

מערכת מעקב אחר דפי אינטרנט ודיוור התוכן ללקוחות

**שרה גורביץ | יהודית כהן |2019**

**שלמי תודה**

בפתח דבר נודה לבורא עולם שהביאנו עד הלום וסייע לאורך כל הדרך.

בנוסף נודה לכל אלו שנתנו רבות מכוחם, מרצם ומתבונתם במהלך העבודה.

🙘תודה **-** לגב' רבקה הנאו ולגב' שרה רבקה פוירשטין - מנחות, שסייעו לנו לאורך כל הדרך, החל משלב התכנון ועד לשלב הביצוע. ייעצו, בדקו ונתנו חוות דעת על מנת להגיע להישגים המקסימליים ולפרויקט ראוי ומושלם.

🙘 תודה - למרכזת גב' הדסה גרוס אשר השקיעה וטרחה רבות למעננו,

🙘תודה - לצוות המורות המסור על המקצועיות, העזרה הרבה, התמיכה וההשקעה.

🙘 לסמינר מרכז בית יעקב, ולעומדים בראשו הרב ישעיה' ליברמן, הרב ישראל ליברמן גב' מינה זלזניק על החינוך, המחשבה, המסירות והמקצועיות הרבה, וההשקעה לטווח הרחוק .

🙘 תודה - להורינו היקרים ולשאר בני המשפחה וידידים על תמיכה ועידוד לאורך כל הדרך החל מתקופת הלימודים ועד להגשת פרויקט גמר.

**תקציר**

המערכת מתחלקת לתת מערכות המממשות צרכים שונים של המנגנון השלם: -

* אתר המסופק ללקוחות ומאפשר רישום, ניהול תוכן ושירותים נגישים. האתר מבוסס Angular ומעדכן את מסד הנתונים של המערכת.
* מערכת .Net להפקת נתונים מאתרים, המתבססת על מסד נתונים sql ומשתמשת בגלישה רובוטית. מערכת זו מכילה אלגוריתמים מורכבים לייעול מהירות ואיכות. מערכת זו מתוזמנת ע"י Windows service.
* extension ל – Google Chrome מיועד לבחירת תוכן האתר ע"י הלקוח, קליטת סיסמאות, רישום פעולות וסינון מידע. התוסף נכתב ב html ו JavaScript..

הפיתוח כולל אפיון, תכנון מסד נתונים, כתיבת שאילתות, הקמת service , בניית אתר לתוכנה, פיתוח תוסף ל Chrome וכן עיצוב גרפי ב CSS.

**1 תיאור הפרויקט**

## 1.1 תיאור כללי

מעקב אחר שינוים ברשת הוא צורך מורכב. על מנת לבנות תוצר יעיל ונוח לשימוש, עלתה הדרישה לחלק את המערכת למספר רכיבים שמתפקדים יחד כמערכת.

* אתר אינטרנט
* מסד נתונים
* מערכת להפקת נתונים מאתרים, איתור שינויים ועדכון המשתמשים.
* extension ל – Google Chrome

משתמש המעוניין להצטרף לשרות נרשם באתר האינטרנט מגדיר פרטים, כתובת מייל ומוריד משם extension ל – Google Chromeהמבצע הקלטה של פעולות הגלישה בדפדפן.

כאשר המשתמש מוצא מידע עליו מעונין לעקוב הוא מעדכן את התוסף על כך וההקלטה נשמרת.  
המשתמש לא יסתפק בכתובת URL בלבד וידרוש הקלטה מכיוון שמעוניין במידע ספציפי שלא מגיע מידית בטעינת האתר אלא לאחר מספר פעולות.

באתר האינטרנט המשתמש מגדיר באילו תדירויות הוא מעוניין להתעדכן, וכן יכול לנהל את הבקשות הקודמות.

נתוני המשתמש והקלטות הגלישה נשמרים במסד נתונים SQL בשרת, הקבצים הן של ההקלטה והן של תוכן הדף נשמרים בטבלה מיוחדת FileTable.

בשרת מתוזמנת בתדירות קבועה מערכת שתפקידה לבצע בפועל מעקב אחר דפים ברשת.  
המערכת שולפת את נתוני הגלישה של הבקשות שמוגדרות לתזמון הנוכחי, מבצעת גלישה רובוטית ומקבלת את דף הHTML המעודכן.  
המערכת משווה בין הגרסאות בודקת אם היו שינויים רלוונטיים ומדגישה אותם בדף הHTML כדי לחסוך קריאה של הדף כולו.  
המערכת שולחת את דפי הHTML הנ"ל לחשבון המייל של המשתמש.

**2 מפרט טכני**

## 2.1 מפרט לעמדת משתמש

* דפדפן Google Chrome
* חיבור לרשת
* חשבון מייל

## 2.2 מפרט לעמדת שרת

* שכבת מסד הנתוניםSQL Server 2014 -
* Imacros
* חיבור לרשת
* Windows service

## 2.3 כלי התוכנה לפיתוח המערכת

Visual Studio 2017

Visual Studio Code

## 2.4 שפת התכנות

* C#
* Angular

**3 יעדי המערכת**

מטרת המערכת לבצע מעקב אחר דפי אינטרנט ודיוור התוכן ללקוחות באופן יעיל ונגיש. לצורך כך הוגדרו היעדים הבאים:

* ניהול נתוני המשתמשים
* הקלטת פעולות הגלישה בדפדפן
* הפעלת המתזמן – Windows service בשרת.
* גלישה רובוטית לקבלת נתונים מעודכנים
* השוואת גרסאות בצורה יעילה אמינה ואיכותית
* הדגשת שינוים בדפי HTML
* שליחת נתונים עדכניים ללקוח

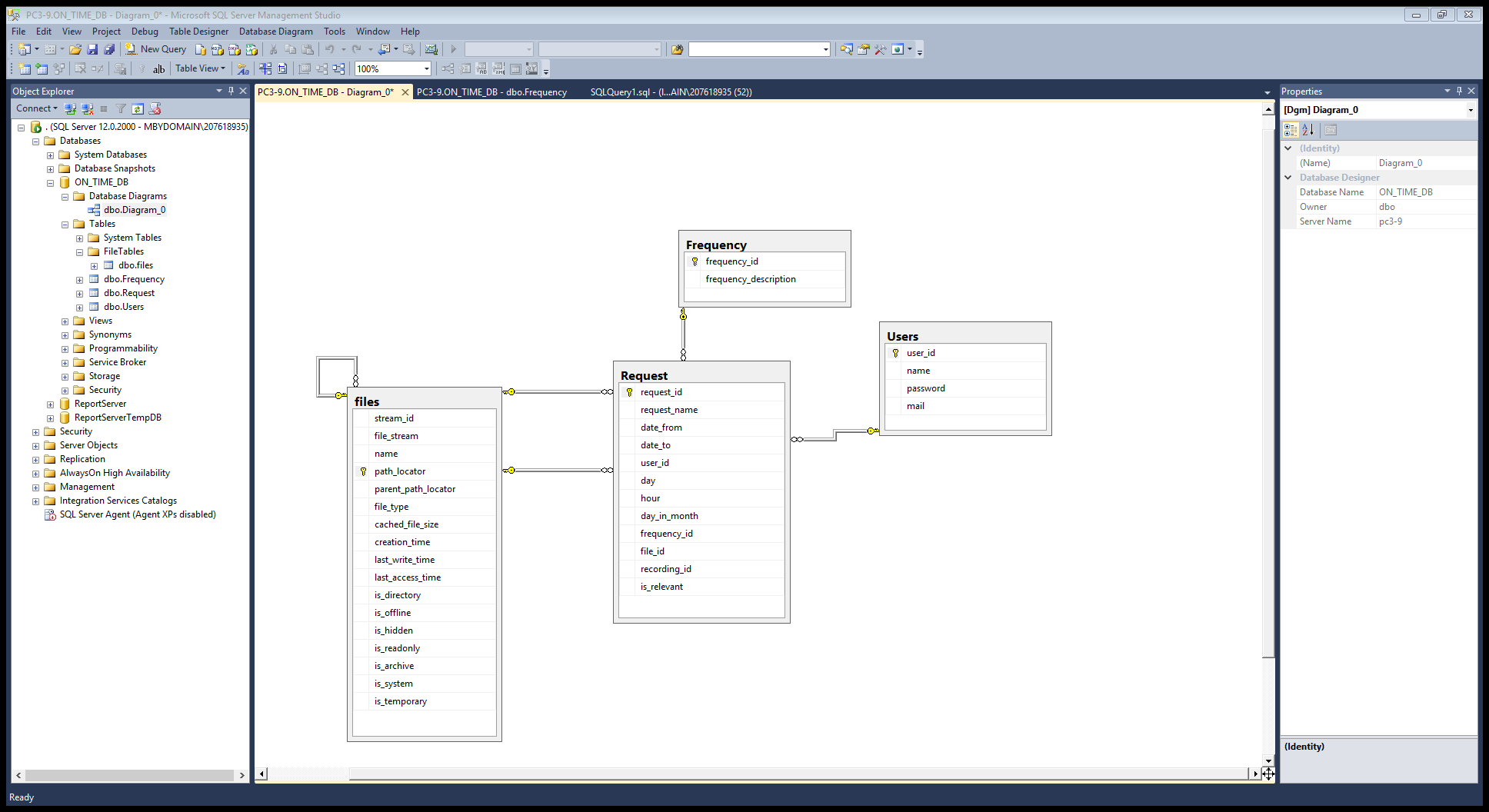
**4 תיחום המערכת**

* הקלטת פעולות הגלישה תתבצע בדפדפן Google Chrome בלבד.
* המערכת לא מטפלת בדפי HTML שאינם תקינים.
* המערכת אינה מתחיבת ל100 אחוזי הצלחה, בדיוק זיהוי השינויים.

**5 אלגוריתם מרכזי**

**6 המודלים והקשרים ביניהם**

6.1 Data Diagram



6.2 Database

**Users** – טבלת משתמשים  
בטבלה זו שמורים כל המשתמשים

|  |  |  |  |
| --- | --- | --- | --- |
|  | שם | סוג | הסבר |
|  | user\_id | int | קוד משתמש - מספר רץ |
|  | name | string | שם |
|  | password | string | סיסמה |
|  | mail | string | מייל |

**Frequency**– תדירות לבקשה

בטבלה זו שמורים רמות תדירויות בהם יבדקו שינויים באתר.

|  |  |  |  |
| --- | --- | --- | --- |
|  | שם | סוג | הסבר |
|  | frequency\_id | int | קוד תדירות |
|  | Frequency\_description | string | תיאור תדירות |

**Request**– טבלת בקשות שרות  
בטבלה זו שמורים פרטי בקשות השרות של המשתמש.

|  |  |  |  |
| --- | --- | --- | --- |
|  | שם | סוג | הסבר |
|  | request\_id | int | קוד בקשת שרות – מספר רץ |
|  | user\_id | int | קןד משתמש |
|  | request\_name | string | תיאור הבקשה |
|  | date\_from | date | תאריך תחילת הבקשה |
|  | date\_to | date | תאריך סיום הבקשה |
|  | frequency\_id | int | קוד תדירות |
|  | hour | int | שעה מסוימת ביום |
|  | day | int | יום מסוים בשבוע |
|  | day\_in\_month | int | יום מסוים בחודש |
|  | file\_id | Guid-uniqueidentifier | קישור לקובץ של תוכן הדף |
|  | recording\_id | guid | קישור לקובץ ההקלטה |
|  | Is\_relevant | Bit-bool | האם הבקשה רלונטית / מחוקה |

**FileTable**– קבצים  
בטבלה זו שמורות הקלטות והדפים של המשתמשים.

ישנן שתי דרכים כיצד לשמור קבצים : לשמור את הקבצים במחשב המקומי ובsql לשמור נתיב.

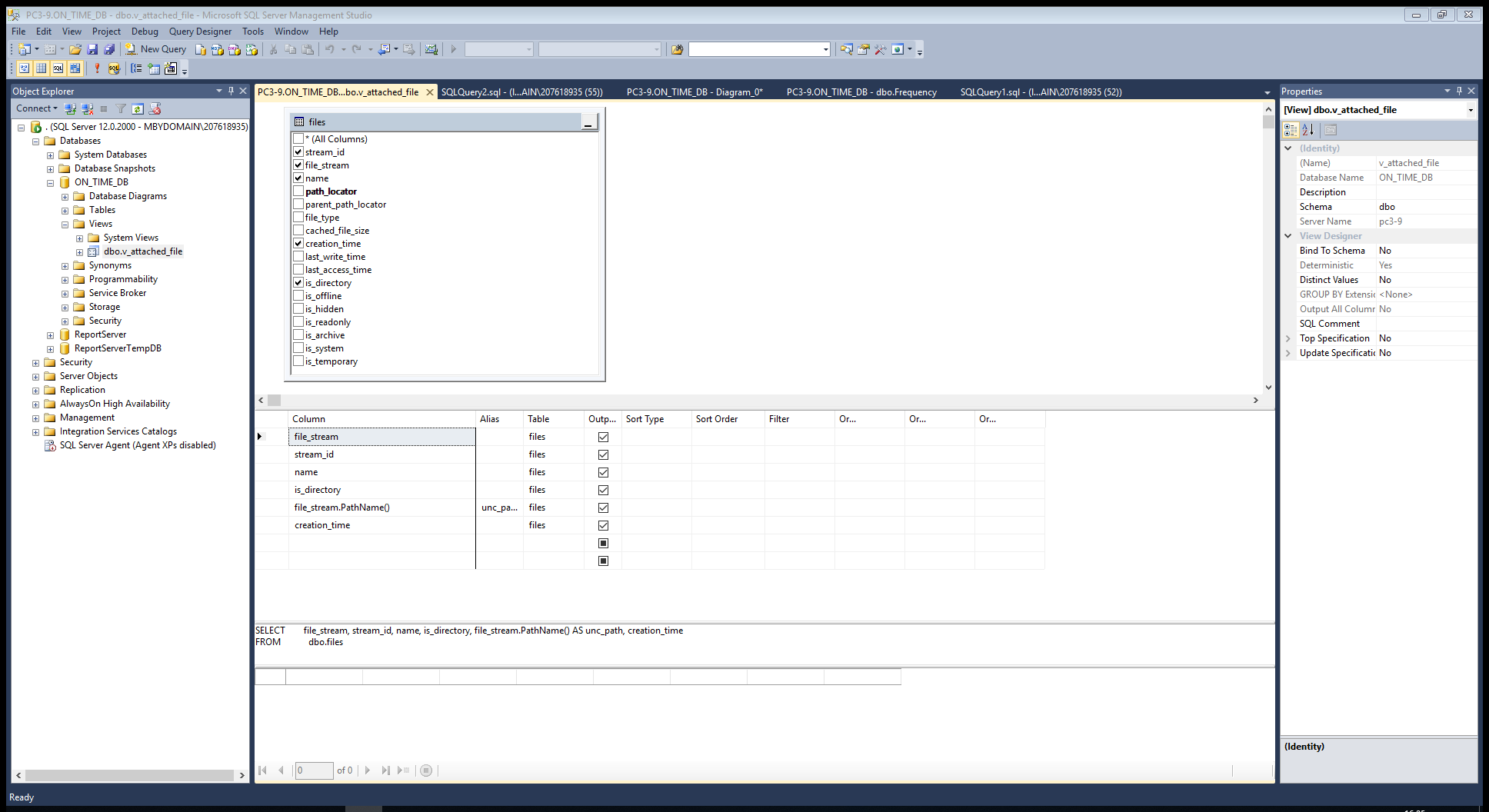
לשמור את הקבצים עצמם בsql בטבלה מיוחדת שנקראת file table

אנו בחרנו בדרך השניה למרות מורכבותה כיון שבה הנתונים נשמרים בבטחה הודות לקשרי גומלין, לעומת קבצים בשרת שפעולה לא אחראית עלולה למחוקם ולאבד את הנתונים החיוניים.

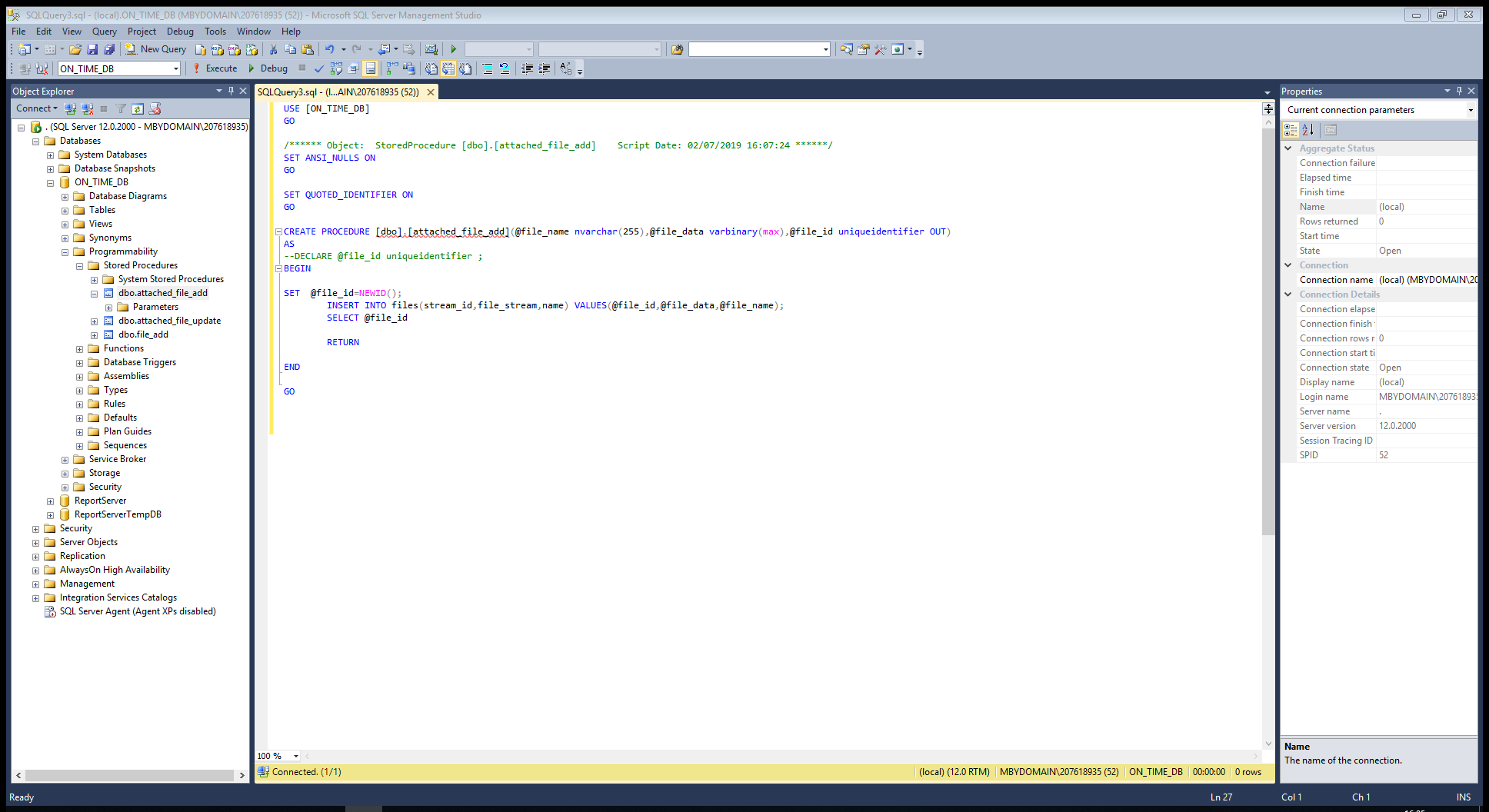
|  |  |  |  |
| --- | --- | --- | --- |
|  | שם | סוג | הסבר |
|  | Stream\_id | uniqueidentifier | מזהה ייחודי עבור נתוני FILESTREAM |
|  | File\_stream | string | מכיל את נתוני FILESTREAM |
|  | name | string | שם הקובץ או הספריה. |
|  | Path\_locator | int | המיקום של הצומת הזה בהיררכיה של- FileNamespace |
|  | Parent\_path\_locator | int | היררכיה של הספרייה המכילה |
|  | File\_type | string | מייצג את סוג הקובץ |
|  | Cached\_file\_size | int | הגודל בבתים של נתוני FILESTREAM. |
|  | Creation\_time | date | התאריך והשעה שבהם נוצר הקובץ. |
|  | Last\_write\_time | date | התאריך והשעה שבהם הקובץ עודכן לאחרונה. |
|  | Last\_access\_time | date | התאריך והשעה שבהם הייתה הגישה האחרונה לקובץ. |
|  | Is\_directory | bit | מציין אם השורה מייצגת ספריה. |
|  | Is\_offline | bit | האם הקובץ לא מקוון. |
|  | Is\_hidden | bit | האם הקובץ מוסתר. |
|  | Is\_readonly | bit | האם הקובץ לקריאה בלבד. |
|  | Is\_archive | bit | האם הקובץ ארכיון. |
|  | Is\_system | bit | האם הקובץ מערכת. |
|  | Is\_temporary | bit | האם הקובץ זמני. |

הערה(כאן או בserver)

על מנת להשתמש בטבלה זו ב entity framework יצרנו view שמאפשר לנו להתנהל מול השדות הרלווטים בטבלה.



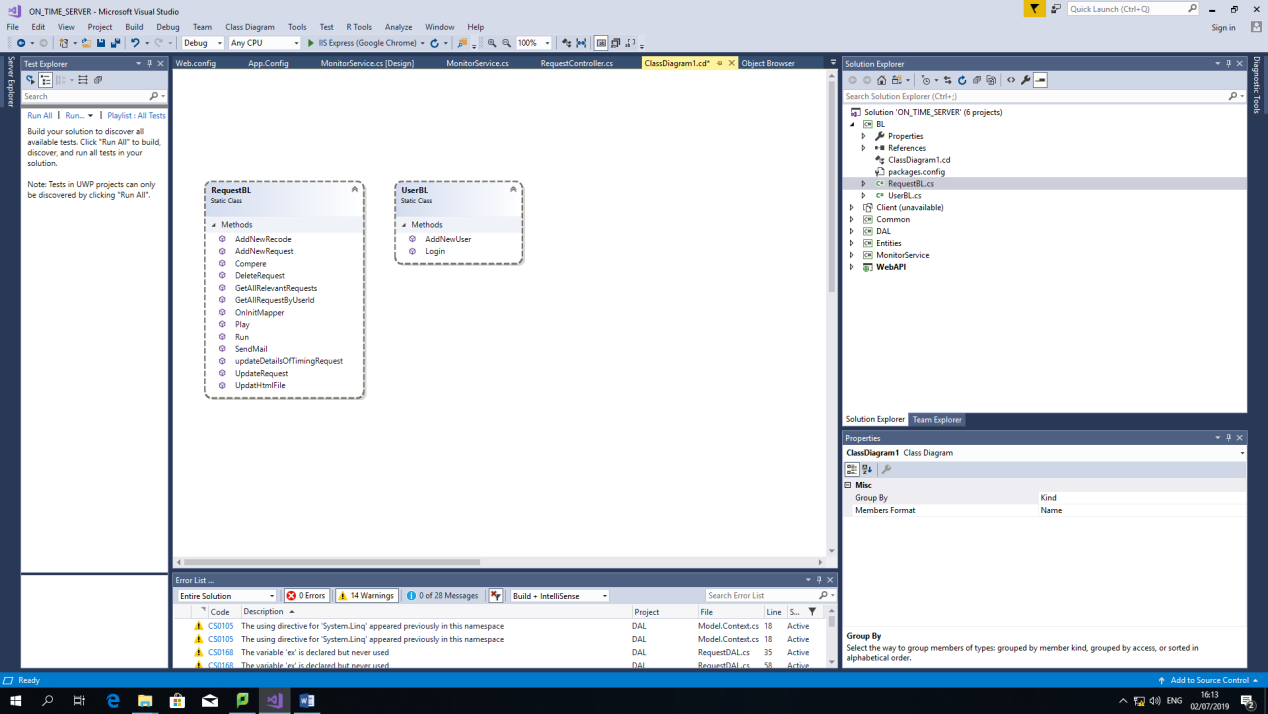
על מנת להוסיף קבצים לטבלה יצרנו פרוצדורה שמקבלת את שם הקובץ ותוכנו. ומחזירה את הקוד.

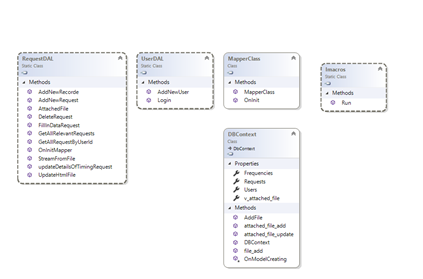


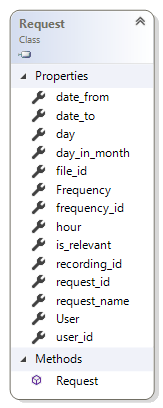
6.3 מחלקות

בנוסף לכל המחלקות הנ"ל, יש מחלקות המתארות את הטבלאות במסד נתונים. (Entity).

(רק לקבל מושגים בעקרון צריך להעתיק כל class בנפרד זה רק לבנתיים)



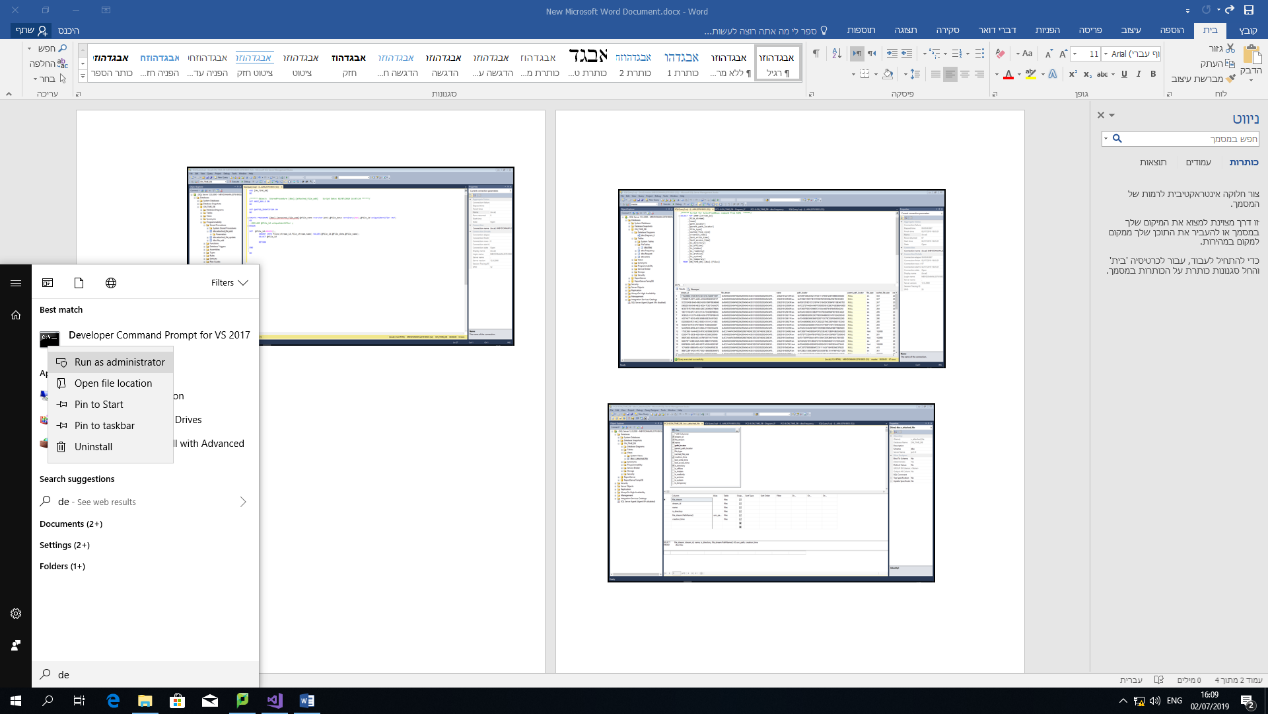


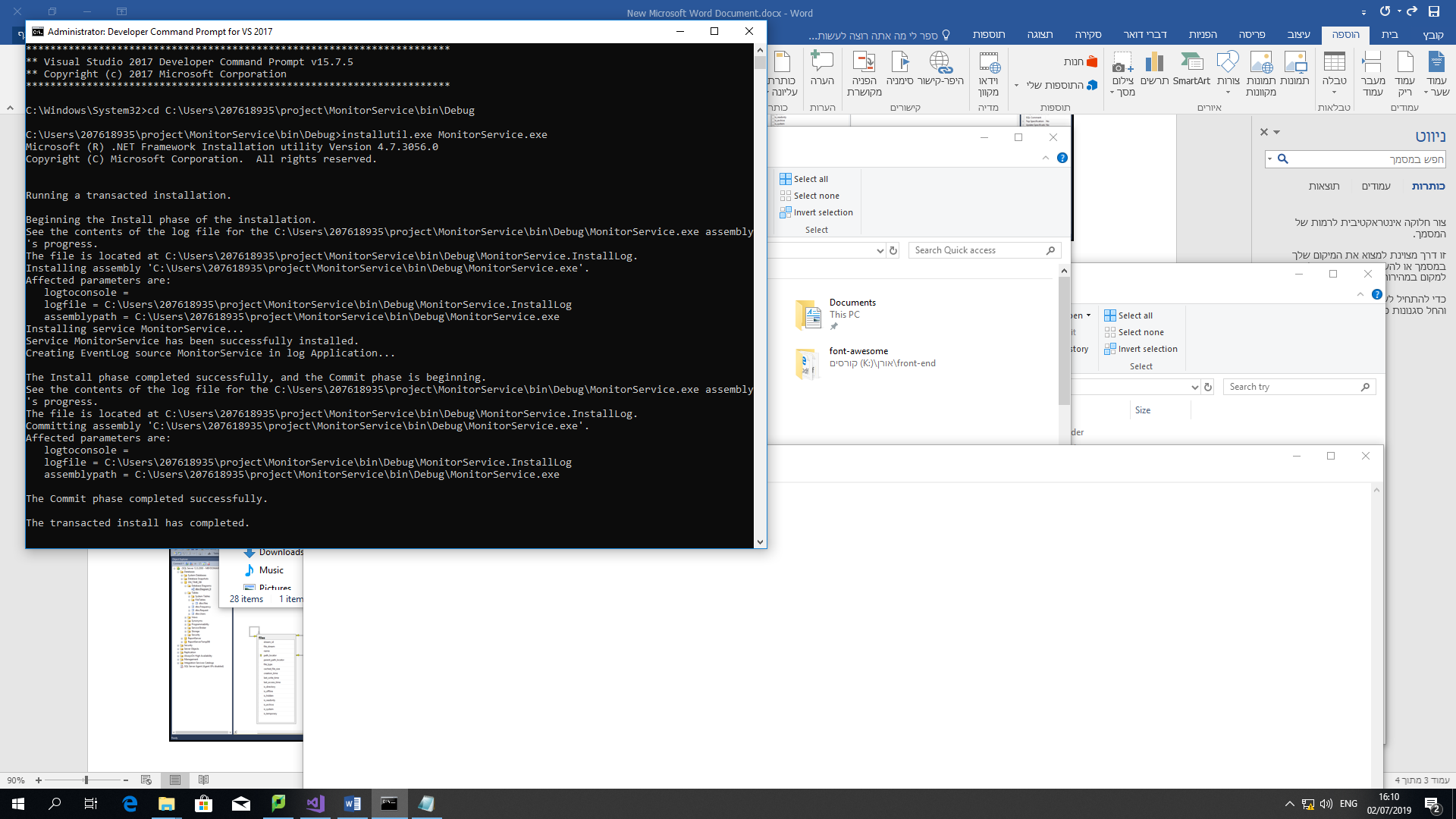
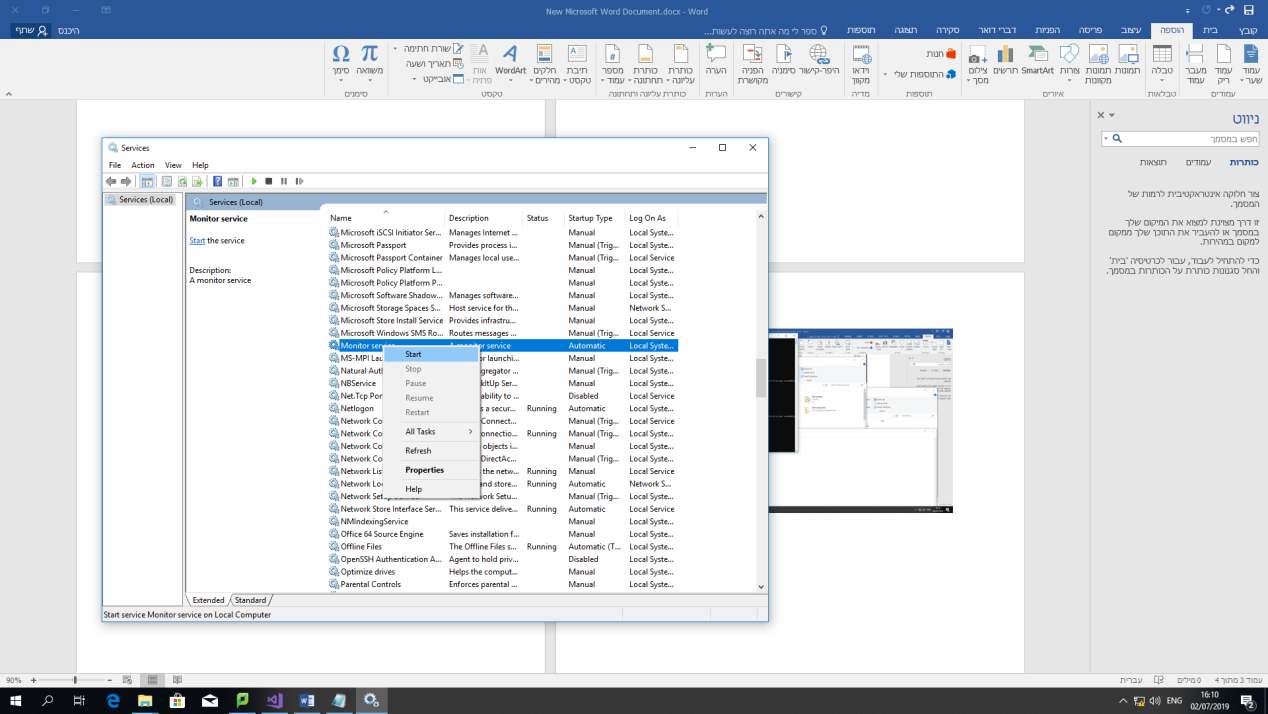
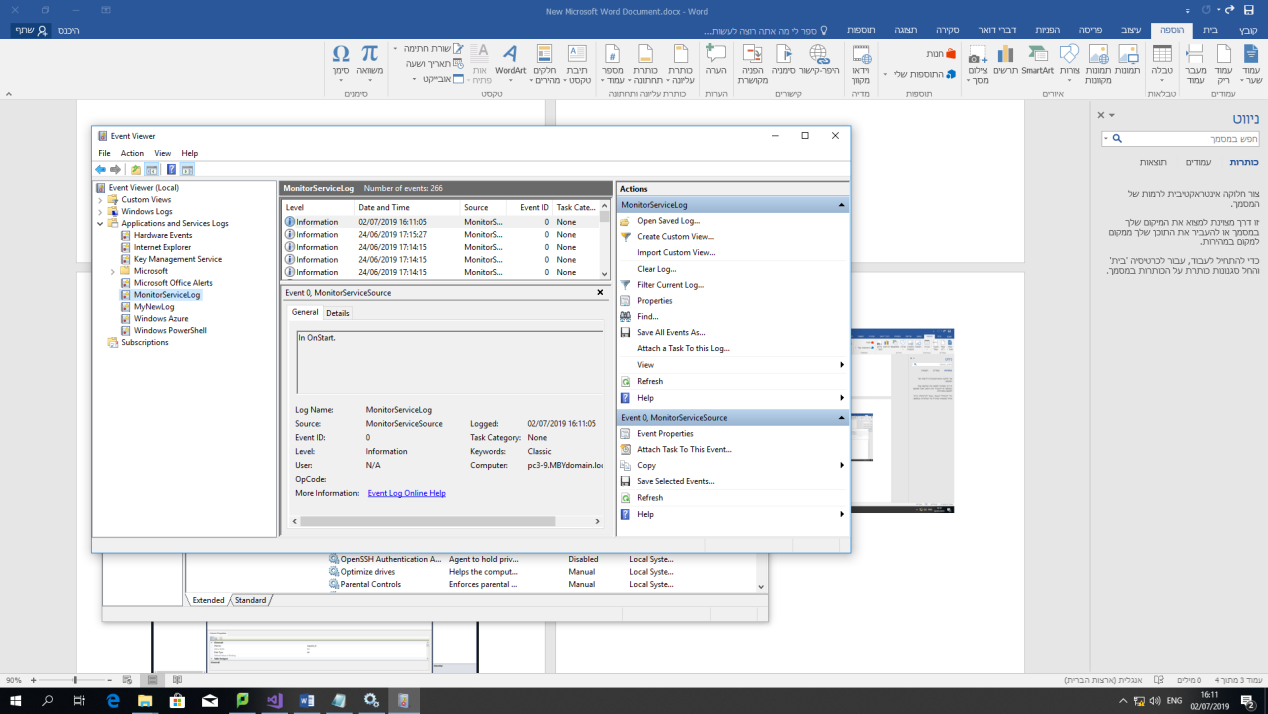


**7 מתזמן**

מעקב אחר שינויים מתבצע בסבב של פעם בחמש דקות, לצורך הפעלת התוכנית בתדירות זו השתמשנו Windows service.

(עדין בלי מלל...רק תמונות מסך)





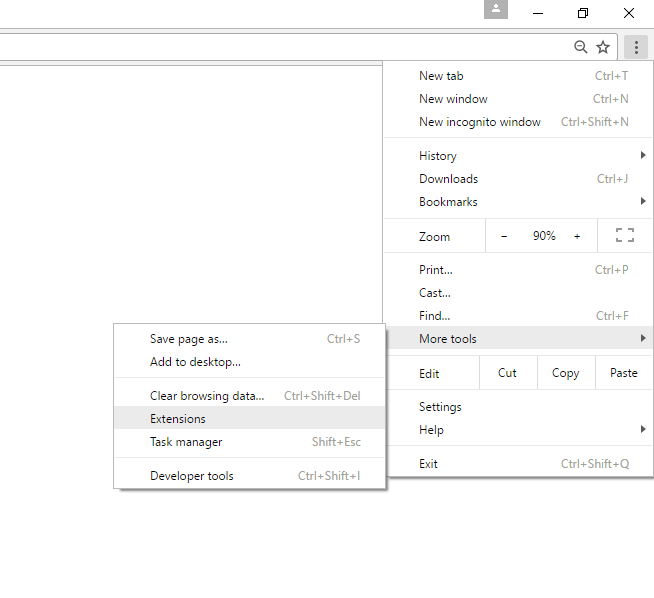
**8 ממשקי המערכת ומדריך למשתמש**

## 8.1 מדריך למשתמש באתר

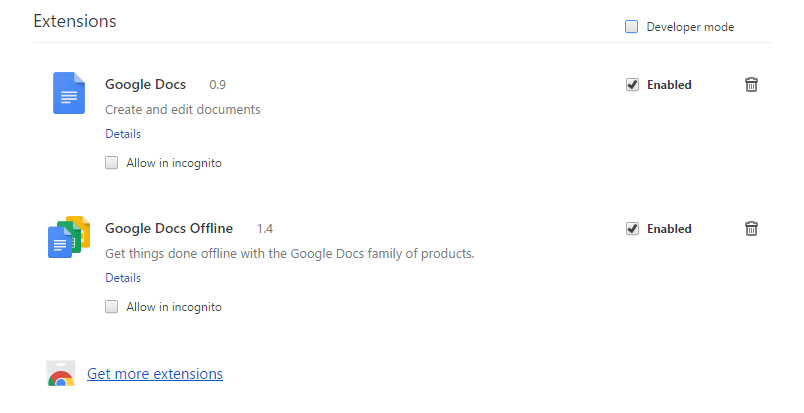
## 8.2 extension ל Google Chrome

### 8.2.1 מדריך התקנת extension ל Google Chrome

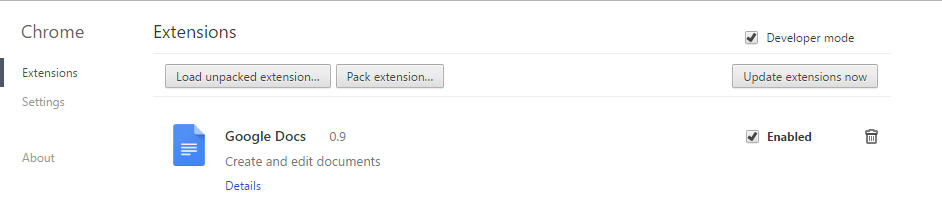
1. הכנס ל **chrome://extensions** בדפדפן שלך או בחר בפינה הימנית של הדפדפן ולבחור **Extensions**  תחת תפריט **Tools** .



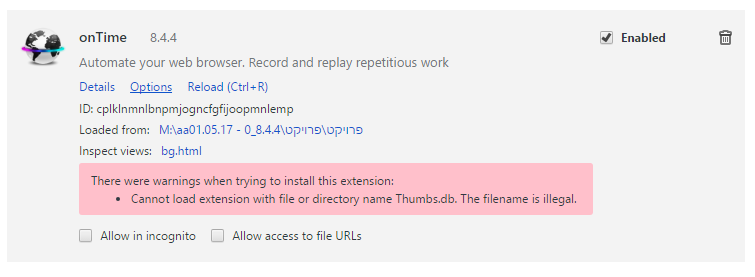
1. וודא כי Developer mode מסומן.



1. לחץ על **Load unpacked extension** ונתב לתיקיה בה התוסף שמור.



בשלב זה התוסף טעון ואמור להתווסף icon לדפדפן, אם אינו מופיע יש לוודא כי התוסף מאופשר.



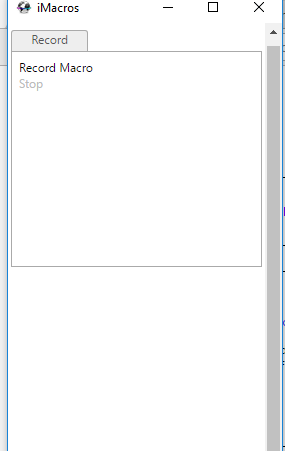
### 8.2.2 מדריך למשתמש בתוסף

הגישה לתוסף היא דרך icon  המופיע בחלק הימני של הדפדפן.

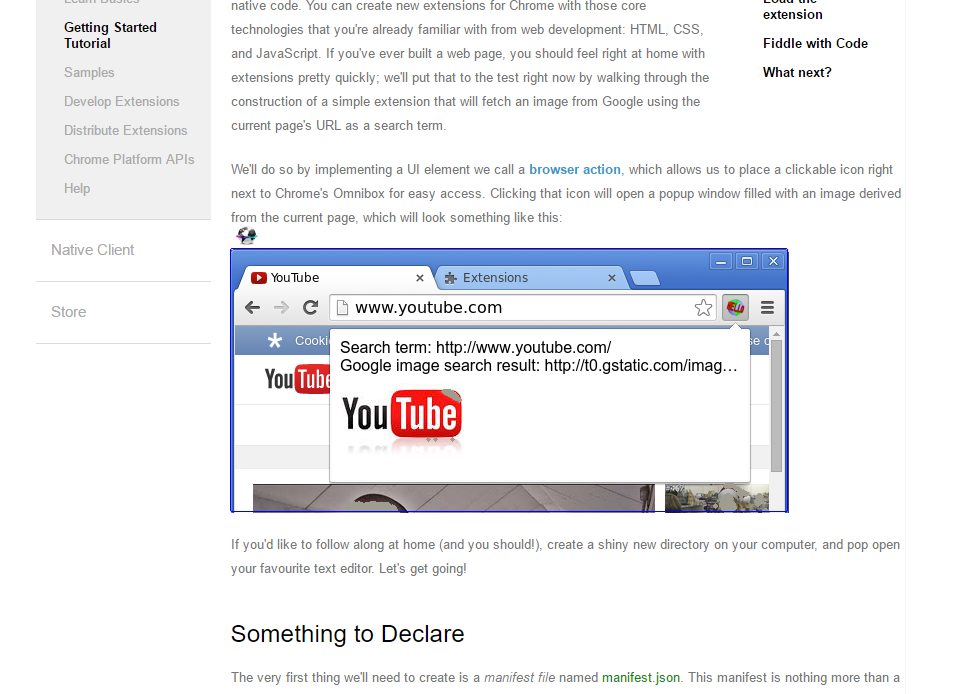
מסך ראשי

תחילת הקלטה

סיום הקלטה



במהלך ההקלטה לאחר כל פעולה שמתבצעת בדפדפן מופיע icon המסמל כי הפעולה נקלטה.

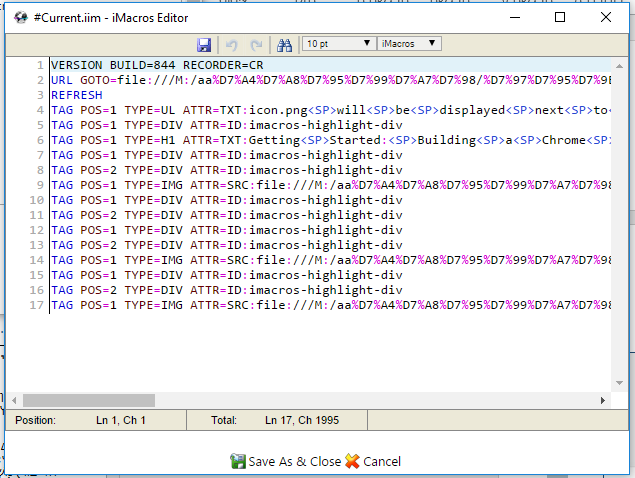


כאשר נמצא הדף המבוקש יש ללחוץ על Stop לסיום ההקלטה.

בסיום ההקלטה מוצגת ההקלטה להתרשמות ועריכה.

שמירה של ההקלטה- ההקלטה נשמרת server.

שמירת ההקלטה



לאחר שההקלטה נשמרה, נפתח האתר וניתן להגדיר את התזמון שלה כך שיוכל להתבצע מעקב בתדירות נבחרת.

**9 מסקנות**

**10 ביבליוגרפיה**

[he.wikipedia.org](https://he.wikipedia.org)

[developer.chrome.com/extension](https://developer.chrome.com/extension)

[stackoverflow.com](http://stackoverflow.com/)

[github.com](https://github.com)

[www.codeproject.com](http://www.codeproject.com)

[internet-israel.com](https://internet-israel.com)

[wiki.imacros.net/iMacros\_for\_Chrome](http://wiki.imacros.net/iMacros_for_Chrome)

[www.seleniumhq.org](http://www.seleniumhq.org)

www.telerik.com

[citeseerx.ist.psu.edu](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.78.8083&rep=rep1&type=pdf)

[www.columbia.edu](http://www.columbia.edu)

Msdn. Microsoft.Com

[www.codechef.com](http://www.codechef.com)

[www.tutorialspoint.com](http://www.tutorialspoint.com)

interactivepython.org

[www.geeksforgeeks.org](http://www.geeksforgeeks.org)

[csharp.net-informations.com](http://csharp.net-informations.com/communications/csharp-smtp-mail.htm)

[www.emailarchitect.net](http://www.emailarchitect.net)

[www.aspsnippets.com](http://www.aspsnippets.com)

[www.c-sharpcorner.com](http://www.c-sharpcorner.com)

en.wikibooks.org

robots.thoughtbot.com

www.ics.uci.edu

code.tutsplus.com

tutorialzine.com

filetable: <https://docs.microsoft.com/en-us/sql/relational-databases/blob/create-alter-and-drop-filetables?view=sql-server-2017>