DA là hệ thống quản lý văn bản.  
- web

- yêu cầu: + đọc các văn bản mẫu do admin/ manager upload convert sang text.

+ Các định dạng:

* Doc, Docx
* Xls, xlsx
* Pdf ( pdf loại có thể cóp được chữ thì convert sang text, loại ảnh chụp thì không xử lý sang text)

                + chuyển đổi sang dạng text để có khả năng search keyword .

+ keywords bao gồm:

* Tên văn bản
* Ngày ban hành
* Ngày ký
* Người ban hành
* Người ký
* Từ ngày – đến ngày
* Số hiệu văn bản
* Nội dung bên trong text của văn bản.

+ Sử dụng thuật toán N-Gram để phân tích văn bản text ra keyword hộ trợ người dung không phải nhâp keywords (viết thành dạng lib càng tốt. đầu vào là text đâu ra là 1 list<String> keywords) .

+ lưu ý cần có 1 bước tiền xử lý văn bản text trước khi đưa vào thuật toán N-Gram. (<https://spark.apache.org/docs/latest/ml-features.html>).

+ Keywords có thể do người dung edit . thêm vào hoặc bớt đi .

+ Đối với pdf file ảnh thì người dung tự tạo keywords.

+ cho phép người dung search theo keywords hoặc nội dung trong text.

+ Giao diện người dung gần giống như 3 ảnh ở ngoài (optional có thể khác nếu hay hơn).

                + cung cấp link down văn bản gốc do admin upload cho user.

+ Có 2 cấp quản lý người dung là admin và trưởng phòng . admin có thể tạo phòng và trưởng phòng . trưởng phòng tạo nhân viên cho phòng(users) . users chỉ search được các văn bản do trưởng phòng upload. ( có thể nhận được mail thông báo có văn bản mới hoặc thay thế nếu có … opitional)

+Role:

-User:

-chỉ được search văn bản do manager add mình vào upload

+download văn bản search được .

-Manager:

+thêm user vào room/department mình quản lý

+Manager users.

+upload văn bản

+Manager văn bản theo catergory (có thể tự tạo catergory để quản lý văn bản)

-Admin:

Tạo manager

Tạo room/department.