

מערכת בינה מלאכותית לסיווג וחלוקה לצברים של שיחות ברשתות חברתיות

מבצעים: רותם לוי, סינדי בובליל

מנחים: ד"ר אמנון רוזנמן, מר אגמון פורת (Itweetlive)

אודות Itweetlive:

חברת סטארט אפ שפיתחה מוצר המאפשר

- מעקב בזמן אמת אודות ציוצים שעלו ברשת

- יכולת הפצת הודעות המוניות

הגדרת הבעיה:

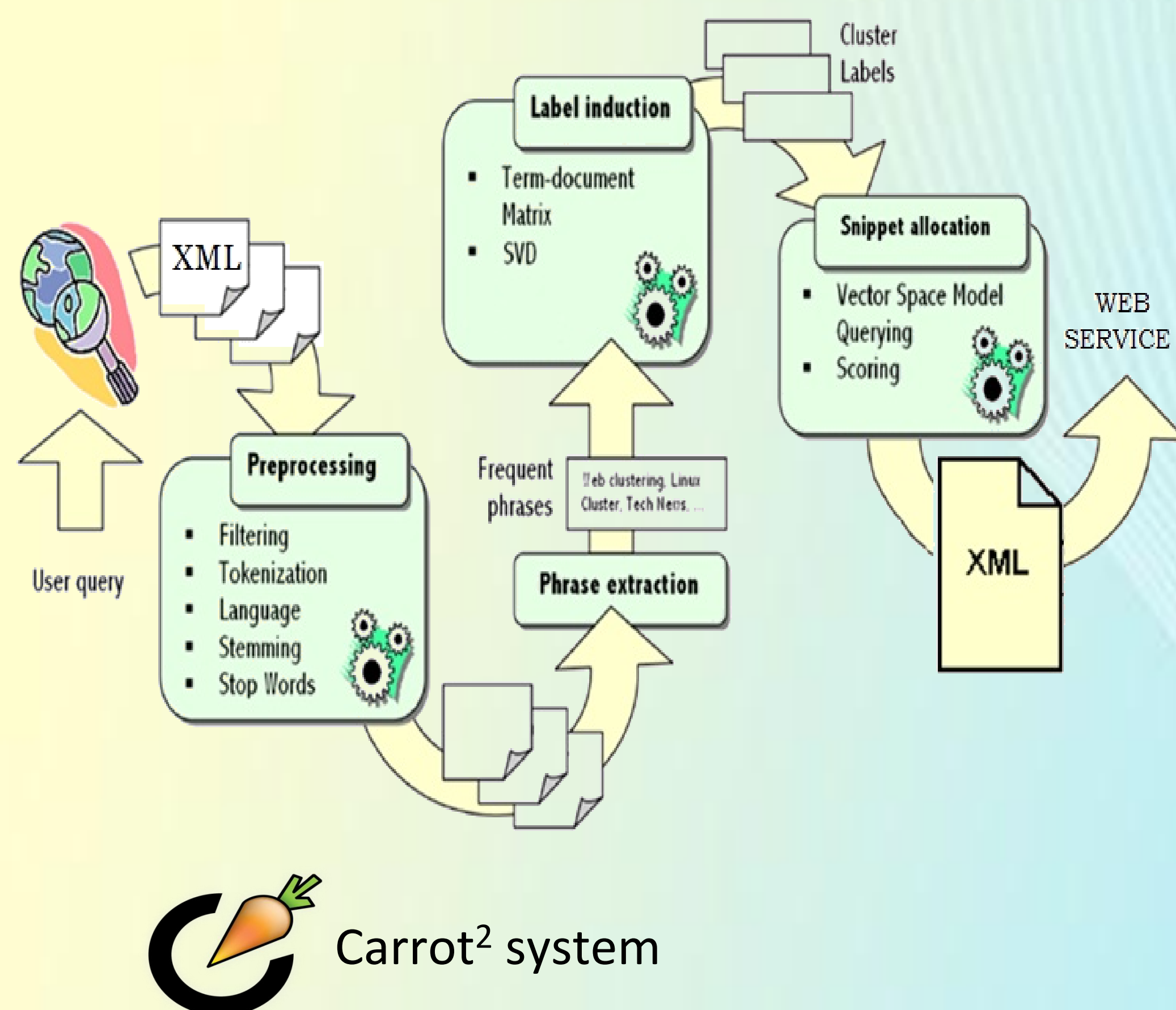
שליחת הודעות המוניות גורמת להצפת רשת twitter בדואר זבל (spam), ושלא במתכוון כופה על אנשים הודעות שאין להם עניין בהם.

מטרת הפרויקט:

לנסות לצמצם את ה spamming כך שהודעות שיישלחו למשתמשים יהיו רלוונטיות להם

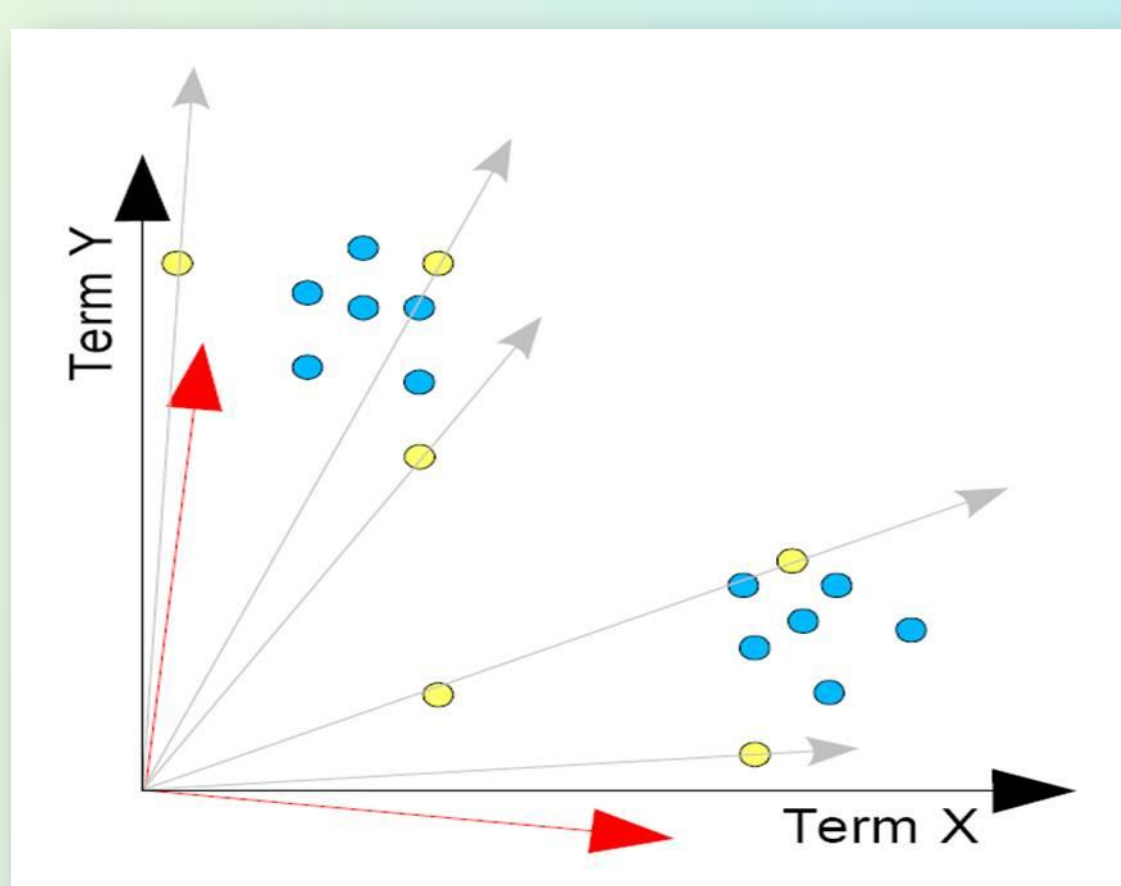
השיטה לפיתרון:

לפתח מערכת המחלקת לצברים (קבוצות) את הציוצים שהתקבלו, כדי לאפשר שליחת הודעות לקבוצות בעלות מכנה משותף גדול ככל הניתן, תוך מימוש אלגוריתם לחלוקה לצברים בסביבת c# המבוסס על מערכת carrot2 (קוד פתוח).



שלבים באלגוריתם לחלוקה לצברים:

- עיבוד מקדים לטקסט
- פרוק לביטויים
- מציאת מטריצת הדמיון בין הציוצים
- הורדת מימדיה של מטריצת הדמיון (RSVD)
- מציאת כותרות לצברים ושיוך הציוצים לצברים
- מתן ציון על איכות הצברים



מסקנות:

אלגוריתם החלוקה לצברים שמימשנו נותן פיתרון לדרישות שהוצבו:

- מתאים לטקסטים קצרים
- מעניק כותרות משמעותיות לצברים
- מחלק לקבוצות הומוגניות בזמן ריצה סביר



חלוקה לצברים