מסמך דרישות ל Article Crawler

טרמינולוגיה:

מאמר:

חבילת מידע שנוצרה על ידי מחבר, קטע עיתונות המכיל גוף וכותרת. נמצא באחד אתרי המידע.

מאמר מפורסר:

מאמר שהורד מהאינטרנט וסודר בצורה אחידה לפי הנקבע ע"י המוריד.

מחבר:

האיש אשר חיבר את המאמר.

אתר מפרסם:

אתר באינטרנט שמפרסם מאמרים בתבנית אחידה לכלל המאמרים שהוא מפרסם.  
כולל לוגו, כתובת ראשית, וחוקים

חוקים:

מכילים מידע של:  
PREFIX וSUFFIX לכתובת  
מיקום המידע