

מבוא לבלשנות חישובית
קורס 67571 תשס"ו

מבחן מסכם מועד א'
19 בפברואר 2006

מותר שימוש בחומר עזר

בהצלחה!

משך המבחן: שעתיים

יש לענות על 4 מתוך 5 השאלות.

1. בגישה הפורמלית הסטנדרטית, משפט יכול להיות תקין או בלתי תקין ביחס לדקדוק נתון. הועלתה הצעה להגדיר סולם של דרגות תקינות תחבירית.
 - א. האם / לשם מה הדבר נחוץ?
 - ב. תארו מודל חישובי שיוכל לתמוך בגישה כזו.
2. בניתוח תחבירי בשיטת GLR (Generalized LR parsing) יתכן שתא בטבלת ההחלטה יכיל הוראות פעולה חלופיות (מצב הנקרא: "action conflict").
 - א. איזה סוג של רב-משמעות משתקף ב- Shift-Reduce conflict ואיזה ב- Reduce-Reduce conflict?
 - ב. האם/איזה שילובים נוספים של הוראות פעולה ייתכנו בתא בטבלה?
 - ג. האם/איזה קושי קיים בניתוח המשפט הבא:
Although Mary forgot her husband didn't seem upset
האם/איך תמיכה במבנה זה תתבטא בטבלת ההחלטה של אלגוריתם GLR?
3. נתונים bottom-up chart parser לדקדוק חסר-הקשר ואלגוריתם האחדה למבני תכונות כלליים (DAG).
 - א. באיזה אופן תשלבו את אלגוריתם ההאחדה עם אלגוריתם ה- chart parsing בכדי לקבל מנתח לדקדוק חסר-הקשר עם אילוצי האחדה? הסבירו את הרעיון בקווים כלליים, רצוי עם דוגמה, אך אין צורך בהגדרה פורמלית שלמה.
 - ב. אנו רוצים להעשיר את המנגנון הנ"ל בפעולות דיסיונקציה ושלילה על מבני תכונות. מהי המסגרת הפורמלית המתאימה? מהי הדרך המעשית לשילוב?
 - ג. האם דקדוק עם חוקי מבנה חסרי-הקשר, אילוצי האחדה ופעולות דיסיונקציה ושלילה שקול לדקדוק חסר-הקשר מבחינת כוח הביטוי שלו?

07.577
16/10/11

4. נתבונן בכמתים המוכללים הבאים:

- few fish
- exactly five fish
- at most five fish

- א. איזה קבוצות יכולות לשמש בתפקיד של "קבוצת עדות" (witness set), כהגדרת בארוויז וקופר) לכל אחד מהכמתים המוכללים הנ"ל?
- ב. האם/איך קבוצות עדות אלה יוכלו לסייע בקביעת ערך האמת של הטענות few fish swim, exactly five fish swim, at most five fish swim ?

5. בשיעור הגדרנו שני סוגים של משלימים לכמת מוכלל Q:

$$\sim Q = \{X \subseteq E \mid X \notin Q\} \quad Q \sim = \{X \subseteq E \mid (E-X) \in Q\}$$

לצורך המענה על השאלות הניחו עולם E המכיל ישויות שונות ובהם חתולים ודגים.

- א. תנו דוגמה לכמת מוכלל Q שעבורו $\sim Q \neq Q \sim$
- ב. קל לראות ש $\sim \sim Q = Q \sim \sim = Q$. האם גם $\sim(Q \sim) = (\sim Q) \sim = Q$?

להלן משימות שעליך לבצע ולהגיש כעבודה מסכמת לקורס, במקום בחינה. יש להגיש את העבודה עד סוף חודש פברואר 1991.

חובה עליך לבצע את משימות (1) ו-(2). משימה (2) משאירה חופש מסוים לגבי היקף הפתרון; אם נתת פתרון מקיף, אין חובה לבצע את משימה (3), אך אם טפלת בבעיות של משימה (2) רק באופן בסיסי, עליך לבצע גם את משימה (3).
כלומר, אפשר להגיע לציון המירבי בשתי דרכים: על-ידי הגשת משימה (1) ותשובה מקיפה על משימה (2), או על-ידי הגשת כל שלוש המשימות (1)-(3), כאשר הטיפול במשימה (2) נעשה ברמה בסיסית בלבד.
מובן שמותר להגיש את משימה (3) גם אם טפלת במשימה (2) באופן מקיף, ובכך לשפר את הסיכוי לציון גבוה, אך אין חובה כזו.

סימון: הכינוי CMU-PC ישמש להלן כשם מקוצר למערכת התוכנה בה אנו משתמשים: CMU: The Generalized LR Parser Compiler, M. Tomita.
החלקים המעשיים של העבודה יבוצעו עם מערכת תוכנה זו, ויוגשו כתדפיסים הכוללים את קובץ הדקדוק המתאים ותעתיק (SCRIPT) של הרצה לדוגמה.

הערה כללית: השימוש בגוף-שני-זכר בדפיס אלה נעשה מטעמי נוחות בלבד, כמקובל, ואין בו כל כוונת אפלייה לגבי תלמידינו ממין נקבה.

ב ה צ ל ח ה !

משימה (1)

קרא את המאמר המצורף בזה
Masaru Tomita and Kevin Knight (1989)
"Pseudo-Unification and Full-Unification"

המאמר מתאר את הגישה המיוחדת לתהליכי האחדה (יוניפיקציה) הננקטת במערכת התוכנה הלשונית CMU-PC. הסבר בכתב במה שונה גישה זו מהגישה ה"טהורה" (קנונית) לתהליכי האחדה. הצבע על יתרונות (בפרט בקשר לעיבודי שפה טבעית) ועל חסרונות. חווה את דעתך האישית לגבי המאזן בין יתרונות וחסרונות (למשל, האם אתה עצמך היית מעדיף פיתוח בסביבה של יוניפיקציה טהורה, והאם דעתך תופשת במידה שווה לגבי תרגיל קטן ולגבי דקדוק רחב היקף לשפה).

67-571
1k/k"Jen

משימה (2)

CMU-FC

השתמשו (י) בחבילה של Tomita ליצירת מודל לניתוח משפטים תוך הבחנה בין סוגי הפעלים הבאים:
א'. פעלים יוצאים, שהם פעלים בעלי שני ארגומנטים, האחד נושא - Subject, והשני מושא - Object.
ב'. פעלים עומדים, שהם פעלים בעלי ארגומנט אחד, הנושא.
ג'. פעלים סבילים או אמצע, שגם הם פעלים בעלי ארגומנט אחד, אלא שתפקידו הוא מושא.

- המערכת שתבנה/י תנתח את המשפטים הבאים ותדחה את המסומנים בכוכב:
- (1) הדובר פגש את העתונאים
 - (2) * הדובר פגש
 - (3) התותחים רעמו
 - (4) * התותחים רעמו את הטילים
 - (5) נפצעו אזרחים
 - (6) * הטילים נפצעו אזרחים

הערות:

א'. כאשר אין נושא, ההתאמה במין ובמספר היא בין הפועל למושא:
(7) * נפצע אזרחים

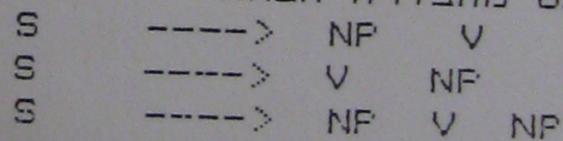
ב'. המערכת חייבת לטפל במשפטים כמו (1) עד (7) לעיל, בהם קודם הנושא לפועל, ואילו המושא מופיע אחרי הפועל. אם יש ביכולתך, נסה/י גם לדאוג לסדרי מילים נוספים, דהיינו כאלה בהם קודם המושא לפועל:

- (8) אזרחים נפצעו
- (9) את העיתונאים פגש הדובר
- (10) את העיתונאים הדובר פגש

שים/מי לב לכך שהנושא בא אחרי הפועל רק בתנאי שיש מושא הקודם לפועל, כמו ב(9) ולעומת (12) ו-(14):

- (11) דני נושם
- (12) * נושם דני
- (13) הטילים מדאיגים את דני
- (14) * מדאיגים הטילים את דני

ג'. מותר לך, בעיקר אם בחרת לענות על ב', לוותר על רכיב ה-VF במשפט, ולכתוב כללי משפט מהצורה הבאה:



בכל מקרה, בין אם ענית על ב' ובין אם לא, נתח(י) את היתרונות והחסרונות שבשמוש ברכיב ה-VF במשפט.

3/111

67.571
"JPS" / k

משימה (3)

כתוב דקדוק (במונחי CMU-PC) המזהה את השפה בה "משפטים" הם ביטויים מהצורה $a^n b^n c^n$ ("אוצר המלים" כולל את a, b ו- c).
כלומר: $\underbrace{a \dots a}_n \underbrace{b \dots b}_n \underbrace{c \dots c}_n$ פעמים.

זו כידוע שפה שאינה חסרת-הקשר, כלומר דרושים אמצעים חזקים יותר מ-Context Free Grammar. הסבר מה ניתן לפורמליזם של CMU-PC את הכוח הנוסף הדרוש.

האם בפתרון שלך השתמשת באמצעים חישוביים של CMU-PC שאינם בגדר האחדה (יוניפיקציה) טהורה? אם כן, האם לדעתך אפשר לפתור את התרגיל גם באמצעי האחדה טהורה?

ב ה צ ל ח ה !

4/...

67.816
תשס"א

ה מ ח ל ק ה ל מ ד ע י ה מ ח ש ב

בחינה בבלשנות ומחשבים

מועד א'

שם המורה: פרופ' עוזי אורנן, ד"ר מיכל אפרת

מס' הקורס: 67816

משך הבחינה: שעתיים

תאריך הבחינה: 20.3.85 שעה: 09:00

מקום הבחינה: מתמטיקה 2

ענה על שאלות 1,2 ועל עוד שתי שאלות

1. רשום מה יהיו מרכיביה של תוכנית-מומחה חד לשונית לתיקון שגיאות כתיב (בעברית) ציין במיוחד אילו חלקים יהיו דומים לתוכנית מומחה אחרת שאתה מכיר (ציין אילו) ואילו חלקים יהיו מיוחדים לתוכנית זו.
2. מכפלים ואותיות יכולים להידון בהגדרת המילים של OCY (WORDS*) במיוחד באחד מכמה מקומות ובכמה אופנים (כתו עצמאי או כשוה לתו אחר). פרט אילו מקומות, דון ביתרון ובחסרון שבכל אחד מהמקומות והאופנים עבור הטקסט העברי הבא (העתק אותו תחילה לכתב פונמי)

רק הסבל שסבל וחדל לסבול
סבר שכל הסבל היה כדאי

3. מה הבעיות ביצירת מכונה קוראת עבור עוורים בעברית.
4. לפניך רשימה של חוקים. רשום ונמק עבור כל חוק לאיזה תחום מתחומי הבלשנות (הגה, צורות, תחביר, סמנטיקה או פרגמטיקה) הוא שייך, וכך באיזו שפת מחשב רצוי לנסחו.
 - א. שם עצם + הַת + שם עצם מופשט
 - ב. עיצור₁ + תנועה + עיצור₂ + עיצור₃ * עיצור₁ + תנועה + עיצור₂ + תנועה + עיצור₃
 - ג. רווק + לא נשוי
 - ד. * מלך + לא יש כתר
 - ה. מלך ראה + מלכים ראו
 - ו. קנה + תי + קניתי
 - ז. דוד האפיל את החדר בניר שחור + הניר השחור האפיל את החדר

5. לפניך אוצר המלים של "לשונית" הדומה בחוקיה לעברית

דוד	אהב	את
יעל	ראה	
רינגו*	נשך	ש

- א. כתוב את חוקי הדקדוק של הלשונית. השתמש בסימני העזר מ (משפט) צפ (צירוף פועלי), עש (צירוף שמני), ש(שם), פ(פועל), י(יחס), ק(קישור)
- ב. הדגם שלושה משפטים (שוני מבנה) המתקבלים מתוך דקדוק זה.
- ג. הוכח שבלשונית זו יש מספר אינסופי של משפטים

* השם רינגו מכוון לשם כלב

67816
משנה
17/1
C1

בחינה בבלשנות ומחשבים

מסי קורס 67816

תאריך הבחינה: 8.9.85

משך הבחינה: 2 ש'

מקום הבחינה: מתי 2

מועד ב'

שם המורה: פרופ' ע. אורנן, ד"ר מ. אפרת

ענה על שאלות 1,2 ועל עוד שתי שאלות.

שאלה-1

עמוד על ההבדלים שבין GUS לבין מערכת אחרת כגון Winograd Schank או Eliza מבחינת הטיפול בנושאים בלשניים וההזדקקות להם.

שאלה 2

יש מספר דרכים לטיפול במיליות בתוכנית OCP. השתמש בטקסט העברי הבא (לאחר שתתעתק אותו-תחילה לכתב פונמי) להדגים את התוצאות של כל דרך תוך שאתה מעיר על יתרונותיה וחוסרונותיה ובמידת הצורך השפעת הדרך על פקודות אחרות ב OCP (די למנות שלוש דרכים שונות זו מזו):

הטקסט: וכשראו שאול ואורן שבכל בית יש אור מיד פנו בשביל שהוא באור החזק של כל הבתים.

שאלה-3

מה החסרונות של כתיבת מנתח תחבירי בלא מנתח סימנטי. עמוד על חסרונות כלליים (עבור כל השפות) ועל חסרונות המיוחדים לעברית.

שאלה 4

לפניך קבוצת משפטים, קבע עבור כל משפט האם הוא תקין או שאין הוא תקין (הוא נוגד את האינטואיציות שלנו כדוברים). לאחר מכן נסח מסגרות פילמור עבור הפועל של כל אחד מן המשפטים ורשימת תכונות עבור השמות, באופן שרק המשפטים שסימנת כתקינים יתקבלו מצירוף מסגרת הפועל לשמות.

- א. דוד קרא שולחן מענין.
- ב. דוד שאל את יוסף שולחן.
- ג. יוסף אמר לדוד שהספר מענין.
- ד. יוסף קרא ספר מענין.
- ה. השולחן נרטב.
- ו. הרעיון נרטב.
- ז. יוסף שאל את דוד אם הרעיון מענין.

שאלה-5

מבין תחומי הבלשנות (הגה, צורות, תחביר, סמנטיקה ופרגמטיקה) בחר שלושה תחומים והדגם עבור כל אחד מהם חוק ריקורסיבי (נסח את החוק ולווחו בדוגמה).

07. 816
תשנ"א

c. 1

החוג ללשון העברית

בלשנות ומחשבים

ע' אורנן

(מבחן סיום הקורס)

ס.מ. 67816

17.7.81

מועד א' תשמ"א

ענה על שאלה 1 או 2 ועל שאלה 3

1. כתוב משפט עברי בן 8-10 מלים בכתב פונמי (אוחיות לסיניות) לפי הדרך שנהגנו בה בשיעורים.
2. שיטת בר-אילן ושיטת המלון ההיסטורי. סקירה וביקורת.
3. כתוב דקדוק או סרטס חרשים זרימה או אלגוריתם בניסוח כלשחו, שיתן את התוצר המבוקש:
פעלים בבנין בפעל, הקלט הוא בעבר; הפלט: אני, אתה בעתיד, גזרות השלמים (כולל עיצורים גרוניים) ול"י.

בתצלחה!