טבלאות:

טבלת הביטויים

מפתח ראשי: ID

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Phrase | LocationFromStart | RowInText | LocationInRow | TextID | WordNumberInText | isDefinedByUser |
| Int identuty | nvarchar | int | int | int | Int FK | int | bit |

טבלת הטקסטים

מפתח ראשי: ID

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | title | totalWords | totalRows | TotalCharacters |
| Int Identity | nvarchar | int | int | int |

טבלת המידע על הטקסטים

מפתח ראשי: TextID + AttributeKey

|  |  |  |
| --- | --- | --- |
| TextID | AttributeKey | AttributeValue |
| Int FK | nvarchar | nvarchar |

טבלת הביטויים המוגדרים אישית:

מפתח ראשי: ID

|  |  |
| --- | --- |
| ID | Phrase |
| Int identuty | nvarchar |

טבלת קבוצות מוגדרות אישית:

מפתח ראשי: ID

|  |  |
| --- | --- |
| ID | Name |
| Int Identity | nvarchar |

טבלת המילים בקבוצה מוגדרת אישית:

מפתח ראשי: GroupID + Phrase

|  |  |
| --- | --- |
| GroupID | Phrase |
| Int FK | nvarchar |

אינדקסים:

אינדקס על הביטויים בטבלת הביטויים לחיפוש של ביטויים (קלסטר אינדקס)

אינדקס על, TextID ,WordNumberInText לאחזור מהיר של רצפים של טקסט

פרוצדורות חשופות החוצה:

AddNewText\_sp – מקבלת טקסט וכותרת ומוסיפה את הטקסט לטבלת הטקסטים ומאנדקסת אותו לטבלת הביטויים תוך כדי חישוב הסטטיסטיקות שלו

מחזירה מזהה לרשומה שנוספה

AddTextMetaData\_sp - מקבלת מזהה טקסט, key ו value ומכניסה לטבלה המתאימה

GetTextMetaDataAndStats\_sp - מקבלת מזהה טקסט ומחזירה את כל הפרטים שלו (כולל סטטיסטיקות)

FindTextByMetaData\_sp מקבלת key ו value ומחזירה טבלה המכילה את כל מזהי הטקסטים שיש בהם את הנתון הזה

GetIndexForPhrase\_sp – מקבלת ביטוי ואופציונלית מזהה של טקסט ומוצאת ומחזירה את כל המופעים של הביטוי בכלל או בטקסט ספציפי שצוין

GetUserDefinedGroupIndex\_sp – מקבלת מזהה קבוצה ואופציונלית מזהה של טקסט ומוצאת ומחזירה את כל המופעים של ביטויי הקבוצה בכלל או בטקסט ספציפי שצוין

GetContextForPhrase\_sp מקבל מזהה ביטוי וטווח של טקסט מביא את הטקסט שמקיף את הביטוי הספציפי עבור התצוגה לפי הטווח שצוין

GetContextForPhraseAllAppearance\_sp מביא את הטקסט שמקיף את כל מופעי הביטוי עבור התצוגה

GetTextByID\_sp מקבלתID של טקסט ומחזירה את הטקסט המתאים

AddUserDefinedPhrase\_sp מקבלת ביטוי חדש ואם הוא לא קיים עדיין אז מוצאת את כל המופעים שלו בטקסטים הקיימים ומכניסה את הרפרנסים לטבלת הביטויים ואת הביטוי עצמו לטבלת הביטויים המוגדרים אישית

DeleteUserDefinedPhrase\_sp מקבלת ID של ביטוי מוגדר ומוחקת אותו ואת המופעים שלו מהDB

GetPhraseByID\_sp מקבל ID של ביטוי ומביא את רשומת הביטוי (מופע ספציפי)

DeleteText\_sp מקבלת מזהה של טקסט ומוחקת אותו ואת כל הציטוטים שלו מהDB

CreateUserDefinedWordsGroup\_sp מקבלת שם קבוצה ומגדירה אותה, מחזירה מזהה קבוצה

AddPhraseToUserDefinedGroup\_sp מקבל ביטוי ומזהה של קבוצה ומוסיף אותו לקבוצה

AddPhraseListToUserDefinedGroup\_sp מקבל מחרוזת של ביטויים ואת התו המפריד בניהם ומזהה של קבוצה ומוסיף אותם לקבוצה

DeleteUserDefinedWordFromGroup\_sp מקבל ביטוי ומזהה של קבוצה ומוחק אותו מהקבוצה

DeleteUserDefinedWordsGroup\_sp מקבל מזהה של קבוצה ומוחק אותה

GetPhraseByRowLocation\_sp מקבל מספר שורה ,מספר מילה בשורה ומזהה טקסט ומחזיר את המילה המבוקשת , יחזיר NULL אם מפנה למיקום שלא קיים בטקסט

GetAllPhrasesInAText\_sp מקבל מזהה של טקסט ומחזיר את כל הביטויים שבו (כולל המוגדרים אישית)

GetStatsForPhrase\_sp מקבל מילה ומחזיר את מספר המופעים הכולל שלה ובכמה טקסטים שונים היא מופיעה

GetStatsForRow\_sp מקבל מזהה טקסט ומספר שורה ומחזיר את מספר המילים והתווים בשורה (התווים כולל רווחים)

GetFrequencyList\_sp מחזיר את כל המילים כשעבור כל אחת מהן מצוין מספר המופעים שלה בDB

GetFrequencyListForText\_sp מחזיר את כל המילים בטקסט ספציפי כשעבור כל אחת מהן מצוין מספר המופעים שלה בDB

SearchPhrase\_sp מקבל ביטוי ומחזיר את כל המופעים שלו שנמצאו ניתן להגביל את החיפוש לטקסט ספציפי ע"י העברה של פרמטר אופציונאלי

ExportIngestedTexts\_sp מייצא לXML את הטקסטים השמורים בDB

ExportUserDefinedGroups\_sp מייצא לXML את קבוצות המילים השמורות של המשתמשים

ExportUserDefinedPhrases\_sp מייצא לXML את הביטויים מוגדרים אישית השמורים בDB

GetImportPathesForXML\_sp מחזירה את הנתיבים המוגדרים לטעינת נתונים מתוך קבצי הXML לDB

ImportIngestedTexts\_sp מייבא טקסטים לDB מתוך XML

ImportIngestedTextsFromPath\_sp מייבא טקסטים לDB מתוך קובץ XML שנשמר בנתיב (אופציה עבור קבצים גדולים במיוחד)

ImportUserDefinedGroups\_sp מייבא קבוצות מילים שמורות של משתמשים לDB מתוך XML

ImportUserDefinedGroupsFromPath\_sp מייבא קבוצות מילים שמורות של משתמשים לDB מתוך קובץ XML שנשמר בנתיב (אופציה עבור קבצים גדולים במיוחד)

ImportUserDefinedPhrases\_sp מייבא ביטויים מוגדרים אישית לDB מתוך XML

ImportUserDefinedPhrasesFromPath\_sp מייבא ביטויים מוגדרים אישית לDB מתוך קובץ XML שנשמר בנתיב (אופציה עבור קבצים גדולים במיוחד)

IngestNewTextMultyLevelProcess\_StepOne\_CreateTextEntity\_sp שלב ראשון בתהליך טעינת טקסט שורה אחר שורה, פרוצדורה זו נקראת פעם אחת כדי ליצור את מסגרת הטקסט ולהקצות לו ID במערכת

IngestNewTextMultyLevelProcess\_StepTwo\_AddNewRow\_sp שלב שני בתהליך טעינת טקסט שורה אחר שורה, פרוצדורה זו נקראת עבור כל שורה בטקסט ומכניסה אותה לDB

IngestNewTextMultyLevelProcess\_StepThree\_ApplyUDP\_sp – שלב אחרון בתהליך טעינת טקסט שורה אחר שורה, פרוצדורה זו נקראת פעם אחת כדי להחיל את הטקסט החדש את המיפוי לביטויים המוגדרים אישית של המשתמשים

GetUserDefinedGroupInfo\_sp מקבל מזהה קבוצה ומחזיר את כל הפרטים שלה ואת כל המילים שבה

GetUserDefinedGroupsList\_sp מחזיר רשימה של כל הקבוצות

פונקציות פנימיות :

[dbo].[GetNextPhraseInText\_f] מביאה את הביטוי הבא בטקסט

[dbo].[GetTextByIndexs\_f] מחזירה בין שני אינדקסים (למשל את הטקסט בין המילה השלישית למילה הארבעים)

GetTextByIndexsNonRecurcive\_f מביאה את הביטוי הבא בטקסט ללא שימוש ברקורסיה (כך שנמנעת המגבלה על מספר תת הקריאות שקיימת בDB )

[dbo].[GetWordCount\_f] מקבלת מילה ומחזריה את מספר התווים שבה

