

Archie

Computerized Textual Artifact Management System

מנחה: ד"ר אמיר ספיר

מגישים: מיכאל כהן, גל אברהמי

רקע:

כיום קיימים ארכיונים רבים לא ממוחשבים. נקודה זו יוצרת בעיה למתעניינים במסמכי הארכיון. בנוסף, ישנם מוצגים לא מפוענחים אשר לא נוחים לקריאה ועיון. קיים קושי בחיפוש אחר מסמכים או אחר תוכנם.

תיאור המערכת:

- כללי: המערכת מיועדת למחקר, ולשם כך מאפשרת קטלוג מסמכים לפי מספר מאפיינים, ובנוסף – אחסון תמונת המסמך ופענוחו לקובץ טקסט. המערכת מיועדת לחוקרים ולאנשי ארכיון, וכדי להבטיח ניהול תקין של המשתמשים – גם מנהל מערכת. להלן פירוט האפשרויות לפי סוג משתמש:
 - עובד ארכיון: העלאת מסמך חדש; עדכון העלאת מסמך קיים; חיפוש מסמכים; עדכון טבלאות כמו קטגוריות.
 - חוקר: חיפוש מסמכים
 - מנהל מערכת: הוספת משתמשים והסרתם; חיפוש מסמכים.
- פעולות עיקריות:
 - הכנסת מסמך חדש לארכיון: כאשר משתמש מעלה קובץ תמונה לאתר, המסמך יומר לקובץ טקסט אשר יקוטלג לפי מילות מפתח, מחבר המסמך, כותרת המסמך וקטגוריה.
 - בניית עץ קטגוריות: הקטגוריות בנויות בצורת עץ ובכך מתאפשר לנהל תתי קטגוריות. כאשר התבצע חיפוש מסמך לפי קטגוריה, מתבצע שימוש באלגוריתם

העובר על עץ הקטגוריות ומחזיר את תוצאות כל המסמכים המתויגים תחת קטגוריה

.12.

- חיפוש מסמך: מתאפשר חיפוש מסמך לפי קטגוריה, מילות מפתח, מחבר המסמך ושם המסמך. לאחר ביצוע החיפוש המערכת תשלוף את תמונות המסמכים מבסיס הנתונים ותציג אותם בממשק web.

אמצעים טכנולוגיים:

- שפות תכנות:
- צד שרת: PHP, AJAX, XAMPP
- צד לקוח: JAVA, HTML, JS
- מערכת הפעלה: Windows, IOS, Linux
- סביבת פיתוח:
- Eclipse: שימוש בספריית OCR
- PHPStorm : פיתוח ממשק web
- OCR