****

**מחלקת הנדסת תוכנה**

שם הפרויקט: סוכן רשת אוטומטי

Project Name: Automatic Web Agent

הצעת פרויקט

|  |  |
| --- | --- |
| שם הסטודנט: | קובי הרשקוביץ |
| מספר תעודת זהות: | 025632654 |
| שם המנחה: | אלעד הוגן |
| תאריך ההגשה: | 25/05/09 |

**תוכן ענינים**

[מבוא 4](#_Toc229036130)

[מטרות, יעדים ומדדים 5](#_Toc229036131)

[סקירה ספרותית 6](#_Toc229036132)

[סקירת מצב קיים 7](#_Toc229036133)

[מסמך דרישות ראשוני ותיחום של המערכת 7](#_Toc229036134)

[גורמי הפרויקט 7](#_Toc229036135)

[תכלית הפרויקט 7](#_Toc229036136)

[לקוחות הפרויקט 7](#_Toc229036137)

[משתמשי הפרויקט 7](#_Toc229036138)

[אילוצי הפרויקט 7](#_Toc229036139)

[אילוצים ראשיים/קריטיים 7](#_Toc229036140)

[הגדרות שמות, קיצורים וראשי תיבות 7](#_Toc229036141)

[עובדות והנחות הקשורות לפרויקט 8](#_Toc229036142)

[דרישות פונקציונאליות 8](#_Toc229036143)

[תחום העבודה 8](#_Toc229036144)

[תחום המוצר 8](#_Toc229036145)

[דרישות מידע ופונקציונאליות 8](#_Toc229036146)

[דרישות לא פונקציונאליות 8](#_Toc229036147)

[דרישות עיצוב 8](#_Toc229036148)

[דרישות שימוש וממשק משתמש 8](#_Toc229036149)

[דרישות ביצועים 8](#_Toc229036150)

[דרישות הפעלה 8](#_Toc229036151)

[דרישות תחזוקה ותמיכה 9](#_Toc229036152)

[דרישות אבטחה 9](#_Toc229036153)

[דרישות תרבותיות ופוליטיות 9](#_Toc229036154)

[דרישות חוקיות 9](#_Toc229036155)

[נושאי הפרוייקט 9](#_Toc229036156)

[נושאים פתוחים 9](#_Toc229036157)

[פתרונות מדף 9](#_Toc229036158)

[בעיות חדשות 9](#_Toc229036159)

[משימות 9](#_Toc229036160)

[העברת המערכת לידי/לניהול הלקוח 9](#_Toc229036161)

[סיכונים 9](#_Toc229036162)

[עלויות 9](#_Toc229036163)

[תיעוד ולימוד משתמשים 10](#_Toc229036164)

[דרישות לדור הבא 10](#_Toc229036165)

[רעיונות לפתרון 10](#_Toc229036166)

[ניתוח פונקציונאלי ראשוני 10](#_Toc229036167)

[ניתוח חלופות מערכתי ראשוני 13](#_Toc229036168)

[תאור דרך הביצוע המתוכנן 13](#_Toc229036169)

[האמצעים/הכלים הנדרשים 13](#_Toc229036170)

[פערי ידע שעל הסטודנט להשלים 14](#_Toc229036171)

[תוצרי הפרויקט 14](#_Toc229036172)

[תכנית בדיקות ראשונית 15](#_Toc229036173)

[ניהול סיכונים 15](#_Toc229036174)

[תוכנית ניהול הפרויקט 16](#_Toc229036175)

[Work Breakdown Structure 16](#_Toc229036176)

[תוכנית עבודה של הפרויקט 17](#_Toc229036177)

[רשימת מקורות: 17](#_Toc229036178)

# מבוא

כיום השימוש באינטרנט נעשה רווח כמעט בכל בית ובאמצעותו ניתן לבצע משימות רבות ומגוונות החל מקבלת מידע בידורי וכלה בפעולות בנקאיות. במרבית מן הפעולות נדרש המשתמש להזדהות, לבצע מעקב אחר נתונים באתר, ולבצע פעולות על סמך ידע שלמד מאתרים אחרים. פעולות אלו במרבית הזמן זהות/דומות בין האתרים וכמו כן ביצוען עלול לדרוש תשומת לב וזמן רב מן המשתמש.

הפרוייקט המוצע הינו סוכן רשת אוטומטי הבא להחליף מטלות שגרתיות וגוזלות זמן של משתמש דפדפן ממוצע/מתקדם. באמצעות המערכת המוצעת יוכל המשתמש לתכנן תסריט של פעולות מסויימות באתרים מסויימים בהתאם למידע הנדגם מאתרים אחרים ולקבוע את זמן ביצוע התסריט.

לדוגמא ניתן לקבוע תסריט שבו נדגם כל 4 שעות אתר סקי שהמשתמש מחבב במיוחד (לדוגמא www.champery.com) , התסריט מכוון לזהות את השדה של גובה השלד באתר וברגע שגובה השלג מגיע ל-110cm התסריט שולף מאתר ה-"דקה ה-90" את מחירי חבילות הנופש לאתר ושולח דוח מפורט עם הנתונים באי-מייל למשתמש.

דוגמא נוספת יכולה להיות תסריט שיאריך את זמן השאלת ספר באתר הספריה של מכללת אפקה ,שאינו מצוייד כיום בסוכן אוטומטי להארכת השאלות, פעם בשלושה ימים.

יחודיות הפרוייקט היא שבניגוד לסוכנים אוטומטיים הנמצאים באתר מסויים ונשלטים מתוך האתר הפרוייקט הנ"ל מציע סוכן שיתוכנת בצד הלקוח ולא בצד השרת, דבר אשר יאפשר לו נוחויות ושליטה מירבית להשגת מטרותיו.

# מטרות, יעדים ומדדים

**מטרות:**

רבות מהפעולות הפיננסיות, הלימודיות והבידוריות נעשות באמצעות דפדפן ברשת האינטרנט:

* רישום לאתרים
* קניה/מכירה של מניות
* הזמנת כרטיסים
* הגשת מטלות
* בחירת קורסים

רבות ממטלות אלו הן מחזוריות, שגרתיות ועלולות להתיש את המשתמש ו/או לגרום לו לבצע שגיאות באופן הכנסת הנתונים.

מטרת הפרוייקט הנגזרת מבעיה זו היא בניית אפליקציה אשר תבצע מטלות שגרתיות וצפויות מראש בצורה אוטומטית במקום משתמש דפדפן אנושי

**יעדים:**

ניתן לפרוט מטרת על זו ל-3 יעדים:

* ניטור אתר בצורה מחזורית
* ביצוע פעולות אוטומטיות בדפדפן
* התראה למשתמש ע"פ בקשתו על אירוע באתר מסויים.

**מדדים:**

המדדים ע"פ תימדד הצלחת הפרוייקט הם:

* אחוז האתרים הנתמכים מתוך עשרת האתרים הראשונים בחיפוש בגוגל יעמוד על 80% בקטגוריות פיננסים, בידור, ספורט (דוגמא לחיפוש יכולה להיות המילה 'פיננסי')
* נוהל כניסה לאתר הספרייה במכללה והארכת השאלה אם קיימים ספרים מושאלים יסתיים בהצלחה.
* נוהל דגימת אתר לשינויים בתדירות של 6 פעמים בדקה יסתיים בהצלחה (ידווח בצורה נכונה על שינוי).
* משתמש חדש יוכל להגדיר תסריט אוטומטי של דגימת 2 אתרים ואינטראקציה ביניהם בתוך כשעה.

# סקירה ספרותית

סוכני רשת הם מערכות תוכנה מורכבות אשר פועלות ברשת האינטרנט העולמית (www), ברשת הארגונית או ברשתות מקומיות. סוכני רשת מתוכננים לבצע מגוון משימות החל מאגירה וניתוב של מידע וכלה בחיפוש קיטלוג וסינון מידע.

הספרות מתארת מסגרת לתכנון וניתוח של סוכני רשת ומערכות סוכנים בהתבסס על מודלים מתמטיים של הסביבה בה הם פועלים. יש 3 אבני בניין מרכזיות לבניית סוכני רשת:

* מודלים סטוקאסטיים של גרף הרשת אשר משקללים את הפילוג והקישוריות של דפי רשת ונותנים קוים מנחים מרכזיים בתכנון סוכנים, ע"י הפקת תכונות ייחודיות של הסביבה בה הסוכנים פועלים.

הרשת יכולה להיראות כגרף מכוון אשר הצמתים בו הם דפי HTML סטטיים והקשתות בו הם קישורים מאתר אחד למשנהו. מאמר [ ‎18] מציע ומנתח מודלי גרפיים אקראיים שנוצרו ע"י תצפיות נסיוניות ברשת. המודלים שונים מהמודלים המקובלים בכך ש קשתות שנבחרו בצורה בלתי תלויה לא משתקפות בסטטיסטיקה שנצפתה ברשת, כך שקשתות במודלים אלו הינם תלויות סטטיסטית ביניהן. כמו כן נוספים צמתים לגרף כתלות בזמן. דבר זה משקף את העובדה שהרשת משתנה עם הזמן.

* שיטות למידה סטטיסטית המאפשרות לסוכנים ללמוד על סביבתם ע"י הפקת מודלים סטוקאסטיים של דף רשת ומבנה קישורים מקומיים. מאמר [ ] מתאר את הרשת כעולם שבו המוני קישורים מחברים בין מסמכי ודפי אתרים כך שנוצר גרף בעל סיבוכיות גבוהה וללא תקדים וזהו גרף הרשת. מאמר זה מציג גישה חדשה ללימוד מודלים הסתברותיים שך הרשת אשר יכולים לשמש כחלק בתהליך חיזוי קישוריות ומידע של מסמכי רשת. השיטה המתוארת במאמר זה היא שיטה הסתברותית לצמצום מימדים (probabilistic dimension reduction) אשר גוזרת ומאחדת את שיטת הניתוח הסמנטי המאוחר (Latent Semantic Analysis) ואת אלגוריתם קליינברג (Hubs-and-Authorities) בסביבה סטטיסטית.

דבר זה אמור להיות הצעד הראשון לפיתוח תשתית סטטיסטית של טכנולוגיות מידע ברשת. למרות שמאמר זה לא מתמקד ביישום מסויים, מגוון אלגוריתמים הפועלים בסביבת רשת יכולים לנצל את היתרונות בטכניקות המוצגות למטרות שונות כגון מנועי חיפוש, "זוחלי רשת" (Web crawlers) ומערכות סוכני מידע.

* אלגוריתמים לתכנון אוטונומי וקבלת החלטות בסביבת רשת לאפשר לסוכנים להשיג מטרות ולהסתגל לסביבה משתנה [16]

ניקח לדוגמא בעיות קלאסיות בתחום בסיסי נתונים/איחזור מידע כגון חיפוש בטקסט, כריית מידע וסיווג. הקלט לבעיות מסוג זה הוא בדרך כלל אוסף של נתונים/מסמכים. הרשת עם המבנה הנוסף שלה כגרף, מאפשרת את השיפור של טכניקות קיימות עם טכניקות מעולם הגרפים. במאמר הודגם שיפור זה באמצעות פתרונות מבוססי גרפים לבעיות הבאות: חיפש נושא, מיספור נושא, סיווג וזחילה (crawling).

# סקירת מצב קיים

כיום כאשר משתמש רוצה לבצע מעקב/שינוי נתונים באתר מסויים פרושות לפניו מספר אפשרויות:

תמיכה של האתר עצמו בסוכן רשת אשר מיידע את המשתמש לגבי נתון ייחודי לאתר אליו הוא שייך. לדוגמא האתר ebay מציע סוכן שיציע במכרז מחיר עבור פריט מסויים עד תקרה שיקבע המשתמש.

מעקב מחזורי אישי של המשתמש אחר שינויים באתר.

שימוש בתוכנות אוטומציה יקרות כדוגמת QFT או RFT.

QTP(Quick Test Pro ) היא חבילת תוכנה מבית מרקורי שבה ניתן לבצע אוטומציה לאפליקציות חלונאיות ורשתיות כגון אוטומציה לכתיבת דואר אלקטרוני ושליחתו או חקירת דף רשת והכנסת נתונים אליו.

RFT (Rational Functional Tester) היא גם חבילת תוכנה לבדיקות אוטומטיות מבית IBM המאפשרת גם לבצע בדיקות על אלמנטים מסוג Flex וטרמינלים כגון VT102.

מחירם של כל אחת מחיבולות אלו מגיע לכמה אלפי דולרים עבור רשיון למשתמש אחד.

# מסמך דרישות ראשוני ותיחום של המערכת

## ***גורמי הפרויקט***

### **תכלית הפרויקט**

להוות כלי למעקב, והתראה על שינויים באתרי אינטרנט וביצוע פעולות מתוכנתות מראש כתגובה למידע הנכרה מהם, באותם אתרים עצמם או אתרי מטרה אחרים.

### **לקוחות הפרויקט**

משתמשי דפדפן מתקדמים הרוצים לבצע פעולות אוטומטיות בדפדפן ללא מגע יד אדם.

### **משתמשי הפרויקט**

ראה לקוחות

## ***אילוצי הפרויקט***

### **אילוצים ראשיים/קריטיים**

תמיכה ב-IE בלבד.

### **הגדרות שמות, קיצורים וראשי תיבות**

|  |  |  |
| --- | --- | --- |
| שם/קיצור/ר"ת | פירוש באנגלית | פירוש בעברית |
| IE | Internet Explorer | דפדפן מבית מיקרוסופט |
| HTML | Hyper Text Markup Language | שפה לתיאור טקסקט עשיר |
| DOM | Document Object Model | מודל אובייקט מסמך |
| אלמנט | Element | פריט בדף HTML |
| URL | Uniform Resource Locator | מאתר משאבים אחיד |
| Hostname | Server name part of the URL | שם השרת ב-URL |
| WWW | World Wide Web | רשת האינטרנט העולמית |

### **עובדות והנחות הקשורות לפרויקט**

הדף הנבדק לא יכיל אובייקטי Flash ו-Java

למשתמש יש ידע בסיסי ב-HTML

הדפדפן יהיה מסוג IE.

## ***דרישות פונקציונאליות***

### **תחום העבודה**

רשת האינטרנט (WWW)/אינטרהנט.

### **תחום המוצר**

ניטור ואוטומציה של אתרים.

### **דרישות מידע ופונקציונאליות**

* בסיס נתונים לשמירת מיפוי אלמנטים בדף לאובייקטים, ולשמירת טבלת מצב-פעולה (trigger-action table)
* ממשק משתמש:
  + מיפוי אלמנטים בדף לאובייקטים בבסיס נתונים.
  + כתיבת תסריט פעולה.
  + קביעת תזמון ריצה.
* יכולת זיהוי אלמנטים בדף HTML.
* יכולת הכנסת נתונים אוטומטית בדף HTML.
* יכולת דגימה מחזורית של שדות מסויימים באתרים מסויימים
* יכולת התראה על אירוע למשתמש במצעות חלון /או דואר אלקטרוני
* המשתמש יוכל להגדיר באילו תנאים תבוצע פעולה, אותה יוכל להגדיר גם כן (לדוגמה לחיצה על כפתורים ומילוי שדות כאשר ערך של שדה מסויים ישתנה)

## ***דרישות לא פונקציונאליות***

### **דרישות עיצוב**

הממשק יעוצב כך שמשתמש דפדפן ממוצע יוכל להבין בצורה אינטואיטיבית את פעולתו.

### **דרישות שימוש וממשק משתמש**

הממשק יהיה מעוצב כממשק לשוניות כאשר:

* הלשונית הראשונה תכיל את ממשק מיפוי האלמנטים
* הלשונית השניה תכיל את ממשק כתיבת התסריטים
* הלשונית השלישית תכיל את ממשק תזמון והרצת התסריטים.

### **דרישות ביצועים**

המערכת תפעל במהירות גבוהה ממהירות עבודה של משמתמש אנושי.

### **דרישות הפעלה**

* חיבור לרשת.
* IE מותקן.

### **דרישות תחזוקה ותמיכה**

תיקוני באגים ובוצעו ע"פ דרישות הלקוח.

### **דרישות אבטחה**

סיסמאות ופרטים אישיים יוצפנו בבסים הנתונים

### **דרישות תרבותיות ופוליטיות**

NA

### **דרישות חוקיות**

NA

## ***נושאי הפרוייקט***

### **נושאים פתוחים**

* האם יוכלו לרוץ כמה תסריטים במקביל
* האם להכניס מנגנון שיתמוך בשינויים באתר/ים

### **פתרונות מדף**

* Newbie (<http://www.newbielabs.com/webautomation.html>)
* iMacros (<http://wareseeker.com/publisher/imacros-website-testing-and-we/38865/>(
* iRobot (<http://www.irobotsoft.com>)
* Selenium (<http://seleniumhq.org/>)
* Rational Functional Tester (<http://www.ibm.com/software/awdtools/tester/functional/index.html>)
* QTP (<http://mercuryquicktestprofessional.blogspot.com/>)

### **בעיות חדשות**

NA

### **משימות**

המשימות הנדרשות מתוארות ב-WBS ובתוכנית העבודה.

## ***העברת המערכת לידי/לניהול הלקוח***

מוצר מדף – יתבצע באמצעות תיעוד ודוגמאות שימוש.

## ***סיכונים***

ראה ‏13 (ניהול סיכונים)

## ***עלויות***

פרוייקט תוכנה בלבד – ההשקעה היא זמן פיתוח בלב

## ***תיעוד ולימוד משתמשים***

התיעוד לפרוייקט יהיה בהערות בגוף הקוד, בדוחות הביניים ובספר הפרוייקט.

## ***דרישות לדור הבא***

לא ידוע עדיין.

## ***רעיונות לפתרון***

ראה ניתוח פונקציונלי ראשוני לפירוט הרעיון

# ניתוח פונקציונאלי ראשוני

פעולת המערכת נחלקת לשלושה שלבים:

* איסוף/מיפוי האלמנטים ה-"מעניינים"

בשלב זה המשתמש גולש לאתר המטרה באמצעות הדפדפן המוטמע באפליקציה, ומסמן אלמנטים רלוונטיים בדף באמצעות העכבר. בלחיצה על מקש במקלדת ייטען השדה המסומן לטופס מיפוי האלמנטים באפליקציה. טופס זה מכיל מיפוי בין האלמנט ב-DOM לבין הייצוג הפנימי שלו במערכת אשר יישמש ביצירת התסריטים, וכן מידע נוסף על האלמנט כגון הטיפוס שלו, הערך הנוכחי שלו וכו'.

מיפוי המחדל עבור שם אובייקט יהיה ה-id של האלמנט, אך המשתמש יוכל לשנות זאת בטופס זה.

* הגדרת תסריט פעולה

בשלב זה המשתמש מגדיר תסריט פעולה המורכב מזוגות סדורים של תנאי קיום ופעולה (להלן יקרא זוג כזה "כלל"). תנאי הקיום יכול לבדוק ערך של אלמנט (המיוצג ע"י אובייקט) ואם הוא עומד בתנאי אז יתבצע חלק הפעולה של הכלל.

לאחר הגדרת כל כלל יתבצע מעבר בדיקה לוודא את תקינות הכללים בתסריט.

* תזמון/הרצה של תסריטים

בשלב זה בוחר המשתמש תסריט להרצה וניתנת לו האפשרות לקבוע מתי ירוץ התסריט ובאיזו מחזוריות.

מפת האלמנטים וכן התסריטים ישמרו בבסיס נתונים לצורך המשכיות (Persistence).

## ***תרשימי DFD***

תרשים הקשר:



תרשים ‏6‑1 Context Diagram(DFD0)

מיפוי אלמנטים:



תרשים ‏6‑2 מיפוי אלמנטים

עריכת מפת אלמנטים:



תרשים ‏6‑3 עריכת מפת אלמנטים

עריכת תסריט:



תרשים ‏6‑4 עריכת תסריט

הרצת תסריט:



תרשים ‏6‑5 הרצת תסריט

# ניתוח חלופות מערכתי ראשוני

|  |  |  |  |
| --- | --- | --- | --- |
| נושא לחלופה | QTP | iMacros | RFT |
| מחיר($) | 35000-70000 | 50-500 | 6500-9000 |
| כיסוי פונקצינאלי | מושלם | חסרה שליטה בתזמון | מושלם |
| קלות שימוש | מסובך – יש ללמוד VBScript | פשוט | בינוני – יש לדעת JAVA ברמה בינונית/נמוכה. |
| Scripting | VBScript | Proprietary | JAVA |

# תאור דרך הביצוע המתוכנן

* הגדרת סכמה לבסיס הנתונים – יש להגדיר את מבנה הטבלאות כגון מבנה טבלת המיפוי ומבנה טבלאות התסריטים
* הגדרת שפת תסריט – יש להחליט על האופן בו המשתמש יתכנת תסריטים במערכת. האפשרויות הנשקלות הן שימוש באבני בניין קבועות ליצירת צעדי תסריט כך שכל צעד יורכב מבחירת מספר אלמנטים קבוע באמצעות פקדים חלונאיים. אפשרות נוספת היא כתיבת תסריט בשפת C# וניתוחו באמצעות מודול הידור תסריטים.
* חלוקת המערכת למודולים – רכיבי המערכת יחולקו למודולים הניתנים להכלה במחלקות.
* הגדרת המחלקות הנדרשות בכל מודול ומודול – כל מחלקה תכיל פונקציה לוגית בודדת.
* קידוד ובדיקת כל מודול בזה אחר זה- הקידוד יעשה בשפת C# ובדיקתו תיעשה ב-NUNIT.
* אינטגרציה של כל המודולים – בשלב זה תיבנה בפעם הראשונה כל האפליקציה כמערכת אחת.
* בדיקות – ביצוע בדיקות המערכת ווידוא שכל המטרות הושגו. כמו כן יזוהו כל הבאגים ויתועדו עבור שלב התיקון.
* תיקוני באגים ובדיקות חוזרות – ניפוי השגיאות עד לפעולה מיטבית של המערכת.
* תיעוד – כתיבת הדוחות וספר הפרוייקט. כמו כן התיעוד בתוך הקוד יעשה במהלך כתיבת הקוד.

# האמצעים/הכלים הנדרשים

**סביבת פיתוח:**

סביבת פיתוח VS2008.

דפדפן IE8

חיבור לרשת

**סביבת ריצה:**

חיבור לרשת

דפדפן IE8

.NET runtime framework 3.5

# פערי ידע שעל הסטודנט להשלים

לימוד הממשקים הנדרשים (כגון mshtml.dll, docview.dll, DOM) בעיקר מאתר MSDN

# תוצרי הפרויקט

**דוח ביניים** **1** – יכיל את כל אשר נדרש ע"פ ההגדרה במסמך "מבנה תוצרים – הנדסת תוכנה"

* + אבטיפוס.
  + תקציר
  + סקירה ספרותית
  + מטרות העבודה
  + סקר מצב קיים
  + ניתוח חלופות מערכתי
  + ניתוח חלופות טכנולוגיות
  + תיאור פיתרון נבחר
  + תוכנית בדיקות תוכנה
  + לוחות זמנים
  + ניהול סיכונים
  + רשימת מקורות.
  + מסמכים מצורפים SRD ,STP , SPMP. הצעת פרויקט.

**דוח ביניים 2** - יכיל את כל אשר נדרש ע"פ ההגדרה במסמך "מבנה תוצרים – הנדסת תוכנה"

* + תקציר
  + מטרות העבודה
  + סקירה ספרותית
  + אלטרנטיבות לתיכון המערכת
  + הפתרון שנבחר לתיכון המערכת
  + תוכנית בדיקות תוכנה
  + לוחות זמנים מעודכנים להמשך העבודה
  + הערכת ביצוע המשימות
  + ניהול סיכונים
  + רשימת מקורות
  + מסמכים: SRD, SDD, STD, SPMP. הצעת פרויקט.
  + אפליקציה מושלמת (קובץ התקנת אפליקציה, חוברת הפעלה).

**ספר פרוייקט** - יכיל את כל אשר נדרש ע"פ ההגדרה במסמך "מבנה תוצרים – הנדסת תוכנה"

* + יסכם את דו"חות הביניים
  + ישלב דרישות מערכת ואפיון מערכת
  + יפנה למסמכים שכתבנו בדוחות הביניים
  + ארכיטקטורת המערכת
  + דווח בדיקות תוכנה
  + דוגמאות להפעלת המערכת
  + יתאר ביצועים עלות ואמינות
  + יתאר את הצלחת המערכת או אי ההצלחה.
  + המסמכים שנעשו במשך העבודה על הפרויקט.

**מערכת**

* + קבצי התקנה
  + קבצי עזרה והדרכה למשתמש
  + אבטיפוס ראשוני
  + מצגת פרויקט
  + מערכת סופית

# תכנית בדיקות ראשונית

* בניית דפי דמה המכילים מגוון רב של סוגי אלמנטים וניסוי המערכת עליהם.
* בדיקת אוטומציה של חיפוש באתר הבית של גוגל.
* בדיקת ניטור שדה באתר ושליחת מייל כאשר שדה זה עובר ערך מסויים
* בדיקת מנוע יצירת תסריטים

# ניהול סיכונים

|  |  |  |  |
| --- | --- | --- | --- |
| סיכון | סיכוי | השפעה | דרך מניעה |
| חריגה מלוחות זמנים | 50% | אי הגעה לאבני דרך בפרוייקט | ניסיון להקדים ביצוע של משימות והקדשת זמן נוסף לפרוייקט |
| כשל חומרתי בסביבת הפיתוח/ריצה | 20% | עיכוב של עד כשבועיים | גיבוי כל איטרציה במערכת ניהול קוד |
| אי יכולת לזהות אלמנטים מיוחדים בדף | 50% | אי יכולת להשתמש באלמנטים אלו – ירידה בפונקציונאליות המובטחת | עדכון סעיף האילוצים בשלב מוקדם ככל האפשר |
| דפים דינמיים (Ajax, DHTML) | 80% | שיבוש יכולת זיהוי האלמנטים | התקנת מנגנוני המתנה לאלמנטים לא קיימים. |

# תוכנית ניהול הפרויקט

## ***Work Breakdown Structure***

## ***תוכנית עבודה של הפרויקט***



תרשים ‏14‑1

# רשימת מקורות:

1. Jon Kleinberg. **Authoritative sources in a hyperlinked environment**. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998. Extended version in Journal of the ACM 46(1999). Also appears as IBM Research Report RJ 10076, May 1997. [[pdf]](http://www.cs.cornell.edu/home/kleinber/auth.pdf)
2. Soumen Chakrabarti, Martin van den Berg, Byron Dom, **Focused Crawling: A New Approach to Topic-Specific Web Resource Discovery**, WWW8 [[html]](http://www8.org/w8-papers/5a-search-query/crawling/index.html)
3. Steve Lawrence and Lee Giles, **Searching the World Wide Web,** Science 1998 [[pdf]](http://www.neci.nj.nec.com/~lawrence/papers/search-science98/search-science98.pdf)
4. Jeffrey Dean, Monika R. Henzinger, **Finding Related Pages in the World Wide Web**, WWW8 [[html]](http://www8.org/w8-papers/4a-search-mining/finding/finding.html)
5. Oren Zamir and Oren Etzioni, **Grouper: A Dynamic Clustering Interface to Web Search Results**, WWW8, [[html]](http://www8.org/w8-papers/3a-search-query/dynamic/dynamic.html)
6. Rajan M. Lukose and Bernardo A. Huberman, **Surfing as a Real Option** [[abstract]](http://www.parc.xerox.com/spl/groups/dynamics/abstracts/InternetEcology/options.html)
7. Bernardo A. Huberman and Rajan M. Lukose, **Social Dilemmas and Internet Congestion** [[abstract]](http://www.parc.xerox.com/spl/groups/dynamics/abstracts/InternetEcology/storms.htm)
8. Bernardo A. Huberman, Peter L.T. Pirolli, James E. Pitkow, and Rajan M. Lukose, **Strong Regularities in World Wide Web Surfing**, Nature ??? [[abstact]](http://www.parc.xerox.com/spl/groups/dynamics/abstracts/InternetEcology/surfing.html)
9. William W. Cohen, Wei Fan, **Web-Collaborative Filtering: Recommending Music By Spidering the Web**, WWW9 [[html]](http://www9.org/w9cdrom/266/266.html)
10. Mike Perkowitz Oren Etzioni, **Towards Adaptive Web Sites: Conceptual Framework and Case Study**, WWW8 [[html]](http://www8.org/w8-papers/2b-customizing/towards/towards.html)
11. Marc Langheinrich, Atsuyoshi Nakamura, Naoki Abe, Tomonari Kamba, Yoshiyuki Koseki, **Unintrusive Customization Techniques for Web Advertising**, WWW8 [[html]](http://www8.org/w8-papers/2b-customizing/unintrusive/unintrusive.html)
12. Jim Hendler, **Is there an intelligent agent in your future?** Nature, Web matters, 11 March 1999. [[html]](http://helix.nature.com/webmatters/agents/agents.html)
13. Steve Kirsch, **The future of Internet search (keynote address)**, SIGIR 1999 [[pdf]](http://www.acm.org/pubs/articles/proceedings/ir/290941/p1-kirsch/p1-kirsch.pdf)
14. G. Pandurangan, P. Raghavan, and E. Upfal. Using PageRank to Characterize Web Structure, Proceedings of the 8th International Computing and Combinatorics Conference (COCOON), 2002.
15. [G. Pandurangan](http://www.cs.brown.edu/people/gopal) , [P. Raghavan](http://theory.stanford.edu/people/raghavan/index.html) , and [Eli Upfal](http://www.cs.brown.edu/people/eli), [**Building Low-Diameter P2P Networks**](http://www.cs.brown.edu/research/webagent/focs-2001.pdf). Proceedings of the 42th IEEE Symp. on Foundations of Computer Science. 2001.
16. S.R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and [Eli Upfal](http://www.cs.brown.edu/people/eli), [**The Web as a graph**](http://www.cs.brown.edu/research/webagent/pods-2000.pdf). Proceedings of the 19th ACM Symposium on Principles of Database Systems, pp 1-10, 2000.
17. Thomas Hofmann, [**Learning Probabilistic Models of the Web**](http://www.cs.brown.edu/research/webagent/Hofmann-SIGIR00.pdf), ACM SIGIR 2000
18. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and [Eli Upfal](http://www.cs.brown.edu/people/eli), [**Stochastic models for the Web graph**](http://www.cs.brown.edu/research/webagent/focs-2000.pdf). Proceedings of the 41th IEEE Symp. on Foundations of Computer Science. 2000.