אוניברסיטת חיפה
החוג למדעי המחשב

עיבוד שפות טבעיות
סמסטר אביב 2001

בוחן 2

1. בעמוד הבא מתוארת סכימת הניתוח של Earley כפי שניתנה בכיתה.

(א) נגדיר סכימת ניתוח חדשה הזהה לסכימת Earley בכל, מלבד העובדה שכללי ההיסק הקשורים לפעולת predict בוטלו. האם הסכימה החדשה שלמה? האם היא נאותה? הסבירו את תשובתכם.

ללא Predict, הסכמה אינה שלמה: על פי האקסיומה היחידה לעולם לא יתמלא התנאי הראשון (השמאלי) של Complete עבור חוק שגופו אינו מתחיל בסמל S.

הסכימה היא נאותה: <u>הפחתת</u> כללים אינה יכולה לגרום לכך שיווצרו <u>יותר</u> פריטים.

(א) נגדיר סכמת ניתוח חדשה המבוססת על Earley, אלא שבמקום כללי ההיסק הקשורים לפעולת predict נוסיף אקסיומות מהצורה $[j, B \rightarrow \bullet\gamma, j]$ לכל חוק $[B \rightarrow \gamma]$ בדקדוק. האם הסכימה החדשה שלמה? האם היא נאותה? הסבירו.

הסכמה היא שלמה, שכן פעולת Predict יוצרת אך ורק פריטים מהצורה $[j, B \rightarrow \bullet\gamma, j]$, וכל הפריטים האפשריים מהצורה הזו נוספו כאקסיומות. הסכימה היא נאותה, שכן האקסיומות החדשות מייצגות טענות לוגיות אמיתיות: משמעותן היא שניתן להניח את החוק $[B \rightarrow \gamma]$, החל ממקום j, אך עדיין לא "נראה" אף חלק מ-γ. טענה זו מקיימת באופן ריק את הפירוש שנתנו לפריטים בסכימה של Earley.

2. אחת הטענות שהועלו פעמים רבות כדי להראות ששפות טבעיות אינן חסרות הקשר נוגעות לקיומם של מבני-חזרה בשפות טבעיות. בצורה כללית ומופשטת, מבני חזרה מעל אלפבית Σ הם מבנים מהצורה $\{ww \mid w \in \Sigma^*\}$.

(א) הראו תת שפה אינסופית של עברית מהצורה $\{s^i = w_1^i \alpha w_2^i \beta \mid i \geq 0\}$ כך ש- s^i שייך לשפה העברית אם ורק אם $w_1^i = w_2^i$.

<u>נצחו</u> או לא <u>נצחו</u>, בכל זאת אני חושב שמכבי חיפה קבוצה גרועה.

<u>נצחו את בית"ר ירושלים</u> או לא <u>נצחו את בית"ר ירושלים</u>, בכל אני חושב שמכבי חיפה קבוצה גרועה.

<u>נצחו את בית"ר ירושלים בטדי</u> או לא <u>נצחו את בית"ר ירושלים בטדי</u>, בכל אני חושב שמכבי חיפה קבוצה גרועה.

<u>נצחו את בית"ר ירושלים בטדי אחרי שהרחיקו להם שחקן</u> או לא <u>נצחו את בית"ר ירושלים בטדי אחרי שהרחיקו להם שחקן</u>, בכל אני חושב שמכבי חיפה קבוצה גרועה.

(ב) בעמוד הבא נתון דקדוק יוניפיקציה היוצר את השפה $\{ww \mid w \in \{a,b\}^*\}$. השלימו את המקומות החסרים בדקדוק.

4. מוצע לממש מבני תכוניות כגרפים על פי ההגדרה הבאה: מבנה תכוניות $A = \langle Q, q, \delta, \theta \rangle$ הוא גרף מכוון, סופי, קשיר ומסומן שבו קבוצה לא ריקה של צמתים Q, שורש $q \in Q$, פונקציה חלקית $\delta : Q \times Feats \to Q$ המתארת את הקשתות ומסמנת אותן בתכוניות, כך שקיים מסלול מכוון מהשורש q לכל צומת אחר, ופונקציה חלקית $\theta : Q \to Atoms$ המסמנת חלק מהבורות בגרף (צמתים ללא קשתות יוצאות) באטומים.

הגדירו את יחס ההכללה (subsumption) במונחים של גרפים. במילים אחרות, בהינתן שני גרפים A ו-B המייצגים מבני תכוניות, הגדירו מתי $A \sqsubseteq B$.

נסמן $A = \langle Q_A, q_A, \delta_A, \theta_A \rangle$ ו- $B = \langle Q_B, q_B, \delta_B, \theta_B \rangle$. אזי $A \sqsubseteq B$ אם ורק אם קיימת פונקציה טוטאלית $h : Q_A \to Q_B$ כך ש:

1. $h(q_A) = q_B$ (הפונקציה ממפה את השורש של A לשורש של B).
2. לכל צומת $q \in Q_A$ ולכל מסלול $q_0, q_1, ..., q_n$ ב-A כך ש-$q_0 = q_A$ ו-$q_n = q$ שקשתותיו מסומנות ב-$f_1, f_2, ..., f_n$, קיים מסלול $h(q_0), h(q_1), ..., h(q_n)$ ב-B שקשתותיו מסומנות באותן תכונות (הפונקציה ממפה כל מסלול ב-A למסלול זהה ב-B).
3. לכל צומת $q \in Q_A$, אם $\theta(q)$ מוגדרת אזי גם $\theta(h(q))$ מוגדרת ו-$\theta(q) = \theta(h(q))$ (אם ערך הוא אטומי ב-A הוא ממופה לאטום זהה ב-B).

בהצלחה!