

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

# ĐỒ ÁN TỐT NGHIỆP

Hệ thống gợi ý tin tức

PHAN THANH LONG

long.pt183587@sis.hust.edu.vn

Ngành Khoa học máy tính

Giảng viên hướng dẫn: ThS. Ngô Văn Linh

Chữ kí GVHD

Khoa: Khoa học máy tính

Trường: Công nghệ thông tin và Truyền thông

HÀ NỘI, 03/2023

# LỜI CẢM ƠN

Lời đầu tiên em xin cảm ơn Thạc sĩ Ngô Văn Linh, người đã giúp đỡ, động viên em nhiều trong suốt quá trình thực hiện đồ án tốt nghiệp. Thầy là người truyền cảm hứng cho em về mảng hệ gợi ý nói riêng và trí tuệ nhân tạo nói chung, giúp em có định hướng phát triển bản thân trong tương lai. Em xin cảm ơn các thầy cô từ Đại học Bách khoa Hà Nội, những năm qua đã nhiệt tình giảng dạy, giúp em dần hoàn thiện các kiến thức, kỹ năng. Con xin cảm ơn bố mẹ, người thân đã luôn ở bên con, là chỗ dựa kể cả về vật chất lẫn tinh thần. Cảm ơn anh em, bạn bè đã giúp đỡ, hỗ trợ tôi trong quá trình làm đồ án. Cuối cùng, xin cảm ơn Công ty VCCorp, cá nhân anh Nguyễn Văn Cường (trưởng bộ phận Platform - Bigdata), cùng các đồng nghiệp đã hỗ trợ tài nguyên tính toán cùng các kiến thức thực tế, để tôi có thể triển khai thành công đồ án tốt nghiệp.

# TÓM TẮT NỘI DUNG ĐỒ ÁN

Báo điện tử hiện nay đang dần phát triển và có lượng truy nhập khổng lồ, đặt ra vấn đề về việc lưu trữ và tính toán trên một lượng lớn dữ liệu. Bên cạnh đó, các trang báo cần xây dựng hệ gợi ý tin tức, đề xuất các chủ đề được quan tâm để giữ chân người dùng trên nền tảng của họ.

Để giải quyết các vấn đề này, đồ án đề xuất xây dựng một hệ thống lưu trữ, tính toán dữ liệu lớn sử dụng Hadoop, Spark, hệ gợi ý sử dụng học sâu. Ngoài ra, phần backend được xây dựng theo kiến trúc Microservice, triển khai trên môi trường Cloud, có khả năng chịu tải cao. Đồ án hướng tới xây dựng một hệ thống với các chức năng cơ bản của một trang báo điện tử, đi kèm với hệ thống lưu trữ dữ liệu người dùng cùng hệ gợi ý tin tức.

## MỤC LỤC

<b>CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....</b>	<b>1</b>
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và phạm vi đề tài.....	1
1.3 Định hướng giải pháp.....	2
1.4 Bố cục đồ án .....	3
<b>CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU.....</b>	<b>4</b>
2.1 Khảo sát hiện trạng .....	4
2.2 Tổng quan chức năng .....	4
2.2.1 Biểu đồ use case tổng quát .....	4
2.2.2 Biểu đồ use case phân rã.....	5
2.2.3 Quy trình nghiệp vụ .....	7
2.3 Đặc tả chức năng .....	7
2.3.1 Đặc tả usecase Login.....	8
2.3.2 Đặc tả usecase xây dựng mô hình gợi ý .....	9
2.4 Yêu cầu phi chức năng .....	10
2.4.1 Tính dễ dùng .....	10
2.4.2 Tính dễ mở rộng .....	10
2.4.3 Tính chịu lỗi cao .....	10
<b>CHƯƠNG 3. CÔNG NGHỆ SỬ DỤNG.....</b>	<b>11</b>
3.1 Microservice .....	11
3.2 Hadoop Ecosystem .....	13
3.3 Apache Spark .....	13
3.4 Apache Kafka .....	13
3.5 Azure Kubernetes Service .....	15

3.6 Gitlab CI .....	15
3.7 ReactJs .....	16
3.8 Hệ thống gợi ý.....	17
<b>CHƯƠNG 4. THIẾT KẾ VÀ XÂY DỰNG HỆ THỐNG .....</b>	<b>22</b>
4.1 Thiết kế kiến trúc.....	22
4.1.1 Lựa chọn kiến trúc phần mềm .....	22
4.1.2 Thiết kế tổng quan cho backend service .....	23
4.2 Thiết kế chi tiết.....	24
4.2.1 Thiết kế giao diện .....	24
4.2.2 Thiết kế chi tiết lớp .....	25
4.2.3 Thiết kế dữ liệu.....	30
4.3 Xây dựng ứng dụng.....	33
4.3.1 Thư viện và công cụ sử dụng .....	33
4.3.2 Hệ thống Hadoop Spark.....	34
4.3.3 Thiết kế Api .....	35
<b>CHƯƠNG 5. KẾT QUẢ THỰC NGHIỆM.....</b>	<b>36</b>
5.1 Kết quả triển khai hệ thống.....	36
5.1.1 Triển khai hệ thống lưu trữ tính toán phân tán trên cụm vật lí .....	36
5.1.2 Triển khai Backend, Frontend, Kafka, Mysql trên Cloud.....	39
5.2 Minh họa các chức năng chính .....	43
5.3 Kiểm thử.....	46
5.3.1 Kiểm thử quy trình CI/CD .....	46
5.3.2 Kiểm thử chức năng tự động scale up khi có tải tăng .....	47
5.4 Đánh giá thực nghiệm .....	49
5.4.1 Đánh giá trên tập dữ liệu MIND .....	49
5.4.2 Đánh giá trên tập dữ liệu từ trang báo Kênh 14 .....	51

<b>CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN .....</b>	<b>53</b>
6.1 Kết luận.....	53
6.2 Hướng phát triển.....	53
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>54</b>
<b>PHỤ LỤC.....</b>	<b>56</b>
<b>A. ĐẶC TẢ USE CASE, MÔ TẢ NGHIỆP VỤ.....</b>	<b>56</b>
A.1 Đặc tả usecase .....	56
A.1.1 Đặc tả usecase Đăng nhập.....	56
A.1.2 Đặc tả usecase Đăng ký .....	57
A.1.3 Đặc tả usecase Tìm kiếm bài báo .....	58
A.1.4 Đặc tả usecase Đọc bài báo .....	59
A.1.5 Đặc tả usecase Tạo bài báo.....	59
A.1.6 Đặc tả usecase Chỉnh sửa bài báo.....	60
A.1.7 Đặc tả usecase Xóa bài viết .....	61
A.1.8 Đặc tả usecase Đăng xuất.....	62
A.1.9 Đặc tả usecase Tạo chủ đề .....	62
A.1.10 Đặc tả usecase Lấy danh sách bài viết theo chủ đề .....	63
A.1.11 Đặc tả usecase Lấy danh sách bài viết được đề xuất .....	64
A.1.12 Đặc tả usecase xây dựng mô hình gợi ý.....	64
A.1.13 Đặc tả usecase Thống kê trực quan hóa dữ liệu.....	65
A.2 Quy trình nghiệp vụ .....	66
A.2.1 Quy trình nghiệp vụ đăng nhập.....	66
A.2.2 Quy trình nghiệp vụ đăng ký .....	67
A.2.3 Quy trình nghiệp vụ tìm kiếm bài báo .....	67
A.2.4 Quy trình nghiệp vụ tạo bài báo.....	68
A.2.5 Quy trình nghiệp vụ chỉnh sửa bài báo .....	69

A.2.6 Quy trình nghiệp vụ xây dựng mô hình gợi ý .....	70
<b>B. NGHIÊN CỨU LIÊN QUAN.....</b>	<b>71</b>
B.1 Tìm hiểu chi tiết về Kubernetes .....	71
B.1.1 Khái niệm.....	71
B.1.2 Kiến trúc .....	72
B.1.3 Các Controller.....	74
B.2 Tìm hiểu chi tiết về Hadoop - Spark.....	75
B.2.1 HDFS.....	75
B.2.2 YARN .....	77
B.2.3 Spark.....	79

## DANH MỤC HÌNH VẼ

Hình 2.1	Biểu đồ usecase tổng quan . . . . .	5
Hình 2.2	Usecase phân rã news management . . . . .	6
Hình 2.3	Usecase phân rã data analysis . . . . .	6
Hình 2.4	Quy trình nghiệp vụ triển khai mô hình gợi ý . . . . .	7
Hình 3.1	Mô hình Microservice . . . . .	11
Hình 3.2	Cấu trúc lục giác . . . . .	12
Hình 3.3	Kiến trúc tổng quan của Kafka . . . . .	14
Hình 3.4	Minh họa kiến trúc kubernetes . . . . .	15
Hình 3.5	Minh họa luồng CI/CD dùng cho gitlab ci . . . . .	16
Hình 3.6	Minh họa mô hình gợi ý dựa trên nội dung . . . . .	18
Hình 3.7	Mô hình gợi ý tin NAML . . . . .	19
Hình 3.8	Ma trận nhầm lẫn . . . . .	20
Hình 3.9	Đường cong ROC . . . . .	20
Hình 4.1	Mô hình tổng quan hệ thống . . . . .	22
Hình 4.2	Thiết kế backend theo kiến trúc microservice . . . . .	24
Hình 4.3	Thiết kế giao diện trang chủ . . . . .	25
Hình 4.4	Thiết kế giao diện trang chi tiết bài báo . . . . .	25
Hình 4.5	Thiết kế service tin tức . . . . .	26
Hình 4.6	Biểu đồ trình tự nghiệp vụ đọc tin tức . . . . .	27
Hình 4.7	Thiết kế service Admin . . . . .	29
Hình 4.8	Thiết kế service Auth . . . . .	30
Hình 4.9	Thiết kế cơ sở dữ liệu . . . . .	31
Hình 4.10	Cài đặt cụm Hadoop - Spark trên YARN . . . . .	34
Hình 5.1	Hệ thống tệp tin phân tán HDFS . . . . .	37
Hình 5.2	Hệ thống xử lý phân tán YARN . . . . .	37
Hình 5.3	Job streaming được submit trên yarn . . . . .	38
Hình 5.4	Job spark streaming . . . . .	38
Hình 5.5	Kết quả ghi log người dùng . . . . .	39
Hình 5.6	Kết quả ghi log người dùng . . . . .	39
Hình 5.7	Các thành phần liên quan AKS . . . . .	41
Hình 5.8	Kết quả deploy các dịch vụ backend . . . . .	41
Hình 5.9	Kết quả deploy API gateway . . . . .	42
Hình 5.10	Kết quả deploy Kafka . . . . .	42



Hình 5.11	Kết quả deploy các Frond End . . . . .	43
Hình 5.12	Minh họa trang chủ . . . . .	43
Hình 5.13	Minh họa trang báo . . . . .	44
Hình 5.14	Đăng nhập Jupyter Notebook . . . . .	44
Hình 5.15	Thống kê lượng user hoạt động theo giờ . . . . .	45
Hình 5.16	Thống kê lượt xem theo giờ . . . . .	45
Hình 5.17	CTR(Click Through Rate) . . . . .	46
Hình 5.18	Gitlab tự động deploy khi có commit mới . . . . .	46
Hình 5.19	K8S khởi tạo lại pod mới khi người dùng update image . . . . .	47
Hình 5.20	Kết quả benchmark sau khi áp dụng cơ chế HPA . . . . .	48
Hình 5.21	Kết quả Autoscale . . . . .	49
Hình 5.22	Minh họa kết quả gợi ý . . . . .	52
Hình A.1	Quy trình nghiệp vụ chức năng đăng nhập . . . . .	66
Hình A.2	Quy trình nghiệp vụ chức năng đăng ký . . . . .	67
Hình A.3	Quy trình nghiệp vụ chức năng tìm kiếm bài báo . . . . .	67
Hình A.4	Quy trình nghiệp vụ chức năng tạo bài báo . . . . .	68
Hình A.5	Quy trình nghiệp vụ chức năng chỉnh sửa bài báo . . . . .	69
Hình A.6	Quy trình nghiệp vụ xây dựng mô hình gợi ý . . . . .	70
Hình B.1	Lịch sử hình thành container . . . . .	71
Hình B.2	Các thành phần của K8s . . . . .	73
Hình B.3	Các thành phần điều phối pod . . . . .	74
Hình B.4	HA Hệ thống HDFS bằng QJM . . . . .	76
Hình B.5	Các thành phần của HDFS . . . . .	77
Hình B.6	Các thành phần của YARN . . . . .	77
Hình B.7	Trình tự thực thi của một ứng dụng trên YARN . . . . .	78
Hình B.8	Các thành phần hoạt động của Spark . . . . .	80

## DANH MỤC BẢNG BIỂU

Bảng 2.1	Danh sách các use case . . . . .	8
Bảng 2.2	Đặc tả usecase Đăng nhập . . . . .	8
Bảng 2.3	Trường dữ liệu cho usecase đăng nhập . . . . .	9
Bảng 2.4	Đặc tả use case Xây dựng mô hình gợi ý . . . . .	9
Bảng 4.1	Các thành phần của hệ thống . . . . .	23
Bảng 4.2	Thuộc tính lớp News . . . . .	27
Bảng 4.3	Phương thức lớp News Repository . . . . .	28
Bảng 4.4	Phương thức lớp News Service . . . . .	28
Bảng 4.5	Phương thức lớp News LogView . . . . .	28
Bảng 4.6	Phương thức lớp News Controller . . . . .	29
Bảng 4.7	Thiết kế chi tiết cơ sở dữ liệu . . . . .	32
Bảng 4.8	Thiết kế chi tiết các trường được gửi đi . . . . .	33
Bảng 4.9	Danh sách thư viện và công cụ sử dụng . . . . .	33
Bảng 4.10	Liệt kê các Api . . . . .	35
Bảng 5.1	Cấu hình cụm Hadoop . . . . .	36
Bảng 5.2	Cấu hình MySql . . . . .	40
Bảng 5.3	Cấu hình AKS . . . . .	40
Bảng 5.4	Cấu hình HPA . . . . .	47
Bảng 5.5	Kết quả thực nghiệm HPA . . . . .	48
Bảng 5.6	Cấu hình sử dụng huấn luyện mô hình . . . . .	49
Bảng 5.7	Mô tả dữ liệu hành vi người dùng . . . . .	50
Bảng 5.8	Mô tả dữ liệu bài báo . . . . .	50
Bảng 5.9	Các tham số lựa chọn cho mô hình . . . . .	51
Bảng 5.10	Kết quả mô hình NAML trên tập dữ liệu MIND . . . . .	51
Bảng 5.11	Kết quả mô hình NAML trên tập dữ liệu công ty VCCorp . . . . .	52
Bảng A.1	Đặc tả usecase Đăng nhập . . . . .	56
Bảng A.2	Trường dữ liệu cho usecase đăng nhập . . . . .	57
Bảng A.3	Đặc tả usecase Đăng ký . . . . .	57
Bảng A.4	Trường dữ liệu cho usecase đăng ký . . . . .	58
Bảng A.5	Đặc tả usecase Tìm kiếm bài báo . . . . .	58
Bảng A.6	Trường dữ liệu cho usecase đăng ký . . . . .	59
Bảng A.7	Đặc tả usecase Đọc bài báo . . . . .	59
Bảng A.8	Đặc tả usecase Tạo bài báo . . . . .	59

Bảng A.9	Trường dữ liệu cho usecase tạo bài viết . . . . .	60
Bảng A.10	Đặc tả usecase Chỉnh sửa bài báo . . . . .	60
Bảng A.11	Đặc tả usecase Xóa bài viết . . . . .	61
Bảng A.12	Đặc tả usecase Đăng xuất . . . . .	62
Bảng A.13	Đặc tả usecase Tạo chủ đề . . . . .	62
Bảng A.14	Trường dữ liệu cho usecase tạo chủ đề . . . . .	63
Bảng A.15	Đặc tả usecase Lấy danh sách bài viết theo chủ đề . . . . .	63
Bảng A.16	Đặc tả usecase Lấy danh sách bài viết được đề xuất . . . . .	64
Bảng A.17	Đặc tả usecase Xây dựng mô hình gợi ý . . . . .	64
Bảng A.18	Đặc tả usecase Thống kê trực quan hóa dữ liệu . . . . .	65

## DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

Thuật ngữ	Ý nghĩa
AKS	Dịch vụ kubernetes của Microsoft Azuze (Azuze Kubernetes Service)
API	Giao diện lập trình ứng dụng (Application Programming Interface)
CD	Triển khai liên tục (Continuous Delivery)
CI	Tích hợp liên tục (Continuous Integration)
CTR	Tỉ lệ người dùng click vào các bài báo được gợi ý trên tổng lượt hiển thị (Click Through Rate)
DDD	Thiết kế hướng theo miền (Design-Domain Driven)
DOM	Mô hình Các Đối tượng Tài liệu (Document Object Model)
HDFS	Hệ thống tệp phân tán Hadoop (Hadoop Distributed File System)
HPA	Cơ chế tự động mở rộng theo chiều ngang (Horizontal Pod Autoscaling)
HTML	Ngôn ngữ Đánh dấu Siêu văn bản (HyperText Markup Language)
K8S	Dịch vụ điều phối container Kubernetes
MIND	Tập dữ liệu về báo của Microsoft (Microsoft News Dataset)
NAML	Mô hình gợi ý dựa trên chú ý đa góc nhìn (Neural News Recommendation with Attentive Multi-view Learning)
YARN	Trình quản lý tài nguyên của Hadoop (Yet Another Resource Negotiator)

# CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

## 1.1 Đặt vấn đề

Trong những năm gần đây với sự phát triển mạnh mẽ của internet, các phương thức truy nhập thông tin truyền thống đã dần lỗi thời, thay vào đó là sự phát triển của báo điện tử vì độ tiện dụng, dễ tiếp cận của nó. Đi đôi với điều này là một lượng truy cập lớn, đặt ra yêu cầu về một hệ thống đủ tốt để đáp ứng nhu cầu của người dùng với độ tin cậy cao, độ trễ thấp.

Cũng như những kênh tin tức truyền thống khác, doanh thu chủ yếu của các trang báo điện tử tới từ việc đặt quảng cáo. Nắm bắt được các đặc điểm về người dùng sẽ giúp nhà phân phối quảng cáo đưa ra các đề xuất quảng cáo tối đúng đối tượng. Vì vậy, cần có công cụ lưu trữ dữ liệu cá nhân, hành vi người dùng. Bài toán đặt ra là cần một hệ thống có thể đáp ứng nhu cầu lưu trữ, tính toán trên lượng dữ liệu lớn này.

Thời buổi kinh tế thị trường phát triển hiện nay đã kéo theo sự cạnh tranh trong kinh doanh. Một sản phẩm công nghệ tốt cần phải đáp ứng đầy đủ các yêu cầu của người dùng, giúp người dùng có những trải nghiệm thoải mái, dễ chịu nhất. Hệ gợi ý ra đời để giúp doanh nghiệp hiểu được khách hàng, đề xuất cho họ những điều mong muốn, giúp tăng trải nghiệm người dùng. Trong lĩnh vực báo thì gợi ý là đề xuất cho khách hàng những bài viết mà người dùng đó có khả năng cao sẽ quan tâm dựa theo dữ liệu lịch sử. Hệ gợi ý đóng vai trò quan trọng với một trang báo điện tử trong việc giữ chân khách hàng, tăng lượng người dùng. Hệ gợi ý tốt sẽ giúp tăng lượng truy cập trang, làm tiền đề tăng doanh thu cho sản phẩm báo ví dụ từ việc bán quảng cáo.

## 1.2 Mục tiêu và phạm vi đề tài

Hiện nay, các ứng dụng cũ thường triển khai theo kiến trúc một khối (monolithic architecture) vì tính đơn giản khi phát triển và triển khai. Monolithic có xu hướng phù hợp với những dự án có quy mô nhỏ, tất cả có thể đóng gói trong một dự án. Tuy nhiên việc gói gọn tất cả trong một sẽ khiến tốn thêm nhiều chi phí trong việc bảo trì hay mở rộng chức năng cho phần mềm, tính ổn định không cao do bất cứ lỗi ở phần nào cũng dẫn tới việc chương trình bị ngừng. Bên cạnh đó là vấn đề về một dịch vụ con trong hệ thống đóng vai quan trọng không thể tự mở rộng một cách riêng biệt, dẫn đến lãng phí tài nguyên vì toàn bộ ứng dụng sẽ phải mở rộng theo. Kiến trúc chia khối nhỏ (microservice architecture) đã dần trở lên phổ biến trong những năm gần đây nhờ khả năng chuyên biệt hóa các thành phần và khả năng mở rộng cho những thành phần cần thiết.

Các hệ thống thông thường sẽ lưu dữ liệu trong các cơ sở dữ liệu quan hệ hoặc không quan hệ như MySQL, MongoDB,... Các cơ sở dữ liệu này chỉ có thể đáp ứng được những nhu cầu lưu trữ cơ bản, làm việc với lượng dữ liệu nhỏ. Big Data là các tập dữ liệu có khối lượng lớn và phức tạp, có thể bao gồm các dữ liệu có cấu trúc, không có cấu trúc và bán cấu trúc, thay đổi nhanh và phức tạp. Để có thể đáp ứng nhu cầu lưu trữ và tính toán khổng lồ này, hệ sinh thái của Hadoop ra đời, được thiết kế để mở rộng từ một máy chủ duy nhất sang hàng ngàn máy khác để có thể xử lý song song hóa các tác vụ.

Hệ thống gợi ý (recommender systems) là một dạng của hệ hỗ trợ ra quyết định, học từ người dùng và gợi ý các sản phẩm tốt nhất trong số các sản phẩm phù hợp, mang tính cá nhân hóa cao. Phương pháp phân tách ma trận (matrix factorization [1]) là thuật toán đơn giản trong việc tách và học được hai biểu diễn của người dùng và sản phẩm dựa trên một lịch sử tương tác. Tuy nhiên phương pháp này hoàn toàn bỏ qua những thông tin người dùng, thông tin sản phẩm mà chỉ quan tâm tới việc người dùng đã đánh giá các sản phẩm như thế nào để đưa ra các gợi ý cho người dùng trong tương lai.

### **1.3 Định hướng giải pháp**

Microservice ra đời để giải quyết vấn đề về sự đóng gói tập trung của kiến trúc Monolithic. Microservice đề cập đến quá trình phát triển độc lập, tương đối nhỏ theo hướng chia hệ thống ra thành các dịch vụ (service). Mỗi service này đều có một nghiệp vụ riêng, một trách nhiệm riêng và có thể được triển khai riêng biệt, có một mối liên hệ lỏng lẻo với nhau. Kubernetes là một mã nguồn mở được dùng để tự động triển khai, mở rộng quy mô (scaling), quản lý các service dưới dạng các container. Kubernetes cho phép người dùng xây dựng các dịch vụ ứng dụng mở rộng nhiều container, lên lịch các container đó trên một cụm máy chủ (cluster), mở rộng các container và quản lý tình trạng của các container theo thời gian.

Hadoop là một mã nguồn mở cho phép phát triển các ứng dụng phân tán (distributed processing) để lưu trữ và quản lý các tập dữ liệu lớn. Hadoop hiện thực mô hình MapReduce, mô hình mà ứng dụng sẽ được chia nhỏ ra thành nhiều phân đoạn khác nhau được chạy song song trên nhiều thiết bị khác nhau.

Với hệ thống gợi ý, cách tiếp cận dựa trên nội dung (content-based) sẽ gợi ý cho người dùng những sản phẩm có nội dung tương tự với những sản phẩm mà người dùng đã có tương tác tích cực. Mô hình sẽ tập trung vào việc xây dựng biểu diễn đặc trưng của sản phẩm rồi dự đoán sở thích của người dùng với một sản phẩm đề xuất theo dữ liệu lịch sử. Các mô hình tiêu biểu có thể kể đến như mô hình gợi ý dựa trên chú ý đa góc nhìn (Neural News Recommendation with Attentive

Multi-view Learning [2]), mô hình gợi ý dựa trên cơ chế tự chú ý (Neural News Recommendation with Multi-head Self-Attention [3])...

### 1.4 Bố cục đề án

Phần còn lại của báo cáo đề án tốt nghiệp bao gồm các chương sau.

Chương 2 sẽ khảo sát thực tiễn của các hệ thống đọc tin tức để thống nhất các yêu cầu của đề án cần giải quyết. Từ đó đồng thời đưa ra các đặc tả chức năng của hệ thống.

Trong Chương 3, đề án trình bày về các công nghệ sử dụng trong đề án và cơ sở lý thuyết cơ bản của một mô hình gợi ý tin tức.

Trong chương 4, đề án trình bày các thiết kế từ tổng quan tới chi tiết các dịch vụ của hệ thống, thiết kế cơ sở dữ liệu, mô hình triển khai các thành phần của hệ thống.

Chương 5 đề án sẽ trình bày về kết quả cài đặt, các minh họa cho sản phẩm, một số kiểm thử phần mềm và phần kết quả đánh giá hệ gợi ý tin tức.

Chương 6 đề án sẽ đưa ra kết luận và hướng phát triển của đề án.