В. О. Сахарчук

Теоретичні засади оцінювання якості машинного перекладу та машинного реферування

Незважаючи на те, що зараз існують деякі загальноприйняті особливості процесу оцінювання машинного перекладу, можна сказати, що універсальним не вважається жоден з них.

Як і в інших оцінюваних галузях лінгвістики, у машинному перекладі (далі МП) визначають 3 типи оцінювання (J. S. White):

– оцінки адекватності, яка призначена для виявлення придатності системи МП у зазначеному оперативному контексті;

– оцінки, що слугує для визначення обмежень, помилок і недоліків, які можуть бути виправлені або вдосконалені;

– оцінки ефективності, яка призначена для оцінювання етапів розвитку систем або різних технічних реалізацій.

Зазвичай оцінка МП включає в себе можливості, відсутні в інших системах оцінювання результатів обробки даних природної мові:

– якість невідредагованого перекладу;

– зрозумілість, точність, доречність стилю;

– зручність використання засобів для створення та оновлення словників, для постредагування текстів і для контролю вхідної мови та для налаштування документів тощо;

– розширюваність системи для нових мовних пар і (або) нових предметних областей;

– витрати і результати порівняння з продуктивністю людського перекладу.

У своєму практичному дослідженні ми використовуватимемо методику оцінки, яка ґрунтується на порівнянні кількості помилок, зроблених машиною-перекладачем, з кількістю помилок в еталоні – перекладі, зробленому людиною (Y. Wilks).

Матеріалом перекладацької частини дослідження слугуватимуть наукові тексти з лінгвістики, перекладені системами МП Google Translate та системою Promt у напрямку англійська – російська мови, рефератної – результати роботи систем автоматичного реферування вищезгаданих текстів: «Автореферат» Microsoft Word 2003, Intelexer Summarizer 2.6, TextAnalist 2.01.

Для визначення обсягу вибірки в описуваному експерименті візьмемо Exel-формулу :

$N = Z\text{^}2 / \delta\text{^}2 * P$,

де Z – табличний, теоретичний коефіцієнт, величина якого залежить від числа вибірок і заданої надійності (імовірності) визначення помилки (у більшості випадків у лінгвостатистиці можна брати коефіцієнт 1,96≈2); N –вибірка (кількість словоформ у перекладених текстах); P – відносна частота словоформи, яка посідає середню позицію в частотному списку словоформ обробленого тексту; δ – відносна помилка вимірювання ($\delta = Z / (N * P)^{1/2}$).

При оцінці автоматичного реферування нами було використано дещо відмінні методи.

Існує можливість установлення деяких загальних принципів побудови методів оцінки результатів автоматичного реферування (С. А. Тревгода), таких як:

1) установлення формального ознакового простору для характеристики потрібних властивостей рефератів;

2) установлення еталонних значень запропонованих ознак;

3) установлення властивостей даного реферату шляхом порівняння значень його ознак з еталонними значеннями;

4) сумарна оцінка якості реферату на підставі комплексного показника якості.

Нехай число результівних лінгвістичних елементів, видаваних даною породжуючою моделлю дорівнює N, а число елементів, визнаних експериментатором релевантними, дорівнює R. Тоді повноту C(completness) функціонування моделі можна подати як відношення числа релевантних елементів до загальної кількості всіх виданих елементів (а не всіх потенційно можливих елементів!), виражене у відсотках:

$C = 100 * R / N$ (1).

Точність P (precision) оцінюваної моделі у свою чергу можна подати як відношення числа $(N – R)$ нерелевантних елементів до загальної кількості всіх виданих елементів, виражене у відсотках:

$P = 100 * (N – R) / N$ (2).

Звернімо увагу, що формули (1) і (2) застосовні також для оцінки якості функціонування породжуючих моделей, алгоритмів

аналізу й синтезу текстів, автоматичних ІПС, включаючи й так звані «пошукові машини» Інтернету.

Оцінка відіграє важливу роль при створенні конкурентного середовища для розвитку комп'ютерної лінгвістики і, таким чином, є важливим засобом сприяння прогресу досліджень у цій галузі (Р. М. Алігулієв).