כשרוצים את ה-N מסמכים הכי טובים, במקום לשמור אותם ברשימה, למיין את הרשימה ולהחזיר את ה-N איברים הראשונים, יעיל יותר לתחזק ערימת מקסימום ולבצע N הוצאות מהערימהלהחזיר רק מסמכים שמכילים לפחות אחת (או יותר) מהמילים בשאילתה, אולי אפילו רק מסמכים שמכילים את כל המילים בשאילתה (אבל אז אתם יכולים להסתכן בלהחזיר פחות מ-N מסמכים)להגדיר איזשהו רף מינימום של ציון idf ולהחזיר רק מסמכים שמכילים term עם ציון שגדול מהרף הזה. אפשר להתייחס למילים עם idf נמוך כ-stopwordsליצור לכל term משהו שנקרא champion list - רשימה של r המסמכים עם ה-tf-idf הכי טובים עבור ה-term הזה (כאשר r>N) ולהחזיר רק מסמכים מתוך ה-champion list הזה (נשמע לי לא ככ מציאותי לעשות preprocessing כזה מוגזם, מצד שני הם אמרו שהם יתנו לקוד שלנו לרוץ במשך יומיים אז אולי כן)