

ניהול נתונים באינטרנט - תרגיל מס' 4 - IE and Description Logic

הנחיות הגשה

ההגשה תתבצע בזוגות, דרך אתר moodle, שם גם מוגדר מועד ההגשה. יש להגיש את הפתרונות בקובץ zip יחיד בשם username_wdm.zip (למשל, daniel2_wdm.zip) שיכיל את answers.pdf ואת קבצי הקוד. רק אחד מבני הזוג יגיש את המטלה (ניתן להשתמש ב-username שלו), אבל הקפידו לציין בראש הקובץ answers.pdf שמות ות.ז. של שני בני הזוג.

שאלה 1

כתבו תוכנית JAVA בשם Royal שמחלצת מויקיפדיה מידע על משפחת המלוכה הבריטית. הפלט צריך להיות קובץ בשם royal.txt שכולל שלשות מהצורה הבאה

xxx related_to yyy

yyy,xxx הם שמות של **חברים במשפחת המלוכה** או לינקים לדפים שלהם בוויקיפדיה.

האלמנטים בשלשה מופרדים ב-TAB, וכל שלשה מופיעה בשורה חדשה. "related_to" הוא קבוע, כלומר אין צורך לחלץ את סוג הקשר. מספיק גם ליצור שלשה אחת עבור כל זוג, ואין צורך לייצר קשרים "טרנזיטיביים" כלומר מספיק לייצר קשרים שהופיעו במפורש בדפים (ראו להלן).

הנחיות

יש להתחיל מהדף

http://simple.wikipedia.org/wiki/Prince_William,_Duke_of_Cambridge

ולעבור לדפים אחרים אליהם יש לינקים מהדף (ודפים אליהם יש לינקים מדפים אלו וכן הלאה, ניתן להגביל את החיפוש ל-50 דפים).

ניתן ורצוי להשתמש בקוד שכתבתם בתרגיל 2, בשינויים המתחייבים.

לצורך זיהוי הקשרים, יש לחלץ מידע רק מפסקאות המבוא שלפני תוכן העניינים.

בקובץ ה- html יהיו אלה טקסטים (ולינקים) המופיעים בין `<p>` ל- `</p>` ולפני התגית

```
<div id="toc" class="toc">
```

חילוך המידע יסתמך על לינקים: אם מהדף של x יש לינק לדף של y, נניח ש- x קשור ל- y (ואלו הקשרים היחידים שעליכם למצוא)

בנוסף, עליכם לנסות ולזהות ש- y הוא אכן אדם ובן משפחת המלוכה. לצורך כך, ניתן להשתמש בכל מידע מתוך דף הויקיפדיה שלו. מצאו היוריסטיקות ונסו לטייב אותן.

הגשה

א. הגישו את הקוד (כולל מחלקות עזר שמימשתם)

ב. הגישו את קובץ הפלט royal.txt.

ג. צרפו לקובץ התשובות ניתוח ידני של ה- precision (אחוז התשובות הנכונות מתוך אלו שדווחו) עבור 10 עובדות שחילצתם (לבחירתכם). עבור עובדות שגויות, הסבירו בקצרה מדוע לדעתכם הן התקבלו.

ד. צרפו לקובץ התשובות ניתוח ידני של ה- recall (אחוז התשובות שחולצו מתוך אלו שהיה ניתן לחלץ מפסקאות המבוא), ביחס לשלושה מהדפים. מצאו עובדה שלא חולצה והסבירו בקצרה מדוע לדעתכם כך קרה.

שאלה 2

הוכיחו או הפריכו את הספיקות של קבוצות הטענות הבאות ב-ALC, תוך שימוש בשיטת ה-tableau. פרטו את כל שלבי ביצוע השיטה.

א.

$$\{[\exists S.C \sqcap \exists S.D](a), [\forall S.(\neg C \sqcup \neg D)](a)\}$$

ב.

$$\{[\exists R.(\forall S.C)](a), [\forall R.(\exists S.\neg C)](a)\}$$