins  
 2005, 2007).

(  
 ,  
 ,  
 .  
 ,  
 ,  
 ,  
 :

$$p(e|f) = \frac{p(e)p(f|e)}{p(f)}, \tag{1}$$

$f$ -  
 ,  
 ,  $p(f|e)$  -  
 ,  
 ,  
 :

$$\arg \max_e p(e|f) = \arg \max_e p(e)p(f|e). \tag{2}$$

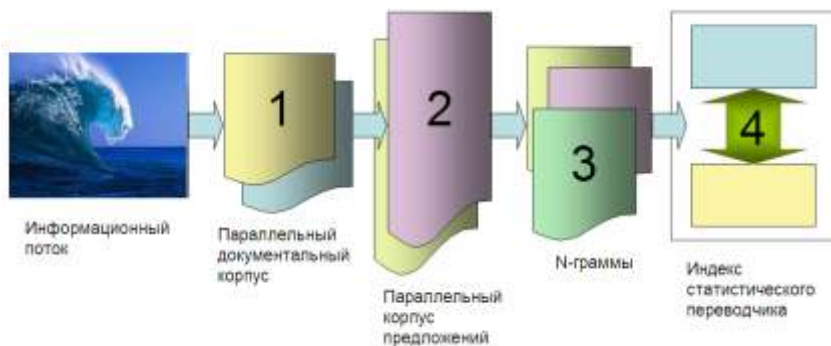
$p(f)$ ,

$p(e)$ ,  $p(f|e)$ ,  $p(e)$ ,  $p(e)$ ,

( . 1):

N-

N-



. 1.

11

2008),

N- . N-

N « 3-

- « »;

- « »;

- « »;

- « ».

N- N-

( ) LDC. 5-

5 DVD- N- Google , Bing.

N- ( ).  
 N-  
 ) N-  
 : a b c d e f g h,  
 )  
 :  
 :  
 ${}_{1eng}(abcde) \sim {}_{1rus}(\quad), {}_{2eng}(bcdef) \sim {}_{2rus}(\quad),$   
 ${}_{3eng}(cdefg) \sim {}_{3rus}(\quad), \dots$   
 N-  
 N- N-  
 N-  
 N-  
 N-  
 N-  
 N-  
 (N-1)-  
 N-  
 ,  
 ,  
 “ ” ( . ).  
 N-  
 Google Translate, Bing Translator  
 IBM,  
 n.Fluent

, ( -  
 ),  
 . 1. -

	<b>Bing Translator</b>	<b>BabelFish</b>
Businesses that hire students can save 8 percent by using a foreign worker over an American because they don't have to pay Medicare, Social Security and unemployment taxes.	8	8 don't Medicare,
	<b>Google Translate</b>	<b>PROMT</b>
	8	8

(SMT),  
 (Rule-Based MT),  
 PROMT (<http://translate.ru/>) BabelFish  
 (<http://babelfish.yahoo.com/>),  
 )

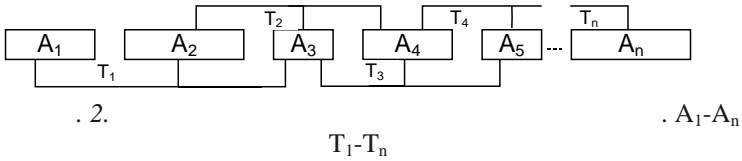
InfoStream

InfoStream 15

InfoStream, -  
 , -  
 , -  
 - ( 2009). -  
 - InfoStream. -  
 . -  
 , -  
 Okapi BM25 -  
 . -  
 . -  
 , -  
 . -  
 : -  
 - 10%; -  
 - , 3; -  
 - . . . -  
 , -  
 . -  
 - -  
 - -

- :  
 - ( (.,!,?;,);  
 ).  
 .  
 .  
 . — .  
 . ,  
 . ,  
 . : , , , . . .  
 « » » , ,  
 . ,  
 .  
 . ,  
 .  
 . ( . 2).  
 N- ,  
 ,  
 N- ,  
 N- N- ,  
 N- N- .  
 InfoStream  
 ,  
 , ,





crc32,

4-

Google Bing Translator 5-

( 2). 2

InfoStream.

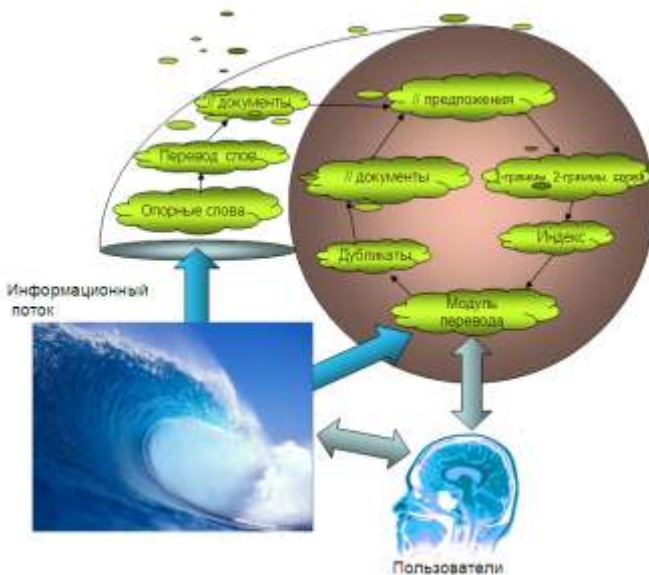
InfoStream

Google Translate.

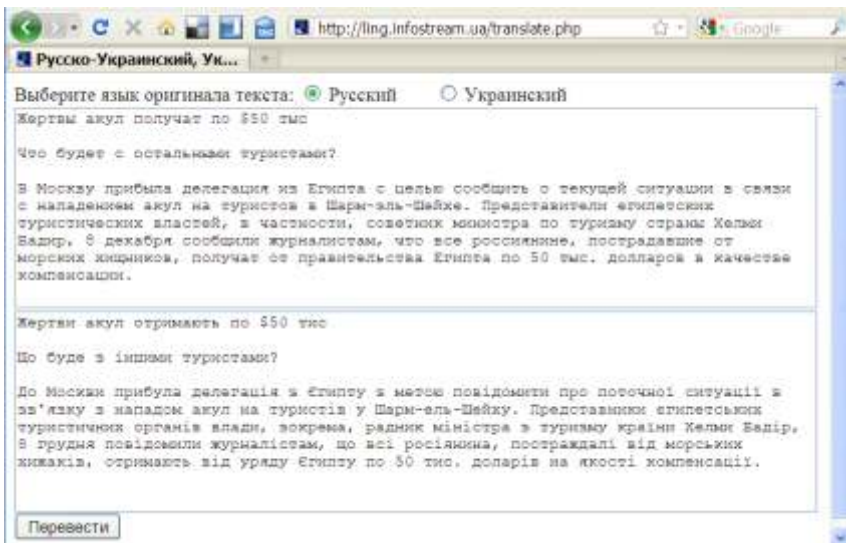


« »  
 ( , ) -  
 , -  
 ( .3). -  
 . -  
 , -  
 , -  
 3-,  
 2- ,  
 , -  
 ( ),  
 InfoStream

<http://ling.infostream.ua/translate.php> ( .4).



. 3.



. 4. - - InfoStream

( , ) - -

,

:

1. ( -

2. ).

3. 10-12 . ,

4. , -

-

« » , -

,

,

1. , : . -

2. , . -  
- , - - - - -

3. , , -

4. , , -

5. PR- , - . -

6. (Deep Web) ). -  
- - - - -

- 
1. // . 53 – . 22–29. . – ., 2009. –
  2. . . . // . – 2009”. – . 8 (15). – . : ., 2009. – C. 278–283.
  3. . . . // . . 156–163. . – ., 2009. – . 53. –
  4. . . . // . : . . . . “ – 2008”. – . : ., 2008. – . 7 (14). – . 431–436.
  5. *Hutchins W. J.* Current commercial machine translation systems and computer-based translation tools : system types and their uses // *International Journal of Translation* vol. 17, No 1–2, 2005, pp. 5–38.
  6. *Hutchins W. J.* Machine translation : a concise history // To be published in *Computer aided translation : Theory and practice*, ed. Chan Sin Wai. Chinese University of Hong Kong, 2007.
  7. *Uszkoreit J., Ponte J.M. Ashok, Popat A.C., Dubiner M.* Large Scale Parallel Document Mining for Machine Translation // *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1101–1109, Beijing, August 2010.
  8. <http://arxiv.org/abs/0807.0311>
  9. <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13>
  10. <http://googleresearch.blogspot.com/2006/08/all-our-n-gram-are-belong-to-you.html>
  11. <http://research.microsoft.com/en-us/collaboration/focus/cs/web-gram.aspx>