

גאורגיה במחשבותיי: ניתוח רשתות חברתיות פוגש עיבוד שפה טבעית
יובל פיינשטיין

SIGTRS,
10 באוקטובר 2024

- איך חוקרים תחום חדש?
- מהם מושגי המפתח?
- איך הייתם לומדים עליהם?

אתן כאן מתכון לכניסה וסיור בתחום חדש, ע"י שימוש בניתוח רשתות חברתיות ועיבוד שפה טבעית. ננצל את העובדה שרשתות חברתיות (גרפים) וטקסט מופיעים בתחומי מחקר ועיסוק רבים. ניתן לממש את המתכון או ב-Python או בכלי no-code (אביא קישורים בהמשך)

המתכון:

1. בחרי **תחום** שאת רוצה ללמוד עליו.
2. אתרי **מקור נתונים** שקשור לתחום.
3. חלצי **רשת חברתית** ממקור הנתונים.
4. זהו **צמתים חשובים** בעזרת מדדי מרכזיות.
5. **נקי והתמקדי** בחלקים מעניינים של הרשת בעזרת אלגוריתמי ליבה.
6. התמקדי ב**רשתות משנה** ע"י זיהוי קהילות ורשת עצמית.
7. הרחיבי טקסט מתוך **צמתים חשובים**.
8. חלצי **רשת חברתית מתוך הטקסט**.
9. אם רוצים ללמוד עוד, חזרי לסעיף 4.

מקורות נתונים אפשריים

- X(טוויטר)
- imdb
- Google Dataset Search
- Wikidata ו- DBpedia
- כל זכות
- ויקיפדיה
- נתוני המחקר שלך
- נתוני החברה/היועץ/הלקוח שלך

חילוץ רשת חברתית

נעשה ע"י זחילה על טקסט המקור, זיהוי ישויות וקישור ביניהן. בדוגמה שלנו, השתמשנו ב-API של ויקיפדיה לצורך חילוץ דפי ויקיפדיה וחלק מהדפים המקושרים אליהם.

מדדי מרכזיות

- PageRank
- Betweenness Centrality
- מאפשרים לזהות צמתים חשובים.

ביקוי הרשת

- התמקדות בחלקים מעניינים של הרשת בעזרת אלגוריתם K-core המשמר חלקים מקושרים היטב

התמקדות ברשתות משנה ע"י זיהוי קהילות ורשת עצמית.

נשתמש באלגוריתם Louvain לצורך זיהוי קהילות ברשת והסתכלות עליהן.
נשתמש בויזואליזציית Ego Network כדי לבחון צמתים חשובים ואת סביבתם.

הרחבת טקסט מתוך צמתים חשובים
בדוגמה שלנו, נעשית ע"י זחילה על דפי הויקיפדיה התואמים לצמתים

חילוץ רשת חברתית מתוך הטקסט
נעשה ע"י זחילה על טקסט המקור, זיהוי ישויות וקישור ביניהן. כאן נזהה ישויות ע"י ספריית spaCy

סיכום
שילוב טכניקות של ניתוח רשתות חברתיות ועיבוד שפה טבעית מאפשר לנו לעשות סקירה ראשונית של תחום עניין חדש.

קישורים

Python

[Network Science with Python \(book\)](#)

[days of networks blog 100](#)

[days of networks github page 100](#)

[/https://networkx.org](https://networkx.org)
[/https://spacy.io](https://spacy.io)
[/https://pypi.org/project/wikipedia](https://pypi.org/project/wikipedia)

No Code

[Netfrinx - פודקאסט רשתות של אסף שפירא](#)
[Gephi - ויזואליזציית רשתות](#)

[בניית רשתות ללא תכנות](#)
/