

תרגיל בית 3

שאלה 1

להלן סדרת משפטים תקינים בעברית:

1. הקופסה היא לבנה.
2. האגרטל הקטן הוא לבן.
3. הקופסה הקטנה היא לבנה.
4. האגרטל הקטן שעל המדף הוא לבן.
5. הקופסה הקטנה שעל המדף היא לבנה.
6. האגרטל הקטן שעל המדף השחור הוא לבן.
7. הקופסה הקטנה שעל המדף השחור היא לבנה.
8. האגרטל הקטן שעל המדף השחור שליד השידה הוא לבן.
9. הקופסה הקטנה שעל המדף השחור שליד השידה היא לבנה.
10. האגרטל הקטן שעל המדף השחור שליד השידה השחורה הוא לבן.
11. הקופסה הקטנה שעל המדף השחור שליד השידה השחורה היא לבנה.

שימו לב שהמרחק בין הנושא (קופסה/אגרטל) לנושא (לבן/לבנה) אינו חסום.

- א. הראו דקדוק חסר הקשר המייצר את המשפטים דלעיל (ואולי גם נוספים שאינם תקינים). על הדקדוק לייחס לכל משפט עץ גזירה סביר מבחינה לשונית.
- ב. הראו את עץ הגזירה המתקבל עבור משפט מספר 11.
- ג. הראו דקדוק עבור המשפטים שלעיל שימנע משפטים לא תקינים בהם תהיה חוסר התאמה (במין) בין שם עצם לשם התואר המצורף לו.

שאלה 2

להלן סדרת משפטים תקינים בעברית:

1. דן גבוה ורזה.
2. דן גבוה או רזה.
3. דן גבוה ורזה או מאושר.
4. דן או גבוה ורזה או מאושר.

- א. עבור כל אחד מן המשפטים קיבעו האם הינו חד משמעי או רב משמעי. במידה והמשפט הינו רב משמעי הראו את כל משמעויותיו השונות.
- ב. הראו דקדוק חסר הקשר המייצר (גם) את המשפטים דלעיל ומבטיח ריבוי משמעות במקומות המתאימים ובהם בלבד (כמובן כי על הדקדוק לייחס לכל משפט עץ גזירה סביר מבחינה לשונית).
- ג. הראו את עצי הגזירה המיוחסים למשפטים ע"י הדקדוק (עבור כל משפט יש להראות את כל עצי הגזירה המתקבלים עבורו).

שאלה 3

הציעו שינוי באלגוריתם הניתוח של Earley כך שניתן יהיה לסרוק את הקלט מימין לשמאל והאלגוריתם יהיה שלם (כלומר, יפיק את כל הניתוחים האפשריים). שימו לב: השינוי צריך להיות באלגוריתם ולא בדקדוק!

שאלה 4

1. ממשו את אלגוריתם הניתוח CYK כפי שהוצג בכיתה.

2. הרחיבו את התוכנית כך שתפיק לא רק תשובות של כן/לא, אלא גם עצי ניתוח עבור משפטים תקינים. במידה וקיים יותר מעץ גזירה אחד, על התוכנית להפיק את כל עצי הגזירה האפשריים. ממשו פרוצדורה המדפיסה עצי גזירה בצורה קריאה.

3. הסבירו כיצד ניתן להרחיב את אלגוריתם הניתוח CYK כך שיטפל גם בחוקי ϵ .

בנוסף: נקודות בנוס יינתנו עבור הרחבת התוכנית שכתבתם כך שתטפל נכונה גם בחוקי ϵ .

הנחיות:

(i). על התוכנית לקבל כקלט דקדוק חסר הקשר כלשהו בצורת CNF. הדקדוק יינתן בקובץ ששמו יועבר לתוכנית כפרמטר בזמן ריצה. הינכם יכולים להחליט על פורמט קובץ הדקדוק כרצונכם. לאחר קריאת הדקדוק ועיבודו, על התוכנית לקבל מספר כלשהו של משפטים (מספרם אינו ידוע מראש!) ולהפיק עבורם ניתוחים. המשפטים יינתנו ברצף בקובץ נפרד.

(ii). יש להגיש:

א. מסמך תיעוד מפורט של התוכנית, כולל פירוט מבני הנתונים והאלגוריתמים בהם השתמשתם. בפרט, הסבירו כיצד הרחבתם את התוכנית כך שתייצר עצי ניתוח עבור משפטים תקינים ומהו פורמט קובץ הדקדוק.

ב. הדפסה של התוכנית - חובה.

ג. עליכם להגיש דוגמאות קלט ופלט מודפסות לפחות עבור דוגמת קלט 1, וכן עבור הדקדוקים שכתבתם בשאלות 1 ו-2 (כמובן כי עליכם להעביר את הדקדוקים שכתבתם לצורת CNF). עבור כל משפט הדפיסו את כל עצי הגזירה האפשריים (במידה וקיימים). במידה ומימשתם את חלק הבנוס הגישו גם קלט ופלט מודפסים עבור דוגמת קלט 2.

ד. דיסקט ועליו כל הקבצים הרלוונטיים, כולל קובץ exe וקבצי הקלטים הנתונים בהמשך בפורמט המתאים. יש לצרף דף ובו הוראות הפעלה מדויקות של התוכנית!

Grammar 1

$$V = \{S, NP, VP, PP, D, N, V\}$$

$$\Sigma = \{the, a, some, cat, hat, mat, mice, in, on, near, sat, smiled, saw\}$$

$$S \rightarrow NP \ VP$$

$$S \rightarrow S \ PP$$

$$NP \rightarrow D \ N$$

$$NP \rightarrow NP \ PP$$

$$PP \rightarrow P \ NP$$

$$VP \rightarrow V$$

$$VP \rightarrow V \ NP$$

$$VP \rightarrow V \ PP$$

$$VP \rightarrow VP \ PP$$

$$D \rightarrow the \mid a \mid some$$

$$N \rightarrow cat \mid hat \mid mat \mid mice$$

$$P \rightarrow in \mid on \mid near$$

$$V \rightarrow sat \mid smiled \mid saw$$

the cat smiled

the cat sat on the mat

the cat saw some mice near the mat

the cat in the hat saw a hat on the mat near some mice

*the cat saw a mat a hat

*the cat in the hat some mice

*the cat saw mice

*the dog saw some mice

Grammar 2

$$V = \{S\}$$

$$\Sigma = \{a, b\}$$

$$S \rightarrow aSb$$

$$S \rightarrow \varepsilon$$

a a b b

a a a a b b b b

* a a b

מועד הגשה:

הגשה עד לתאריך 12.1.04. הגשה ביחידים או בזוגות.