

ניהול נתונים באינטרנט – תרגיל מס' 4 – PROB DB and DL

הוראות:

יש לעלות את הפתרונות בקובץ ZIP לMOODLE, שכולל קובץ PDF בשם answers.pdf ובו הפתרון וקבצי קוד נוספים (HTML, XML או Python) לפי הדרישה של כל סעיף וסעיף. ההגשה היא בזוגות, ורק אחד מבני הזוג יגיש את התרגיל, אך יש להקפיד לכתוב את השמות והת.ז. של שני בני הזוג בתוך הקובץ. שם של הקובץ ZIP צריך לכלול את הת.ז. של אחד מהמגישים (למשל: HW1_123.zip). תאריך הגשה: 10.06.2018

שאלה 1:

הוכח או הפרך, תוך שימוש בשיטת ה tableau את הטענות הבאות לגבי לוגיקת ALC:

$$\alpha. \{ [\exists R. ((\forall R. C) \sqcap C')] (a), [\neg (\exists R. (C \sqcup C'))] (a) \}$$

$$\beta. \{ [\exists R. ((\exists R. C) \sqcap C')] (a), [\neg (\exists R. (C \sqcap C'))] (a) \}$$

תזכורת: לוגיקת ALC כוללת את הפעולות הבאות

- ▶ **conjunction** $C_1 \sqcap C_2$,
- ▶ **existential restriction** $\exists R.C$
 - ★ $\exists Y (R(X, Y) \wedge C(Y))$
- ▶ **negation** $\neg C$.

וכנגזרת גם את הפעולות

- ▶ **disjunctions** $C_1 \sqcup C_2 (\equiv \neg (\neg C_1 \sqcap \neg C_2))$,
- ▶ **value restrictions** $(\forall R.C \equiv \neg (\exists R. \neg C))$,
- ▶ $\top (\equiv A \sqcup \neg A)$ and $\perp (\equiv A \sqcap \neg A)$.

שאלה 2:

נתון בסיס נתונים הסתברותי שבו טבלה של שחקני כדורגל הכוללת שם, מדינה בה נולד וקבוצה בה הוא משחק, וטבלה של קבוצות כדורגל הכוללת את שם הקבוצה, עיר ומדינה בה הקבוצה נמצאת:

Players	Player	Birthplace	Team	P
	Ivan Rakitic	Spain	Real	0.5
	Kiril Nababkin	Russia	CSKA	0.9
	Aleksandrs Cauna	Latvia	CSKA	0.6
	Nikola Mitrovic	Serbia	Shanghai Shenxin	0.4
	Dennis Cheryshev	Russia	Real Madrid	0.5
	Alan Dzagoev	Russia	CSKA	0.7
Teams	Team	City	Country	P
	Real Madrid	Madrid	Spain	0.7
	CSKA	Moscow	Russia	0.8
	Shanghai Shenxin	Shanghai	China	0.9

השאלתה: "כל המדינות, בהן יש קבוצה שבה משחק שחקן שנולד באותה המדינה".

בשפה פורמלית: $Q(B) := \text{Players}(P, B, T), \text{Teams}(T, C, B)$

מה ההסתברות לקבל כחלק מהתשובה את ה- "Russia" tuple?

שאלה 3:

כתבו בפירוט (פסאודו קוד) אלגוריתם שבהינתן HMM (של תיוגי חלקי דיבר) ומשפט, מוצא את top-k התיוגים הסבירים ביותר של המשפט.

מה סיבוכיות האלגוריתם שכתבתם כפונקציה של גודל ה- HMM, אורך המשפט, ו- k ?