Bài báo đưa ra cách rút trích các đặc trưng theo các topic, mà ta có thể dùng các đặc trưng này để phân lớp tài liệu theo topic .

+ Bài đưa ra cách rút trích các đặc trưng cho một tài liệu, mà những đặc trưng này có thể dùng các thuật toán như NB và SVM để phân lớp .

Điều kiện để phân lớp topic là :

* Topic đó không được dùng nhiều trên internet như (computer,… có thể kiểm tra điều kiện này bằng cách search các từ khóa trên search engine ), .
* Topic không được nhiều nghĩa trong từ điển.
* Tập topic cần phải rõ nghĩa với nhau, không được bao hàm nhau.

Khi rút trích đặc trưng dựa vào tập dữ liệu LDC dataset (Linguisitic Data consortium ), IF và IDF để tính và đánh trọng số