

תרגיל בית 5

מטלה זאת מחולקת לשני חלקים - (1) התנסות במערכת תרגום אוטומטית ו (2) הפקת עצי גזירה של משפט והבנתם.

(1) מערכות תרגום אוטומטיות

מטרת חלק זה היא להתנסות במערכת תרגום אוטומטית, תוך הבנת יתרונותיה וחסרונותיה. למרות האיכות ההולכת וגדלה של מערכות אלו, נופלות לא מעט טעויות בתרגום, במיוחד כאשר משפטי הקלט אינם טריוויאליים, ארוכים, או עושים שימוש במילים נדירות.

במשימה זו השתמשו במערכת תרגום קיימת - Google translate מערכת זו משתמשת במודל למידה עמוקה שפרטים נוספים עליו הם לא במסגרת הקורס. אמנם לא למדתם את המבנה של המערכת לפרטים אבל במהלך הקורס עסקנו בקשיים והבעיות שיש למערכות המתמודדות עם שפה טבעית.

עליכם למצוא טעויות בתרגום של המערכת הנ"ל ולאפיין אותן. סווגו את טעויות התרגום לסוגים שונים, ומצאו 3-4 משפטי קלט אשר גורמים לטעות מכל סוג. המטרה היא כמובן למצוא כמה שיותר סוגי טעויות שונים (ומנומקים), ולדייק בדוגמאות לסוגי טעויות אלה.

עליכם להתמקד בזוג השפות אנגלית - עברית, בשני כיווני התרגום.

להלן מספר דוגמאות למשפטים ומבנה תשובה מצופה:

טעויות מסוג "זיהוי לא נכון של צירוף מילים":

המשפט: "translation is not as easy as you make it out to be"

תורגם ל: " התרגום הוא לא קל כמו שאתה עושה את זה כדי להיות"

תיאור הבעיה: המערכת לא מצליחה לזהות את הביטוי "make it out to be" ולתרגם אותו למשמעות אחת ובמקום מתרגמת אותו באופן מילולי.

...

טעויות מסוג "פתרון לא נכון של עמימות לקסיקלית":

המשפט: "נעצר המורה שניצל ופיתה את תלמידיו"

תורגם ל: "The teacher Shnitzel was arrested and seduced by his students"

תיאור הבעיה: למילה "שניצל" יש שני פירושים אפשריים, כאשר אחד מהם (ש-ניצל) דורש פירוק מורפולוגי. כמו כן, הטעות יכולה לנבוע מהתפלגות שונה לשני הפירושים האפשריים בקורפוס האימון.

כמו כן, יש בעיה של ניתוח תחבירי, במשפט רשום ש "המורה פיתה את תלמידיו" ואילו בתרגום המשמעות ההפוכה (התלמידים פיתו את המורה).

...

(2) פענוח עצי גזירה

מטרת החלק השני היא להתנסות בהבנת עצי גזירה של משפטים באנגלית. למטרת ביצוע חלק זה השתמשו ב stanford parser, אשר זמין online כאן - [/http://nlp.stanford.edu:8080/parser](http://nlp.stanford.edu:8080/parser)

שימו לב שה-parser תמיד מפיק עצי גזירה, לכל קלט, ולעיתים מפיק מספר עצי גזירה עבור אותו משפט אשר מדורגים בצורה הסתברותית. ה-parser לעיל מציג את עץ הגזירה הסביר ביותר (לדעתו) למשפט נתון. אנו מעוניינים לדעת האם עץ הגזירה שנבחר תואם את המשמעות של המשפט.

א. התבוננו ברשימת המשפטים מטה, הפיקו עץ גזירה עבור כל אחד מהם וסמנו האם הוא תקין (תואם את המשמעות) או לא. אם לא - נמקו.

1. The dog ate the bone thrown by Paul
2. The dog ate the bone encouraged by Paul
3. The horse that was raced past the barn fell
4. The horse raced past the barn fell
5. She told me a little white lie will come back to haunt me
6. We painted the wall with the cracks
7. We painted the wall with a paintbrush
8. We painted the wall with a happy heart and a big smile
9. Blonde on Blonde is the seventh studio album by American singer-songwriter Bob Dylan released on May 16 1966 by Columbia Records

10. The honor is given each year for collections of fantasy stories by a single author published in English

ב. חשבו על מספר דוגמאות נוספות אשר מפיקות עץ גזירה אשר לא תואם את משמעות המשפט. המשפטים צריכים להיות תקינים מבחינה דקדוקית. הביאו דוגמאות לבעיות שונות ככל הניתן, אין טעם לתת מספר משפטים אשר מתארים את אותה תופעה בדיוק.

הגשה: עליכם להגיש דו"ח hw5report.pdf מפורט עם שני חלקי התרגיל. הקפידו על מבנה מסודר וברור, ומלל תמציתי.

הציון יינתן באופן יחסי, כלומר העבודה הטובה ביותר תקבל את כל הנקודות ושאר העבודות ינוקדו באופן יחסי לאותה העבודה. זאת על מנת לעודד עבודה עצמאית.

מועד ההגשה: 07.01.19, 23:59.