

הצעת פרוייקט במדעי הרוח הדיגיטליים – ירין קגל

האוסף הישראלי שבחרתי הינו מרכז המידע לאמנות ישראלית. האוסף מתאר מידע על כ- 6,000 אמנים ישראלים, לכל אחד משויך דף המתאר רקע כללי על חייהם – איפה נולדו, מתי התחתנו, מתי פתחו גלריות וכו', פרטים לימודיהם ועל קריירת ההוראה שלהם, ופרסים שזכו בהם.

מאפיינים:

1. נגישות- האוסף נגיש מאוד, בפשטות ניתן לגשת לאתר לפי url, לבקש גישה לרשימת כל האומנים ששם משפחתם מתחיל באות כלשהי (ספציפית) ולקבל רשימה של כולם (מופיע באוסף כלינק, גישה תתבצע באמצעות selenium לאלמנט של כל אחד מהאומנים / גישה ל-class של רשימת האומנים או באמצעות משיכת הנתונים מקבצי ה-XML).

בנוסף האוסף נגיש מאוד למשתמש, קל מאוד לגשת לכל אומן שהמשתמש מעוניין לצפות בו ולהגיע לדף שלו. האוסף מסודר לפי אותיות הא"ב, כדי להגיע לאומן נלחץ על האות ששם משפחתו מתחיל בה, ובצד ימין של האתר תופיע רשימת האומנים ששם משפחתם מתחיל באות זו, נגלול (לפי סדר הא"ב) לשם הפרטי של האומן ונלחץ על השם שלו.

2. פתיחות- ניתן בקלות לגשת לרשימת השמות בכל דף של אות במאגר (א'-ת') ומכאן לשלוף את השמות ולהיכנס לדף של האמן. באמצעות named entity recognition אוכל לבצע גישה לאנשים המקושרים לאומן ולמקומות (מקום לידה וכו') ומכך להציג את הנתונים בצורה נוחה למשתמש.

3. סטנדרטים- באוסף אין סטנדרטים קבועים או מינימליים להצגת המלל. קיימים אומנים שהמלל שמוצג לגביהם הוא מקום ושנת הולדתם, או אמנים שהפירוט לגביהם לא מציג שיש להם תערוכות אך ניתן ללחוץ על

קישור לצפייה בתערוכות שלהם לכן מבחינת הגישה לנתונים מהקוד יהיה צורך לתמוך בתוכן שחלקים ממנו קיימים וחלקים אחרים לא קיימים, לכל גישה לאמן.

4. אחידות- קיימת אחידות כלשהי האוסף אך לא אחידות מושלמת. למשל לאומנים מסויימים יש תמונה ולחלק אין. לחלק יש סעיף של הוראה ולחלק אין (היינו מצפים שיהיה ריק), אצל חלק מהאומנים ניתן לצפות בעבודות – קיים לינק קישור לעבודות, ואצל חלק לא קיים ועוד פרמטרים נוספים שמראים על חוסר אחידות בין הפריטים באוסף.

הגדרת המטרה:

בדיקת קורלציה בין מקום הולדתם של האומנים, מקום הלימודים שלהם ומקום ההוראה שלהם (מרכז, פריפריה, חו"ל), לבין כמות התערוכות והעבודות שלהם כפי שמוצגות באוסף.

הגדרת ה- data:

הדאטה מחולק לפי אותיות הא"ב, המייצגות את שם המשפחה של האומן, "בתוך" כל אות ישנה רשימה של כל האומנים ששם משפחתם מתחיל באות.

שדות התיאור- אות שבה השם משפחה מתחיל, שם האומן, לכל אומן יש דף אומן ובו שדות נוספים: תמונה, עבודות, תערוכות, רשימת סרטים ומאמרים, פסקת הסבר על האומן, לימודים, הוראה, פרסים, פסלים במרחב הציבורי .

סיכום פרוייקט במדעי הרוח הדיגיטליים – ירין קגל

- בחרתי להציג בפרוייקט שלי ניתוח נתונים לגבי אמנים ישראלים כפי שמוצגים באתר של מרכז מידע לאמנות ישראלית.
- הפרוייקט מסכם מידע של יותר מ-6100 אמנים ישראלים. ניתחתי את המידע והצגתי נקודות מעניינות לגבי מקום הולדתם/ מחייתם של האמנים והמגדר שלהם לעומת הכרה שצברו כמו פרסים שזכו בהם וכד'.
- על מנת ליצור את הדאטה לקראת הניתוח, השתמשתי בקוד Python שכתבתי ונעזרתי ב – selenium על מנת לבצע סריקה של הנתונים מהאתר, ולהעביר אותם לקובץ EXCEL.
- כדי למצוא את עיר / מקום הולדת של האמנים השתמש בעיבוד שפה טבעית בעברית (HebrewNLP) כדי לנתח את המידע שמופיע על כל אמן ולמצוא את העיר / מדינה ממנה הוא מגיע. (יש לציין שהכלים שבאתר עדיין בשלב בטא לכן היו מספר מצומצם של תיקונים ידניים שהיו צריכים להתבצע).
- לאחר שהיו לי את כל הנתונים מהאתר, השתמש בחבילות של Pandas ו- Geopy כדי לבצע הכנה טובה יותר לנתונים לקראת הצגת הנתונים ב- Tableau (המרת המיקומים בהם חיים / נולדו האמנים לכדי קוארדינטות על מנת להציגם על מפה ועוד) .
- יצרתי 2 מסכי תצוגה ב – Tableau:
 1. Map View – מסך שמציג על מפה את פריסת האמנים: בישראל לפי ערים, ובעולם לפי מדינות. כמות האמנים מכל מקום מיוצגת על המפה כעיגול (ככל שהעיגול יותר גדול, כך כמות האנשים גדולה יותר) בצבעים- ערים במרכז(כחול) , ערים בפרפריה (אדום), ומדינות מחו"ל (ירוק). המסך גם מציג גרף של כמות האנשים המגיעים מכל מקום, ובנוסף כאשר לוחצים על מקום(עיר/מדינה) מסוים, נפתחת רשימה של אמנים שמגיעים מאותו מקום, וכשנלחץ על אמן ספציפי תיפתח רשומה של כל הנתונים אודותיו.
 2. Gender and Origin Dashboard – מסך שמציג השוואות בין נשים לגברים בנושאים כמו מספר פרסים בהם זכו, מספר תערוכות ועבודות שמוצגות באתר, ומציג בנוסף גרף המראה את אחוז האנשים שזכו בפרסים מכל אזור (מרכז/ פריפריה/ חו"ל).

- המידע שונה מזה שהיה לפני העבודה

בכך שניתן לראות בצורה ברורה השוואות בין גורמים שונים, גרפים המציגים הבדלים / שוויון בתחומים בהם המידע שהיה באתר מדבר על כל פרט בנפרד, כאשר בפרויקט הצגתי את ההתייחסות לקבוצות (מקור , מגדר). בנוסף הפרויקט הרבה יותר ויזואלי וקל להבנה גם למי שלא מבין בדאטה ורוצה להבין ולהסיק מסקנות לגבי אמנים ישראלים.

מקורות וכלים:

1. מקורות:

א. מרכז מידע לאמנות ישראלית -

[/https://museum.imj.org.il/artcenter/newsite/he](https://museum.imj.org.il/artcenter/newsite/he)

ב. עיבוד שפה טבעית בעברית – HebrewNLP –

[/https://hebrew-nlp.co.il](https://hebrew-nlp.co.il)

2. כלים:

א. Python

ב. ספריות פייתון:

(1 Selenium

(2 Pandas

(3 Geopy

ג. Tableau