

**ПЛОВДИВСКИ УНИВЕРСИТЕТ
„ПАИСИЙ ХИЛЕНДАРСКИ“**

Факултет по Математика и Информатика
Специалност „Информатика“
Задочно обучение

Курсова работа по „Компютърна лингвистика“

Тема:

25. Машинен превод

Изготвил:

Емил Григоров Медаров

Ф№:

2001262013

Увод

В епохата на бързо развиващите се технологии и глобализацията, машинният превод изпълнява ключова роля в облекчаването на комуникацията между хората, използващи различни езици. Важността на този вид технология не може да бъде подценена, тъй като тя улеснява обмена на информация в различни сфери на обществото, включително бизнес, наука, право и много други.

С разрастването на световната търговия и мобилността на хората, необходимостта от ефективен и бърз превод на информация става належаща. Машинният превод предоставя решение за този предизвикателен проблем, като съкращава времето и усилията, които биха били изразходени при традиционните методи на превод.

В този контекст, машинният превод играе важна роля в работата на много професионалисти, като юристи, бизнесмените, учени и дори обикновени потребители. За юристите и правителствените служители, този вид технология предоставя възможност за бърз и точен превод на юридически документи. В сферата на бизнеса, машинният превод е от съществено значение за глобалните компании, които имат нужда от бърз и ефективен обмен на информация с партньори и клиенти от различни култури и езици.

В тази курсова работа ще проучим не само техническите аспекти на машинния превод, но и как той оказва влияние върху различни области на обществото, подчертавайки необходимостта от постоянно развитие и подобрене на тази технология.

1.Какво е „Машинен превод“?

Машинният превод представлява съществена дисциплина в областта на компютърната лингвистика, занимаваща се с автоматизирания процес на превеждане на текстове или реч от един естествен език на друг, използвайки компютърен софтуер. В своята основа, този метод обикновено включва заместване на думи от един език с думи от друг, като използва текстови корпуси и сложни техники за отчитане на различията между езиците, фразово разпознаване, превеждане на идиоми и разпознаване на аномалии.

Съвременният софтуер за машинен превод предоставя възможност за персонализация, позволявайки избор на конкретни области като право или метеорология. Това ограничаване на значението на думите има за цел подобряване на качеството на превода, като се насочва към стандартизирани и шаблонни езикови конструкции. Например, в случаите на превод на правителствени или правни документи, машинният превод често достига по-сполучливи резултати в сравнение със свободния разговор или текстове с по-ниска степен на стандартизация.

Въпреки че човешката намеса може да подобри качеството на машинния превод, съвременните системи все още не успяват да конкурират с човешките преводачи, особено при текстове с художествен или разговорен стил. В контекста на тази курсова работа, ще проучим различните аспекти на машинния превод, включително технологичните предизвикателства и възможностите за подобрене на резултатите, както и разграничаването между машинния превод и компютърно подпомагания превод, където човекът запазва водещата роля.

Съществуват различни форми на машинен превод, включително тези, които се основават на правила, статистика или невронни мрежи, като се използват в зависимост от конкретните изисквания. Технологията на машинния превод продължава да напредва и да се усъвършенства непрекъснато, като съществуват програми, които се доближават до изключително висока степен на съвършенство.

Освен това, тези софтуерни решения не само спестяват време, но и предоставят икономическа изгода.

Растящата необходимост от бързо и ефективно превеждане на разнообразно съдържание води до широкото разпространение на

машинния превод, включително приложения като Google Translate и Amazon Translate.

Това е ключовата причина, поради която множество доставчици и клиенти предпочитат тези технологии.

Машинният превод предоставя възможност за незабавно превеждане на текст в различни езикови комбинации и тематики, като същевременно се спестяват ресурси и време. Въпреки това, резултатът може да включва пропуски и грешки, които могат да представляват препятствие за постигането на крайната цел. В този контекст, агенциите за превод изпълняват важна роля.

Освен използването на специализиран софтуер за превод, който позволява създаването на терминологични бази данни в различни езици и за различни бизнес сфери, важно е да се разбират новите технологии, които позволяват интегрирането на машинния превод в преводаческия софтуер.

2.История на машинния превод

Историята на машинния превод започва през XVII век, когато философи като Лайбниц и Декарт предполагат съществуването на код, свързващ думите от различни езици. Въпреки хипотетичния характер на всички тези предположения, до средата на 30-те години на XX век се подават първите патентни заявления за "преводаческа машина". Френският изобретател Ж. Арцруни представя едно от тях, създавайки автоматичен двуезичен речник на перфолента. Друго заявление се подава от съветския учен Петър Троянски, който представя по-подробен модел, включващ двуезичен речник и система за работа с граматични роли между два езика, базирана на Есперанто. Въпреки че тази система включва три етапа, схемата ѝ остава неизвестна до средата на 50-те години, когато се появява електронноизчислителната машина.

Истинското развитие на машинния превод обаче започва през 50-те години на 20 век след Втората световна война. На 7-ми януари 1954 г. експериментът, проведен от университета "Джорджтаун" и фирмата "Ай Би Ем", успешно демонстрира превод на няколко десетки изречения от руски на английски. Този успех стимулира сериозно финансиране на изследвания в областта на машинния превод, като авторите му предвиждат решението на задачата в срок от три до пет години.

Системата от ранните периоди на машинния превод се отличавала значително от съвременните детски играчки, като използваше ограничен речник от 250 думи и правила за превод от руски на английски за 49 предварително избрани изречения с тематика в областта на химията. Въпреки този ограничен обхват, демонстрацията подтикна разпространението на идеята за неизбежното развитие на машинния превод, като привлече вниманието на финансиращи институции в САЩ и по света.

Въпреки тази оптимистична прогноза, реалният напредък е по-забавен.

Ранните системи за машинен превод използваха големи двуезични речници и ръчно програмирани правила, за да подредят думите в правилния ред. Въпреки че този метод се счита за ограничен, развитието на лингвистични науки като генеративна лингвистика и трансформационна граматика беше насочено към подобряване на качеството на превода. В същото време възникнаха операционни системи, като BBC на САЩ и комисията по атомна енергетика в САЩ и ЕВРОАТОМ в Италия използваха различни системи, разработени от различни университети. Въпреки ниското качество на превода, тези системи станаха популярни сред потребителите поради по-голямата им скорост.

Към края на 1950-те години Йегошуа Бар-Хилел, по искане на Американското правителство, демонстрира, че "Напълно Автоматичен Висококачествен Превод" от машините не е възможен. Доказателството се основаваше на двусмислеността в езика, като пример се предоставяше изречение, в което думата "rep" може да има две различни значения. В наши дни, такива семантични неясноти могат да бъдат решени чрез използване на контролиран език при написването на изходния текст за машинен превод, включващ думи с ясно дефинирани значения.

През 1966 г. Специалният комитет по приложна лингвистика към Националната академия на науките на САЩ издава доклад, известен като "Черната книга на машинния превод", който изразява съмнения относно възможността за сериозен напредък. Този доклад довежда до драстично намаляване на финансирането на изследванията в тази област в САЩ. Впрочем, в Канада, Франция и Германия се продължава с разработките.

През 1970 г. BBC на САЩ внедряват системата Systran, последвана през 1976 г. от Европейската комисия. През 1977 г. в Канада се внедрява системата МЕТЕО, разработена от университета в Монреал и работи и до днес, превеждайки около 80 000 думи на ден или 30 милиона думи на

година. В Европа от 1970 до 1994 г. проектът Eurotra за превод между езиците на Европейската общност се финансира от Европейската комисия, въпреки че не успява да достигне до практически използваем продукт. С развитието на технологиите през 80-те години, интересът се насочва към статистическите модели и превода чрез примери. В края на 80-те години нараства изчислителната мощност и компютрите стават по-достъпни, като се наблюдава бурен разцвет и поява на множество комерсиални софтуерни продукти в областта на машинния превод. Сред българските програми са Translat, Bultra и WebTrance.

3.Качество на машинния превод

Изготвени са разнообразни методи за оценка на функционалността на системите за машинен превод. Начинът, който включва участието на хора, като преводачи, за определяне на качеството на превода, представлява най-старата методика. Сред по-съвременните автоматизирани методи за оценка се включват BLEU, NIST и METEOR.

Съвременните системи за машинен превод не са в състояние да се състезават с човешките преводачи по отношение на качеството на превода, особено когато става въпрос за текстове с художествен или разговорен характер. Обикновено, преведените от тях текстове предоставят повърхностно познание на темата, но не и пълно разбиране на смисъла.

В определени специализирани приложения, обаче, машинният превод вече може значително да замести човешкия, особено при превода на климатични прогнози или технически инструкции, където използваният терминологичен набор и изразни средства са изключително ограничени. Възможно е този раздел да съдържа оригинални изследвания.

Добавяне на връзки към източници е препоръчително, в противен случай информацията може да бъде изтрита. Допълнителни детайли могат да бъдат открити на страницата за обсъждане.

Най-често, като на шега се забелязва, че програмата не успява да разпознае контекста на фрази, превеждайки термините и изразите дума по дума без разграничаване на собствени имена от общи думи. Например, преводачът PROMT представя "Лев Толстой" като «Lion Дебела» («Лев Толстой»), "сутиен-КЕТ нотация" като "Забележка Кейт сутиен", "Lie алгебра" като "алгебра на Lie", "ексцентричност вектор" като "вектор оригиналност", "Шоуни Смит" като "Шоуни Смит индианско племе" и т.н. В

сравнение, Google Translator често грешно превежда думата "ориз" като име на член на секретариата на Организацията на обединените нации.

4.Процедура

Превеждането може да се определи като последователност от две ключови стъпки:

I.Декодиране на значението на изходния текст:

Този етап включва разшифроване и разбиране на смисъла на текста, който трябва да бъде преведен. Преводачът трябва да анализира всички аспекти на изходния текст, включително граматика, семантика, синтаксис, идиоми и други езикови особености. Това изисква обширни познания не само за структурата на изходния език, но и за културните нюанси, свързани с него.

II.Кодиране на откритото значение в целевия език:

Втората стъпка включва предаване на разбраното значение в нов текст на целевия език, като се запази смисъла, стилът и естетиката на изходния текст. Тук отново се изискват високи компетенции по отношение на граматиката, семантиката и културните особености на целевия език.

Този прост процес съдържа сложна когнитивна задача.

Предизвикателството пред машинния превод се състои в това как да се програмира компютърът, така че да може "да разбира" текста с подобна на човека маневра и да създава нов текст на друг език със запазен смисъл и естетика, подобно на този, съставен от човек. Този проблем може да бъде решен чрез различни подходи.

5.Видове машинен превод

I.Машинен превод, базиран на речник

В най-простия вид, този метод машинен превод замества думите от един език с техните съответстващи в друг, без да прави корекции върху структурата на изречението или да отчита различните значения, които думите могат да придобият в различни контексти. Този вид "превод" има ограничено приложение.

II.Машинен превод, базиран на правила

В случая на машинния превод, базиран на правила, програмата ползва граматични правила за анализ и синтез на изречения и фрази. За успешното решаване на задачата за машинния превод, често се предполага, че трябва първо да се разреши задачата за разбиране на естествения език. Програмата анализира оригиналния текст, преобразува го в специфично вътрешно представяне и генерира преведения текст. За този вид превод се използват обширни речници и правила, включващи подробна морфологична, синтактична и семантична информация.

III.Статистически машинен превод

При статистическия машинен превод, информацията за съответствието между думите и изразите в два езика се събира автоматично от двуезични корпуси. Тези корпуси съдържат текстове, паралелно записани на два езика, като например парламентарни протоколи на английски и френски. След като програмата бъде "обучена", статистиката се използва за избиране на думи и фрази по време на превода.

IV.Машинен превод, базиран на примери

Машинният превод, базиран на примери, използва двуезични корпуси по време на самия превод, като непознати изречения се превеждат чрез аналогия с най-близкия познат на програмата пример. Този метод се приближава до машинното самообучение.

V.Машинен превод с помощта на междинен език

В случая на машинния превод чрез междинен език, изходният текст се трансформира в интерлингва - форма, независима от изходния и целевия език. След това се генерира текст на целевия език от това представяне.

6.Какви проблеми могат да възникнат при машинния превод

I.Многозначност

Думите, които имат повече от едно значение, са предизвикателство за автоматичния превод. В настоящия момент съществуват разнообразни подходи за преодоляване на този проблем, които могат да се разделят на две основни категории: "плитки" и "дълбоки".

При "плитките" подходи, които са по-популярни в практиката, няма нужда от разбиране на контекста на текста - вместо това се използват статистически методи за анализ на контекста на многозначните думи. При "дълбоките" подходи се предполага, че системата разполага с обширни познания за думите.

II.Имена

Често системите за машинен превод се сблъскват с трудности или дори са неспособни да различат имена или съкращения, които трябва да бъдат транслитерирани, транскрибирани или прехвърлени без промяна, вместо да бъдат преведени.

III.Неграматичен изходен текст

Недостатъците в изходния текст, като правописни или говорни грешки, лоша пунктуация или недовършени изречения, могат сериозно да затруднят автоматичния превод.

В някои случаи, за да се подобри работата на системата за машинен превод, оригиналният текст се подготвя ръчно, като се маркират думите, които не трябва да бъдат преведени (имена и съкращения), се премахва многозначността и други подобни. Проблемът с машинния превод възниква преди повече от две десетилетия. Същността му се състои в създаването на автомат, който получава текст на един естествен език (например на немски) и генерира текст на друг език (например на английски). Вместо този автомат, днес се използват електронни изчислителни машини, за които се разработват алгоритми за превод и речници.

Съвременният етап от развитието на машинния превод може да се определи като етап на синтактичен превод на думи. Основната единица за смисъла е думата, а граматичните форми и редът на последователност на думите в преведения текст се определят въз основа на синтактични връзки между думите в оригиналния текст. По аналогия с традиционните речници за автоматичен превод се създават главно като речници на думи (словосъчетанията в тях са малко).

В същото време в оригиналните езици думата не е единствената единица на смисъла. В тях се използват едновременно няколко нива на семантични единици (ниво морфема, ниво дума, ниво фраза, ниво изречение и други), като смисълът на единиците от по-високите нива не винаги може да бъде определен извън контекста на единиците от по-ниско ниво (това се вижда най-ясно в отношенията между нива морфема и нива дума). Следователно и речниците за автоматичен превод трябва да включват единици от различни нива.

Многовековният опит в комуникацията между народите, говорещи и пишещи на различни езици, показва, че добрият преводач трябва не само да разбира смисъла на превеждания език, но и да владее фразеологичното богатство на езика, на който се осъществява превода. Затова ръчният превод на текстовете е по-скоро фразеологичен, отколкото превод на думи. Този аспект трябва да се има предвид при автоматичния превод. Качественият машинен превод на научно-технически и делови текстове може да бъде осъществен чрез използването на комплекси от речници, които включват семантични единици от различните нива на думата, фразата, изречението и др. На предно място трябва да се отделят фразеологичните единици, тъй като в текстовете те изразяват най-често наименованията на понятията, представени чрез словосъчетания, а значително по-рядко - чрез отделни думи. При превода трябва да се спазва правилото за предпочитания, което предвижда първо изолиране и превод на семантичните единици от по-високите нива, а към единиците от по-ниските нива внимание да се обърне след това, и само ако няма други възможности за превод.

Важно е автоматизираната речникова служба да заема място в системите на машинен превод. Задачата ѝ е създаването и попълването (коригирането) на речници. Ролята на човека в автоматизираната речникова служба се състои в маркирането на чуждоезичните текстове с цел отделение на речниковите елементи и присвояването на съответните им еквиваленти на превеждания език. Този процес на маркиране може да се извършва и от машина.

В настоящия момент проблемът с машинния превод не е напълно решен и изисква дълбоки теоретични изследвания и обширна експериментална работа. За решението на този проблем са необходими големи, добре координирани научни екипи и мощна електронна изчислителна техника с достатъчно математически (програмни) елементи за изпълнение на задачата. Изследванията в областта на машинния превод трябва да бъдат

насочени към фундаментални изследвания, тъй като това би предоставило възможност за спестяване на ресурси за превод на чуждоезични текстове и улеснение на обмена на научно-техническа и делова информация между различните страни. Провеждането на такива изследвания е от съществено значение, тъй като те представляват един от най-ефективните пътища за откриване на техническо решение на множество проблеми в автоматичната обработка на информация, а развитието на тези средства и методи пряко влияе върху темпото на научно-техническия прогрес.

7.Разликата между машинния превод и „човешкия превод“

Въпреки че машинният превод предоставя бързи резултати, той не може да конкурира с качеството, достигнато от истински преводачи, които прилагат нюанси и професионален контрол. Клиентите, които обаче избират услугата машинен превод с последващо редактиране, могат да се възползват от по-професионално изгладен текст.

Инструментите за машинен превод демонстрират впечатляваща способност да обработват големи обеми текст със скорост и точност почти на всякакъв език. Въпреки това, те все още не могат да заменят човешката способност да улови нюанси, тон, хумор, сарказъм и значение.

Ако се ползва само машинен превод, роботизираният тон може да компрометира посланието, като предизвика впечатление за ниско качество на предлаганите продукти или услуги. Преводите с обществено значение изискват задълбочено редактиране от професионален преводач, за да бъдат разбрани естествено от потребителите. В търсенето на високо качество и локализирано съдържание се налага вложение в професионални преводачи или носители на езика, които да извършат редакцията след машинния превод.

Представената перифраза подчертава важността на хуманния аспект при превода и изразява загрижеността, че машинните системи все още не могат да достигнат нивото на разбиране на езиковите финеси, които са възможни чрез човешко вмешателство.

Вярвам, че от самото начало сме склонни да опростяваме своя поглед към друг език. Затова се насочваме към краткия и лесен път, който обаче може да ни доведе до неизвестни места. Като преводач аз насърчавам хората да се ангажират, да четат между редовете и да взаимодействат с реални

личности, които са специалисти в областта на превода, било то литературен, научен, икономически, кино и други.

Допълнително, изразявам убеждението си, че културното познание е от съществено значение за успешно изпълнение на задачите. Например, за да преведеш испанската дума "Guantanamo", трябва да разбереш, че това е жена, живееща в Гуантанамо.

С нарастването на броя на приложенията, които улесняват превода, възникват и конфликти в разбирането. Езиците не са статични таблици с малки лексикални единици, а по-скоро динамични и живи системи.

В контекста на моя роден език - български, видях многократни опити за дигитализация на думите, което считам за положително. Въпреки това, осъзнавам основния проблем, свързан с изоставането в контекстуализацията на значенията и изразите.

Заклучително, предизвикателствата на езиковото разбирателство ме насочват към мисълта, че докато можем да опростим усилията си с езика, ние никога няма да постигнем истинска красота, която се крие зад дестинацията на истинското разбиране и взаимодействие.

8. За и против машинния превод

Познато е, че Google Translate предоставя най-икономичното решение за преводи. Въпреки че услугата е напълно безплатна, използването на този инструмент прави преведената информация собственост на Google, което може да бъде рисковано, особено при фирмени документи или такива, които съдържат лична и чувствителна информация.

Също така, както беше посочено, точността на софтуерния превод е далеч от професионалната степен и не е подходящ за сериозни ситуации. За точни и надеждни преводи, предназначени за формални документи, се препоръчва използването на професионални преводачески услуги.

При големи обеми текстове, където точността не е от решаващо значение, софтуерният превод може да бъде подходящ. Също така, ако превода е предназначен само за вас и не е предназначен за широка аудитория, софтуерните решения могат да се окажат удовлетворителни. Въпреки това, за точни и професионални преводи, предназначени за външна употреба, е

препоръчително да се обърнем към професионални преводачески агенции.

Скоростта на машинния превод е несравнима, но качеството варира в зависимост от сложността на текста и езиковата комбинация. Най-евтиният вариант, въпреки това, се сблъсква с трудности в запазването на стила и идентичността на бранда. Четящият може да изпитва противоречиви усещания, тъй като текстът е разбираем, но на места губи смисъла си.

Използвана литература

1. https://bg.m.wikipedia.org/wiki/%D0%9C%D0%B0%D1%88%D0%B8%D0%B%D0%B5%D0%BD_%D0%BF%D1%80%D0%B5%D0%B2%D0%BE%D0%B4
2. [https://artedoc.com/%D0%B1%D0%BB%D0%BE%D0%B3/%D0%BF%D0%BE%D0%BB%D0%B5%D0%B7%D0%BD%D0%B8-%D1%81%D1%82%D0%B0%D1%82%D0%B8%D0%B8/%D0%BF%D1%80%D0%BE%D1%84%D0%B5%D1%81%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D0%B5%D0%BD-%D0%BF%D1%80%D0%B5%D0%B2%D0%BE%D0%B4-%D0%BE%D1%82-](https://artedoc.com/%D0%B1%D0%BB%D0%BE%D0%B3/%D0%BF%D0%BE%D0%BB%D0%B5%D0%B7%D0%BD%D0%B8-%D1%81%D1%82%D0%B0%D1%82%D0%B8%D0%B8/%D0%BA%D0%B0%D0%BA%D0%B2%D0%BE-%D0%B5-%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%B5%D0%BD-%D0%BF%D1%80%D0%B5%D0%B2%D0%BE%D0%B4-%D0%B8-%D0%BA%D0%B0%D0%BA%D0%B2%D0%B8-%D0%BF%D0%BE%D0%BB%D0%B7%D0%B8-%D0%BD%D0%BE%D1%81%D0%B8-%D0%BD%D0%B0-%D0%B1%D0%B8%D0%B7%D0%BD%D0%B5%D1%81%D0%B0.html)
3. <https://oltrans.bg/kakvo-predstavlyava-mashinen-prevod/>
4. <https://prevodachi.eu/mashinen-prevod-istoriya-v-1-tchast/>
5. <https://prevodachi.eu/mashinen-prevod-kakvi-problemi-mogat-da-vazniknat/>
6. <https://artedoc.com/%D0%B1%D0%BB%D0%BE%D0%B3/%D0%BF%D0%BE%D0%BB%D0%B5%D0%B7%D0%BD%D0%B8-%D1%81%D1%82%D0%B0%D1%82%D0%B8%D0%B8/%D0%BF%D1%80%D0%BE%D1%84%D0%B5%D1%81%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D0%B5%D0%BD-%D0%BF%D1%80%D0%B5%D0%B2%D0%BE%D0%B4-%D0%BE%D1%82->

[%D0%B0%D0%B3%D0%B5%D0%BD%D1%86%D0%B8%D1%8F-%D0%B8%D0%BB%D0%B8-%D0%BC%D0%B0%D1%88%D0%B8%D0%BD%D0%B5%D0%BD-%D0%BF%D1%80%D0%B5%D0%B2%D0%BE%D0%B4.html](#)

7.<https://polyglotclub.com/help/translations/human-vs-machine-translation/translate-bulgarian>

8.<https://bgkids.hallowedgaming.com/experts/mashinen-prevod-edno.html>