Design Document

for

Development of Mining Minds Core Technology Exploiting Personal Big Data

(Project No. 10049079)

Professor Sungyoung Lee

Ubiquitous Computing Laboratory
Kyung Hee University, Korea

Kyung Hee University

May 16, 2014

Contents

1. Introduction 9
   1.1. Definition 9
   1.2. Motivation 9
   1.3. Related Work (Problem Statement) 9
   1.4. Proposed Idea 11
   1.5. Abstract View 12
   1.6. Functional View 13
   1.7. Use Case Diagram 14
   1.8. Sequence Diagram 15
   1.9. Macro Level Interfaces 16
   1.10. Deployment Diagram 17
   1.11. Development Environment (Tools) 18
   1.12. Comparison of Existing and Our Solution 19
2. Data Curation Layer 21
   2.1. Data Curation Layer Functional Diagram 21
   2.2. Data Curation Layer Use Case Diagram 22
   2.3. Data Curation Layer Sequence Diagram 22
   2.4. Data Curation Layer Component Diagram 23
   2.5. Data Streaming and Communication 23
      2.5.1. Introduction 23
      2.5.2. Related Work 23
      2.5.3. Components Description 24
      2.5.4. UML Diagrams 25
      2.5.5. As Is - To Be Systems 26
      2.5.6. Uniqueness 26
      2.5.7. Summary 26
   2.6. Data Representation and Mapping 27
      2.6.1. Introduction 27
      2.6.2. Related Work 27
      2.6.3. Components Description 28
      2.6.4. UML Diagrams 29
      2.6.5. As Is – To Be Systems 30

2.6.6. Uniqueness 30
2.6.7. Summary 30
2.7. Data Curation 30
2.7.1. Introduction 30
2.7.2. Related Work 31
2.7.3. Components Description 31
2.7.4. UML Diagrams 34
2.7.5. As Is – To Be Systems 35
2.7.6. Uniqueness 35
2.7.7. Summary 35
2.8. Visualization 35
2.8.1. Introduction 35
2.8.2. Related Work 35
2.8.3. Components Description 36
2.8.4. UML Diagrams 37
2.8.5. As Is – To Be Systems 38
2.8.6. Uniqueness 38
2.8.7. Summary 38
2.9. Security and Privacy 39
2.9.1. Introduction 39
2.9.2. Related Work 39
2.9.3. Components Description 39
2.9.4. UML Diagrams 41
2.9.5. As Is – To Be Systems 42
2.9.6. Uniqueness 42
2.9.7. Summary 42
2.10. Overall Data Curation Layer Uniqueness 43
3. Introduction to Information Curation Layer 45
3.1. Information Curation Layer Functional View 45
3.2. Information Curation Layer Use Case Diagram 46
3.3. Information Curation Layer Sequence Diagram 46
3.4. Information Curation Layer Component Diagram 47
3.5. Personalized Feature Extraction 47

3.5.1. Introduction 47
3.5.2. Components Description 47
3.5.3. UML Diagrams 48
3.5.4. Uniqueness 49
3.5.5. Summary 49
3.6. Low Level Context Awareness 49
3.6.1. Introduction 49
3.6.2. Related Work 50
3.6.3. Components Description 53
3.6.4. UML Diagrams 56
3.6.5. As Is – To Be Systems 57
3.6.6. Uniqueness 58
3.6.7. Summary 59
3.7. High Level Context Awareness 59
3.7.1. Introduction 59
3.7.2. Related Work 60
3.7.3. Components Description 61
3.7.4. UML Diagrams 62
3.7.5. As Is – To Be Systems 63
3.7.6. Uniqueness 63
3.7.7. Summary 64
3.8. Long/ Short Term Behavior Analysis 64
3.8.1. Introduction 64
3.8.2. Related Work 64
3.8.3. Components Description 65
3.8.4. UML Diagrams 67
3.8.5. As Is – To Be Systems 68
3.8.6. Uniqueness 69
3.8.7. Summary 69
3.9. Information Curation Layer Overall Uniqueness 70
4. Introduction to Service Curation Layer 72
4.1. Service Curation Layer Functional Diagram 72
4.2. Service Curation Layer Use Case Diagram 73

4.3. Service Curation Layer Sequence Diagram 73
4.4. Service Curation Layer Component Diagram 74
4.5. Reasoner and Predictor 74
4.5.1. Introduction 74
4.5.2. Related Work 74
4.5.3. Components Description 75
4.5.4. UML Diagrams 77
4.5.5. As Is – To Be Systems 78
4.5.6. Uniqueness 78
4.5.7. Summary 78
4.6. Knowledge Maintenance Engine 78
4.6.1. Introduction 78
4.6.2. Related Work 79
4.6.3. Components Description 80
4.6.4. UML Diagrams 82
4.6.5. As Is – To Be Systems 83
4.6.6. Uniqueness 83
4.6.7. Summary 83
4.7. Feedback 83
4.7.1. Introduction 83
4.7.2. Related Work 84
4.7.3. Components Description 84
4.7.4. UML Diagrams 85
4.7.5. As Is – To Be Systems 87
4.7.6. Uniqueness 87
4.7.7. Summary 87
4.8. Recommendation 87
4.8.1. Introduction 87
4.8.2. Related Work 87
4.8.3. Components Description 88
4.8.4. UML Diagrams 89
4.8.5. As Is – To Be Systems 90
4.8.6. Uniqueness 90

4.8.7. Summary 90
4.9. UX/ UI Authoring Tool 91
4.9.1. Introduction 91
4.9.2. Related Work 91
4.9.3. Components Description 92
4.9.4. UML Diagrams 94
4.9.5. UI/ UX Authoring Tool Service Scenario 95
4.9.6. As Is – To Be Systems 99
4.9.7. Uniqueness 99
4.9.8. Summary 99
4.10. Service Curation Layer Overall Uniqueness 100
5. Service Scenarios 102
5.1. 전체 서비스 시나리오 102
5.1.1. 신체활동 관리를 위한 Exercise curation 서비스 104
5.1.2. 수면관리를 위한 Quality of Sleep Curation 서비스 105
5.1.3. 물 섭취를 위한 Drinking Water Curation 서비스 106
5.2. 서비스 시나리오별 As is – To be 전략 107
5.2.1. 활동량 기반 4050 Healthy life management 서비스 107
5.2.2. 당뇨관리 서비스 108
5.2.3. 수면 모니터링 서비스 109
5.2.4. 물 섭취 관리 서비스 110
5.2.5. Survey platform 111
5.2.6. SNS 기반 빅 데이터 분석 113
6. Conclusion 114

Chapter 1
Introduction to Mining Minds

Abstract

Human interactions are generating a variety of data, such as text, location, sound, sensory, and multimedia data, which are dynamically produced from multiple sources with different data structures or even no data structures. Therefore, it is necessary to find a way to effectively analyze, integrate and mine personal big data that records people's information behaviors and social activities in both cyber space and real world.

Traditionally, analysis has been performed on archived or stale data; however, with the Zeta bytes of personal data being produced, utilization of this data for real-time needs has become of utmost important. Existing systems only focus on the volume of such data, and not focus much on the velocity and variety of data. These aspects are important for the maintenance of high quality of data to provide high quality of service.

The biggest hurdle for the success of legacy systems is maintenance of high quality of content and services. The quality of content can be maintained with evolution of the knowledge in the knowledge bases. While, the quality of services can be provided with improved and adaptive user interfaces to the users. Therefore, we propose a Mining Minds platform that is based on layered architecture to support high quality personalized services.

Personalized services provision should evolve with the evolution in the knowledge bases that is based on maintaining contextual information of the users. Therefore, we proposed secured mining minds platform in the layered architecture that can curate the data, information, and services with adaptive user interfaces. The two prominent and unique modules to facilitate the objectives of evolution and adaption are Knowledge Maintenance Engine (KME) and Adaptive User Interface/ User Experience (AUI/ AUX) respectively.

Mining minds platform takes benefit from the technology of big data with respect to variety as well as volume, mapping of life events through sensory environment and reasoning and prediction to process the real-time data for providing personalized services. The platform will provide the benefits to users in the form of personalized life quality improving services, silver business services, proactive way to control the chronic disease services, and life care services.

1. Introduction

Human generates lot of data by using sensory devices, cameras, social media, smart phones, and web. This generated is diverse in nature, high volume and velocity. It can be utilized for provision of personalized healthcare services to the users. In order to provide the services, we need to find **a way** that can **effectively** analyze, integrate and **mine personal big data**. Furthermore, it should provide people's behaviors information and social activities in both **cyber space** and **real world**. This can be made possible by meaningful interpretation of the collected information and providing services that can help in the management of their life styles.

1.1. Definition

"Mining Minds is a collection of **services**, **tools**, and **techniques**, working collaboratively to investigate human's **daily-life routine** data generated from **heterogeneous resources**, for people's wellbeing and health-care support"

1.2. Motivation

Our motivation is to model the **daily life events** using heterogeneous input sources and provide **personalized services** to **enhance human life style.** The proposed platform will find human life event patterns and facilitate user by providing customized services. We want to develop a novel platform based on our design philosophy that revolves around the concepts of evolutionary knowledge maintenance engine, adaptive user interface, and adaptive user experience. These paradigms are equally important to provide the quality of services and adoption factor in daily routines. If quality is so good but presentation factor is not so impressive then the service has very low probability to adopt in daily routines and vice versa. In order to improve the quality of the services, maintenance of the system plays and important role. We proposed the concept of evolutionary knowledge bases, that is supported by automatically rules updating, experts involvement using easy to use authoring environment, and retrieving evidences from online resources for experts' education with state of the art methodologies. Another important aspect is provision of customized and personalized user interfaces that captures user preferences, platform usage and user's contextual information. The user interface should be adapted enough to engage the user's and provide services in terms of virtual caregiver.

1.3. Related Work (Problem Statement)

- **DNANexus** is a fine-grained authentication and access controls system that provides end-to-end encryption, strict production system access control, auditing, and scanning. It also provides API-based integration and reproducible and version-controlled analysis results. The customers include research scientists and clinical research partners in pharma and biotech.
- **NetBio** assembles vast amounts of curated and annotated clinical and molecular data. This technique enables clients to make unique discoveries that would not be possible with their own private datasets alone. It uses Big Data technology to make correlations between the billions of data points from the public domain with private genomic and clinical data sets. A rich set of APIs enable clients to integrate NextBio within their workflows. Their current clients include Pharmaceutical R&D and academic medical centers. Also, the system initial focus was on oncology, now expanding into metabolic and autoimmune diseases.
- **Prefixition Software** pulls data from a variety of sources, using data mining, machine learning and mathematical algorithms to power predictions. It uses a predictive analytics algorithm to risk score patients upon admission and throughout their hospital stay, to identify those at risk of readmission before they leave the hospital, with 86% accuracy. The current project is applying analytics to

prevent MRSA infections and deaths in the hospital setting. Also, it is working to use predictive analytics as a tool for prevention of chronic disease – e.g., diabetes.

- **Well Doc** system is based on clinical requirements definition and provides user experience-driven solution design. It works on multi-party program coordination and management and mobile and web software development. The other services provides by this system includes integration and acceptance testing, device and network porting and certification, and regulatory filing and management.
- The comparison based on the use of technologies is given in Table 1 of the existing systems discussed above.

Table 1: Technologies Use of Existing Systems

Characteristcs	DNAnexus	NextBio	Prefixition Software	WellDoc	PKS	Proposed Platform
Cloud-based Solutions	Yes	Yes	Yes	Yes	No	**Yes**
Big Data	Yes	Yes	Yes	Yes	No	**Yes**
Domain	Healthcare	Healthcare	Healthcare	Healthcare	CDSS	**Lifecare**
Specialized Security	Yes	No	No	No	No	**Yes**
Personalization	No	No	No	No	No	**Yes**
Feedback Mechanism	No	No	No	No	Yes	**Yes**
Knowledge Management	No	No	No	No	Yes	**Yes**
Data Mining	Yes	Yes	No	No	No	**Yes**
AI Techniques	No	No	Yes	Yes	Yes	**Yes**

1.4. Proposed Idea

Most of the big data solutions are focusing on only volume of the data, whereas our design philosophy is to handle velocity and variety of data in addition to volume of data. Another key aspect with handling these big data challenges is related to securing the data, therefore we are dealing with the security and privacy of data as well. The main objective is to provide quality of personalized services to the users that is the motivation for our layered architecture. The data will pass through these layers and thus we define protocols for conversion between unstructured, semi-structured and structured data. Maintaining the personalized services requires evolution of knowledge bases that is performed with our Knowledge Maintenance Engine (KME). Maintenance Engine is also equipped with human experts as well as evidence supports. Presenting personalized services to the users in customized manner is another challenge. This aspect is handled with our UI/ UX Authoring Tool that defines personalized adaptive user interfaces for the users.

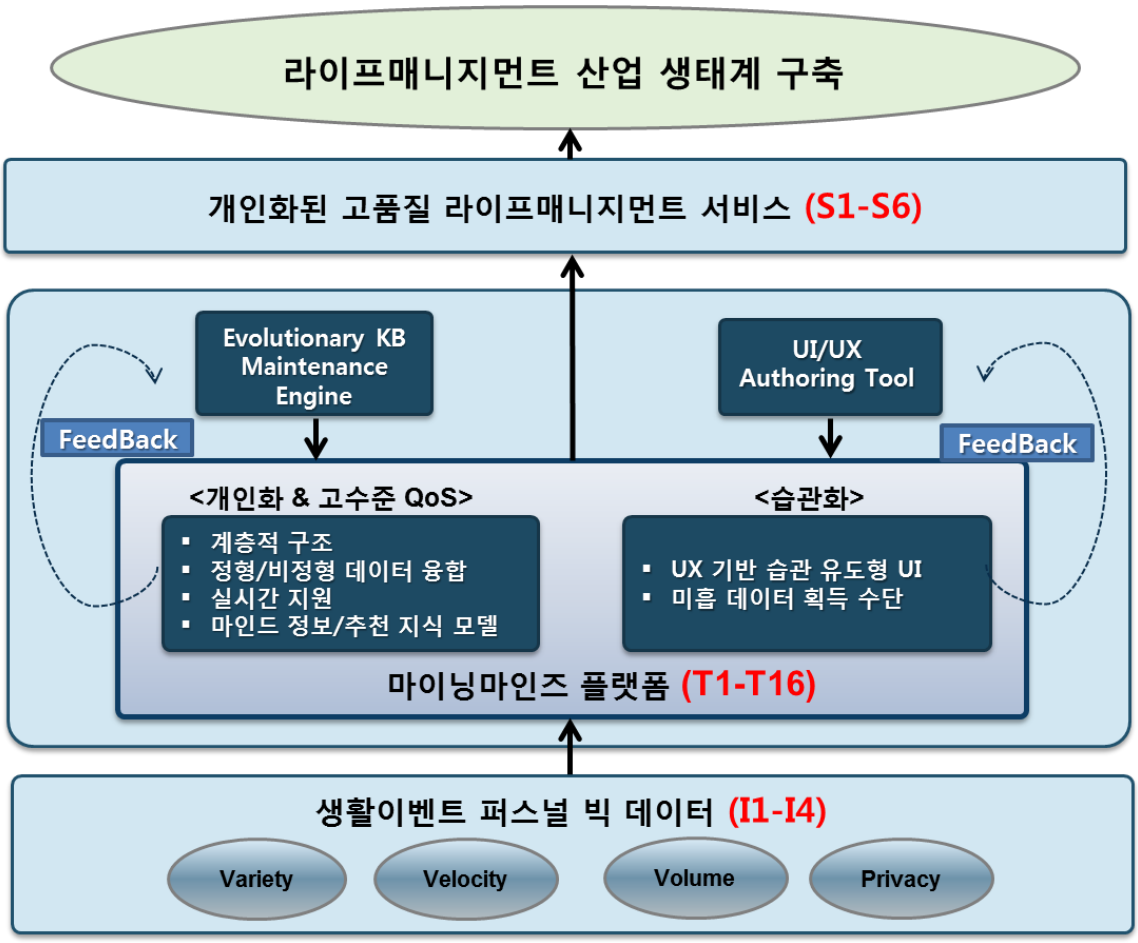


Figure 1.1: Philosophy of Mining Minds

1.5. Abstract View

Mining Minds platform uses heterogeneous data sources for the generation and storage of personalized big data. The platform is categorized into layered architecture for handling data volume, velocity, and variety for efficient big data management. The three layers are Data Curation, Information Curation, and Service Curation. Each layer process the data and stores it into its own knowledge base for provision of the personalized services. In order to improve the quality of services, feedback of the user is obtained to maintain the knowledge bases for personalized services.

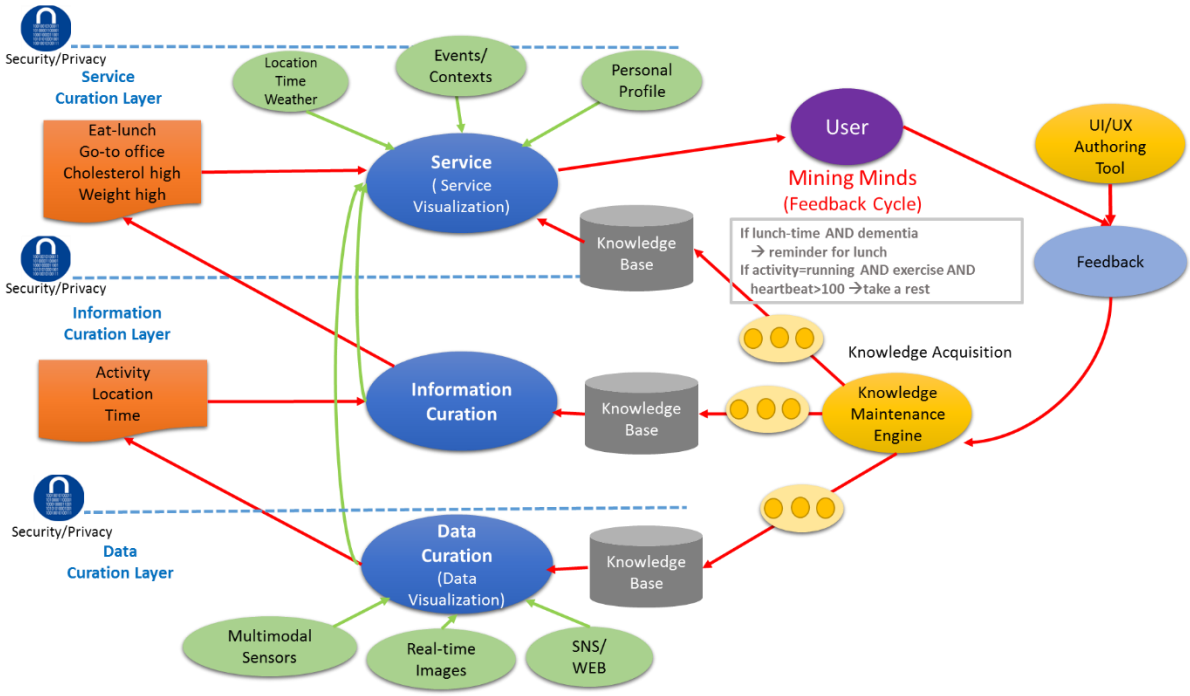


Figure 1.2: Philosophy of Service Scenario

1.6. Functional View

The Mining Minds platform is divided into three layers: Data Curation, Information Curation, and Service Curation. The data curation layers is responsible for accessing the real time data and curated through our proposed method and storing the information into Hadoop-based distributed file system. The data stored in the Hadoop-based distributed file system is processed for converting in to structured format to be processed by the upper cores in the form of RDBMS that is intermediate data. The service layers utilizes the information from the data layer and process the low level context data to convert to high level contextual information. The service Curation layer process the data with reasoning and prediction to find out the recommendations, that is presented to the user in a more elaborated form by UI/UX Authoring Tool. The evolution of the knowledge in the system is maintained by taking user feedback and all the layers data, information as well as processes are made secure by providing security and privacy. The architecture is shown in Figure 1.3.

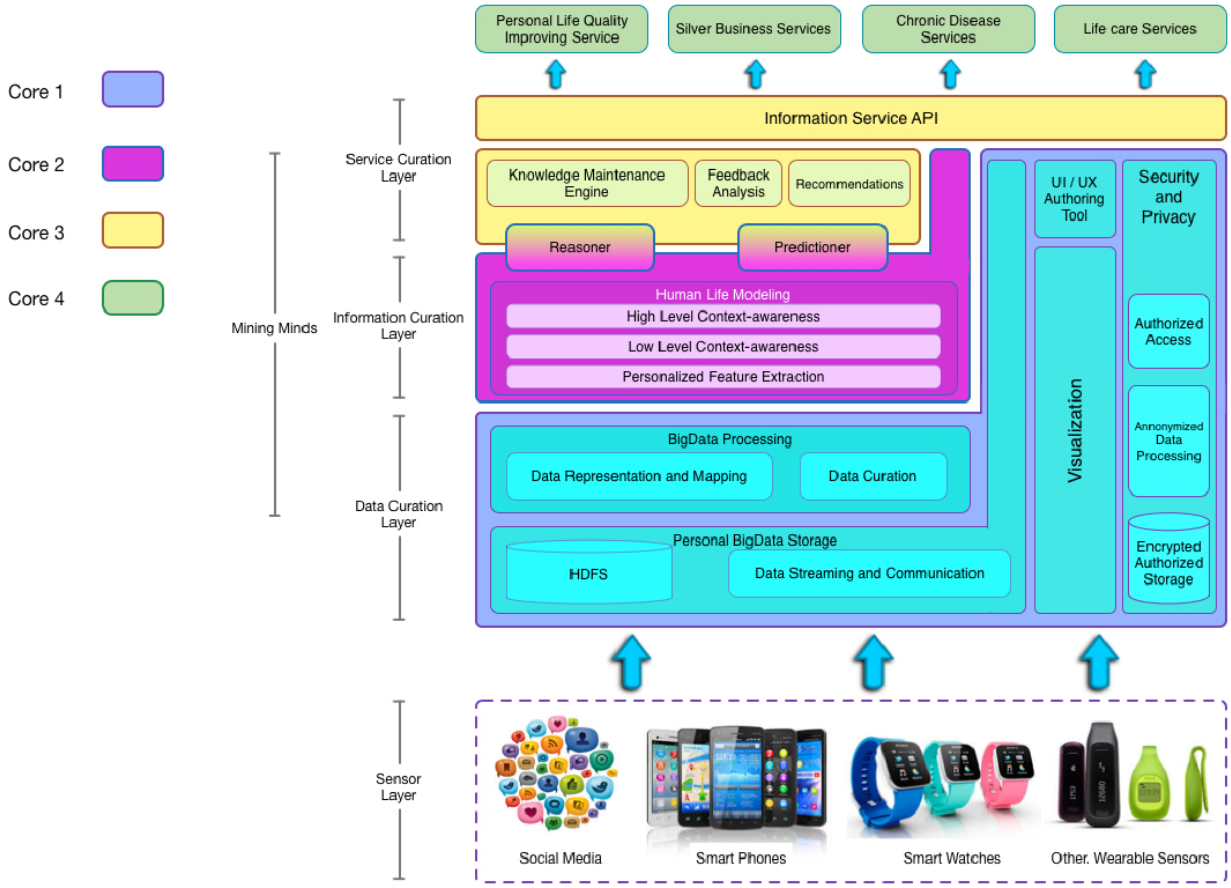


Figure 1.3: Mining Minds Framework

1.7. Use Case Diagram

The overall use case diagram of the proposed Mining Minds platform is shown in the following Figure 1.4.

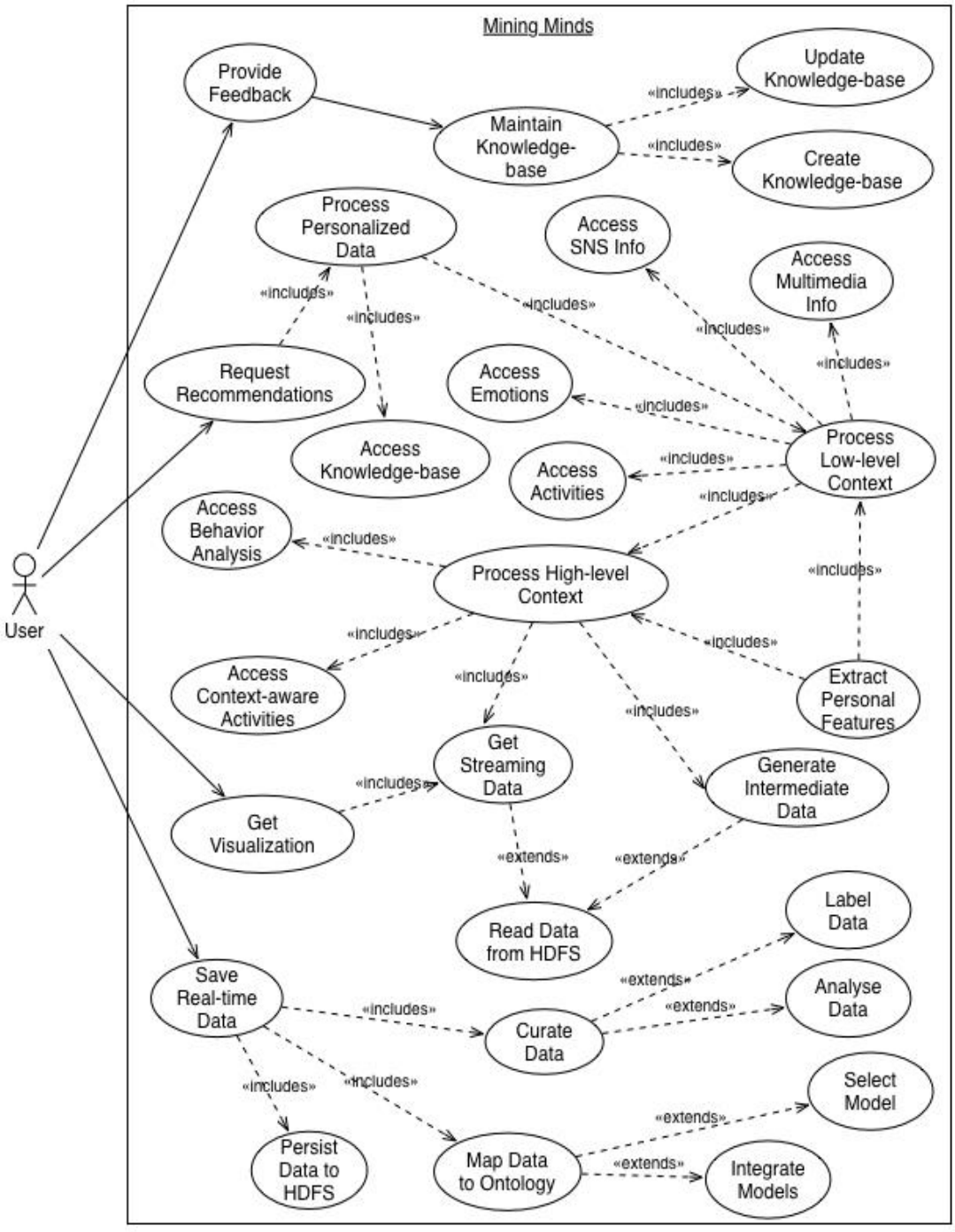


Figure 1.4: Overall Use Case Diagram

1.8. Sequence Diagram

The overall use case diagram of the proposed Mining Minds platform is shown in the following Figure 1.5.

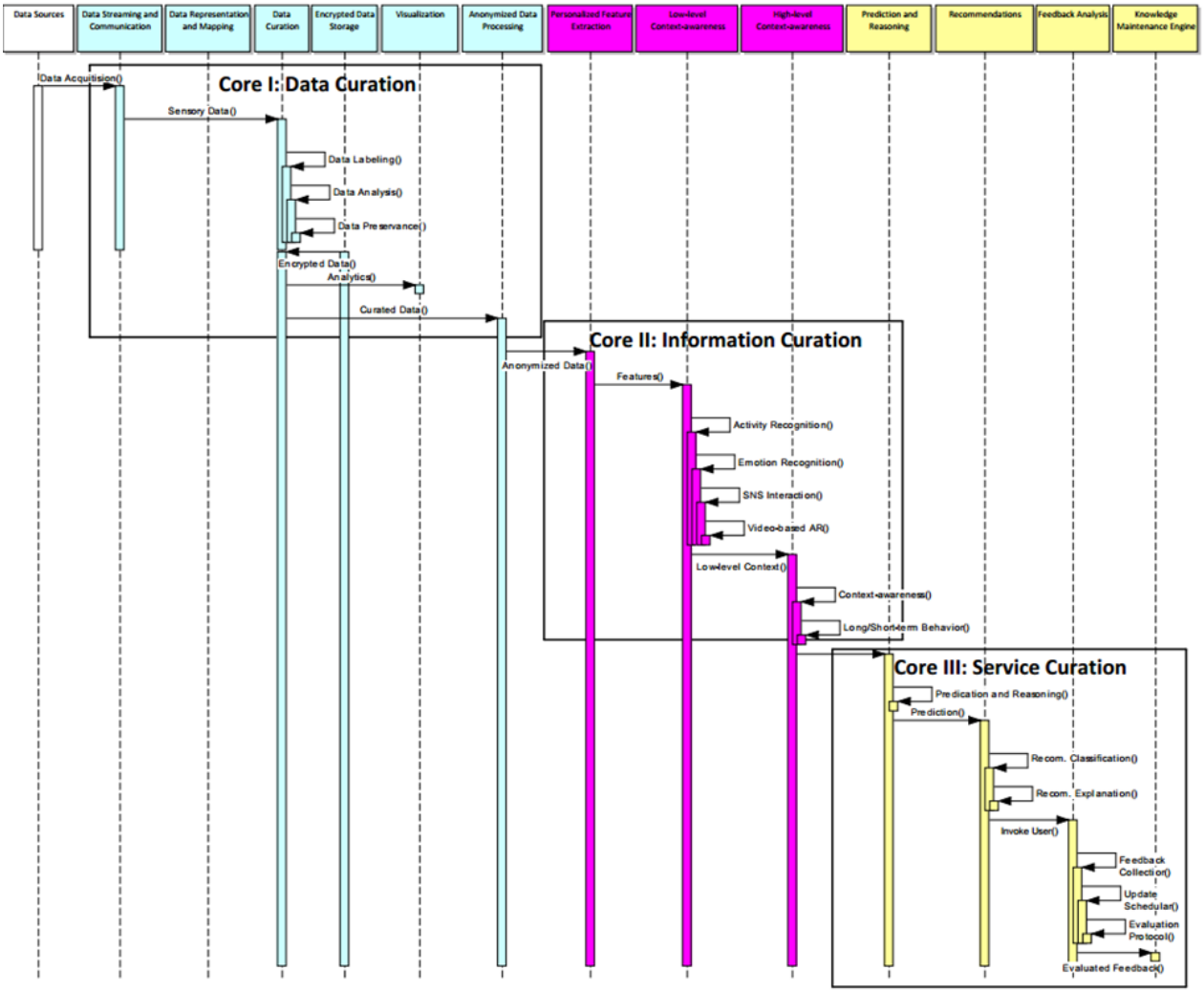


Figure 1.5: Overall Sequence Diagram

1.9. Macro Level Interfaces

The macro level interfaces of the proposed Mining Minds platform is shown in the following Figure 1.6. Initially, data is gathered about the user from different sources that includes sensors, smart phones, smart gear, and social media. These personalized data is curated in the data curation layer using data mapping and representation. This data is stored in the HDFS using Data Streaming and Communication in secured encrypted form through our security and privacy module. Data can be visualized to the user by the visualization module. Information layer can process the data stored in the data curation layer in anonymized form. Personalized Feature extraction extracts the data and processes it for Low-level Context Awareness and High-level Context Awareness. In the service curation layer, process authorization will take place and data is utilized by Reasoner and Prediction module to generate recommendations. These recommendations are further elaborated for user understanding using Recommendation Manager. Finally, UI/ UX Authoring Tool presents the final recommendations in users personalized format. Feedback Analysis provides the interface to the user to provide his feedback over the platform recommendations. The knowledge is maintained with evolution in the knowledge-bases and is maintained using Knowledge Maintenance Engine. Only the authorized users can accessed the platform through Security and Privacy module to handle the user authentication and authorization.

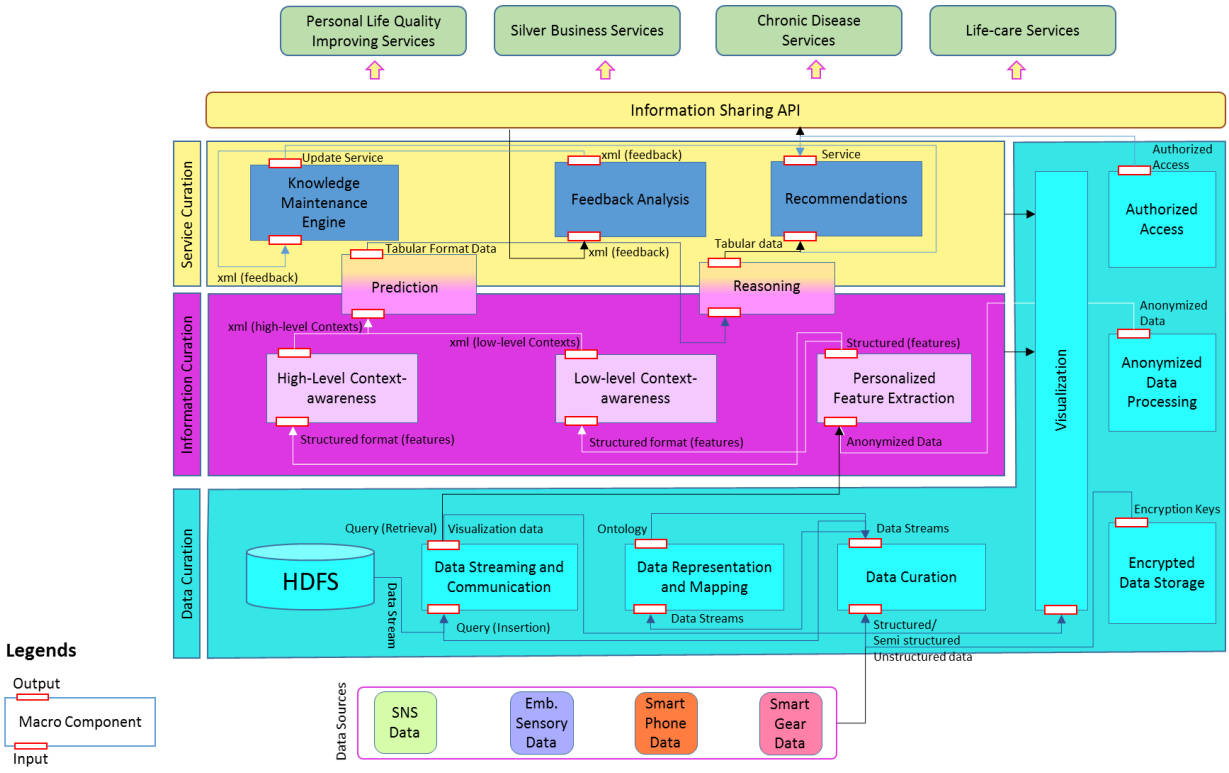


Figure 1.6: Macro Level Interfaces

1.10. Deployment Diagram

Our Proposed platform is deployed on a distributed cloud computing environment, where 3 families of computing virtual machines are collaborating as participating nodes. Each family of nodes hosts a respective core from mining minds architecture, Primary nodes 4 and 5 provide abstraction over the internal deployments of data curation. Data for information curation is integrated and provided to Node 8 by Node 5 and Node 4. Similarly for Service curation data is also integrated and provided to Node 10 by Node 5 and Node 4. The deployment diagram is shown in Figure 1.7.

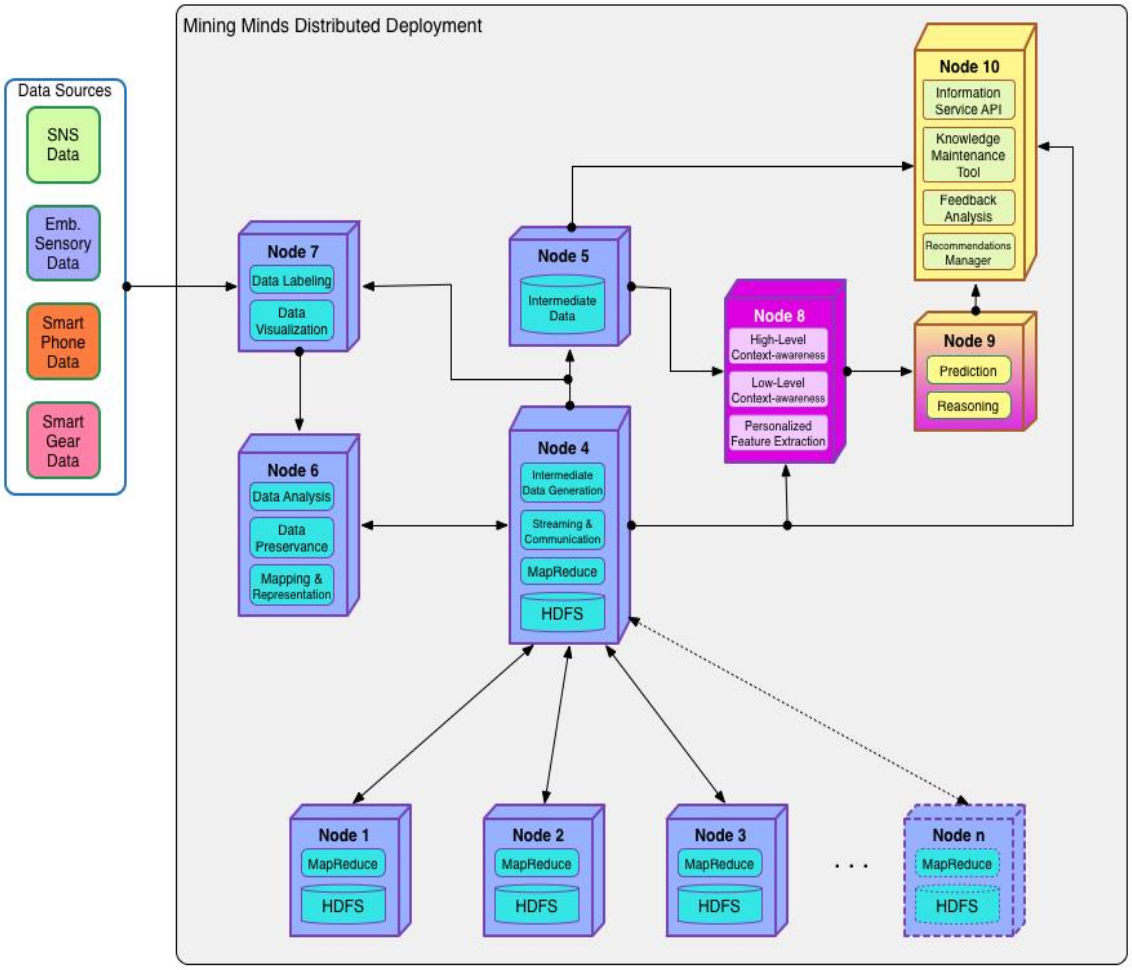


Figure 1.7: Deployment Diagram

1.11. Development Environment (Tools)

The under consideration technologies and development kits to develop Mining Minds are mentioned in Figure 1.8. As Mining Minds platform is a collaboration of several components hosted among the three layers, constituting specific goals and responsibilities. To ensure the unique contribution of individual components, several tools and techniques facilitate the platform at macro and micro level. Figure 8.1, provides an overview of candidate tools and techniques proposed to be utilized collectively for the development of Mining Minds Platform.

For deployment infrastructure, we will be using a cloud platform based on 150 node Dell PowerEdge R610 Rack Cluster, equipped with total of 1800 Cores of CPU, 3600 Gb of Memory , and 450 TB of disk storage. This infrastructure is already equipped with numerous open-source software and platforms; including Big Data support by HDFS with batch processing of MapReduce. For In-Memory and real-time query processing, Spark and Impala frameworks are used respectively.

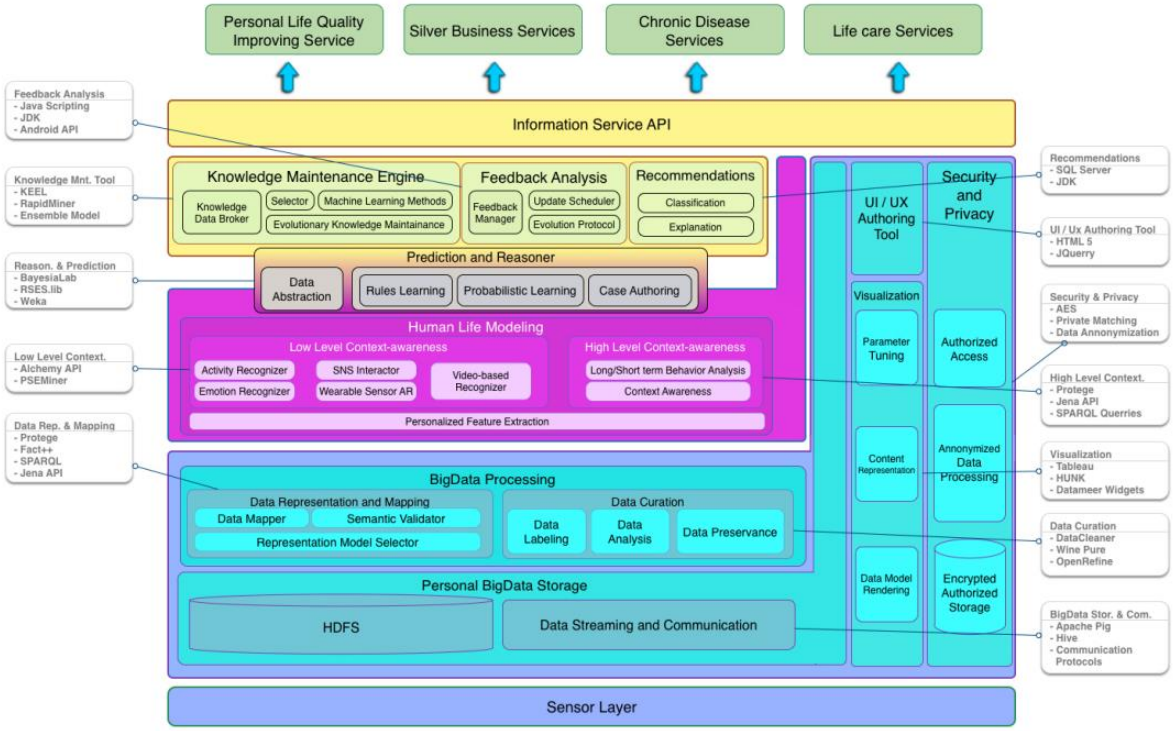


Figure 1.8: Technology Stack

1.12. Comparison of Existing and Our Solution

The comparison of existing approach and our proposed solution is shown in Figure: 1.9. Current status of the research community and already developed system focused on either data mining or artificial intelligence track. Data mining track can process only stale data to get the statistics of the past events. In case of artificial intelligence track, processing is done for the specific and independent source. Curation can play and important role to generate the quality of services and improve the quality of life. Our solution is considering heterogeneous aspects of sensory and SNS data to get the human life events. We are processing in a secured way to process the personalized data by data mining, artificial intelligence and involving minimal interaction of human experts. Consequently, our solution provides the high quality of services to the user through personalized user interface. The comparison is depicted in Figure 1.9.

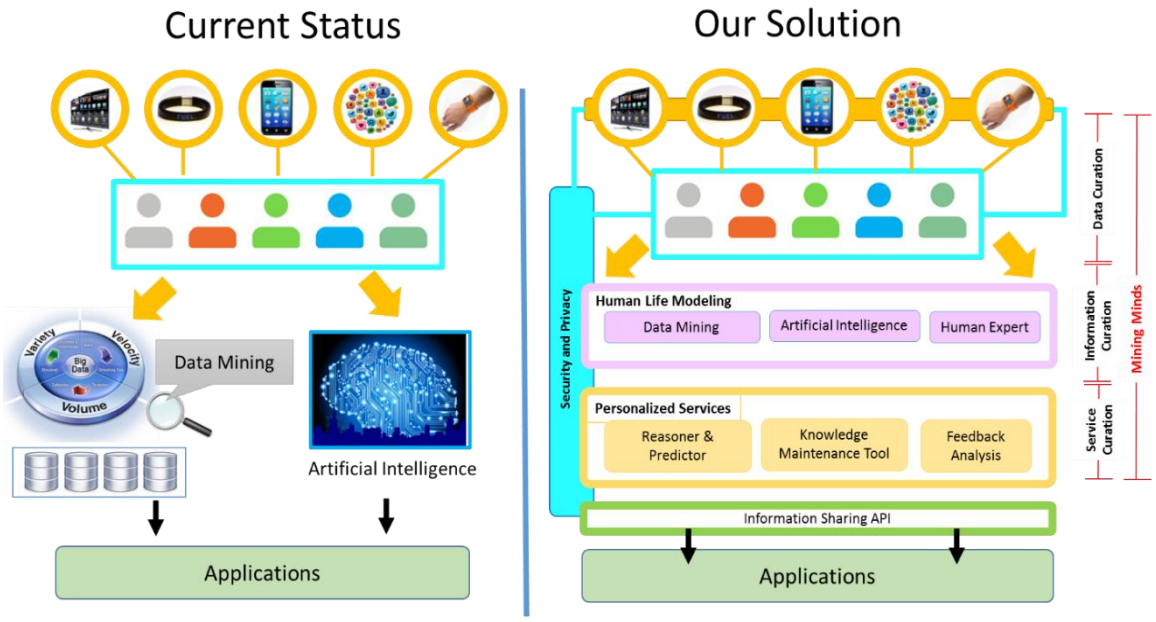


Figure 1.9: Existing Technology Comparison

Chapter 2
Data Curation Layer

2. Data Curation Layer

The data curation layer curates the data and securely stores into Hadoop with the help of data representation and mapping. The details of the key methods and technologies are mentioned in the following Figure 2.1 and details are provided in the sub-sequent sections.

2.1. Data Curation Layer Functional Diagram

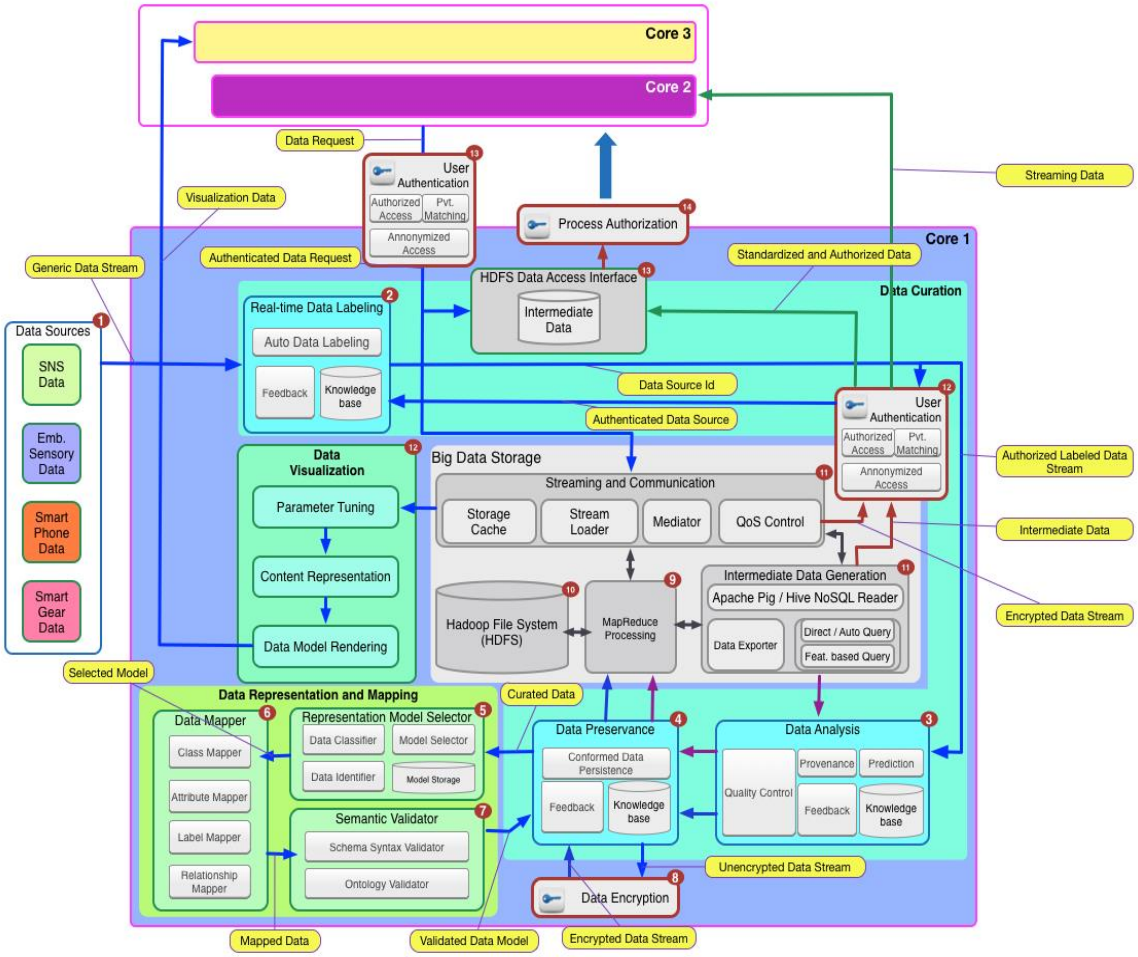


Figure 2.1: Data Curation Layer Functional Diagram

2.2. Data Curation Layer Use Case Diagram

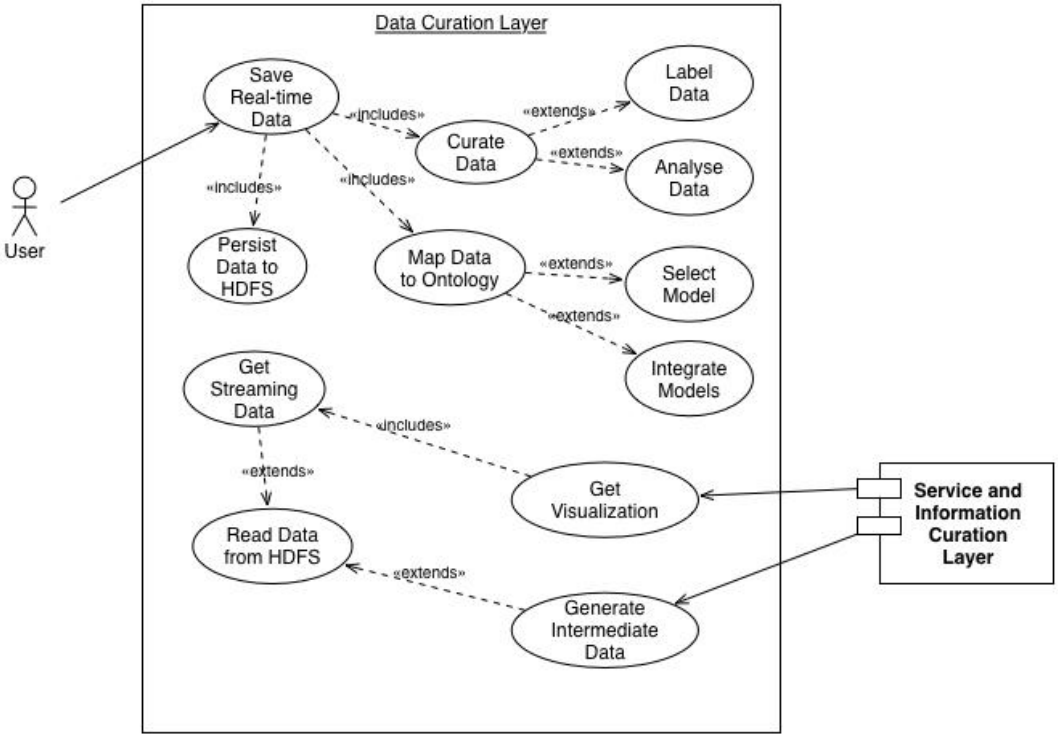


Figure 2.2: Use case diagram of data curation layer

2.3. Data Curation Layer Sequence Diagram

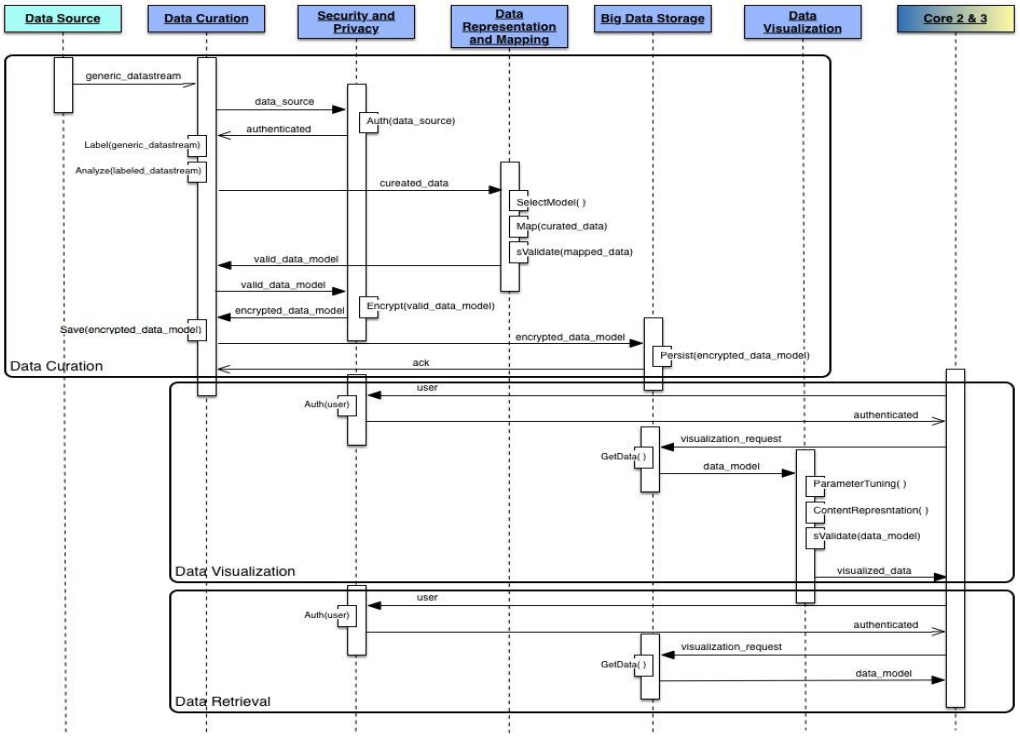


Figure 2.3: Sequence diagram of data curation layer

2.4. Data Curation Layer Component Diagram

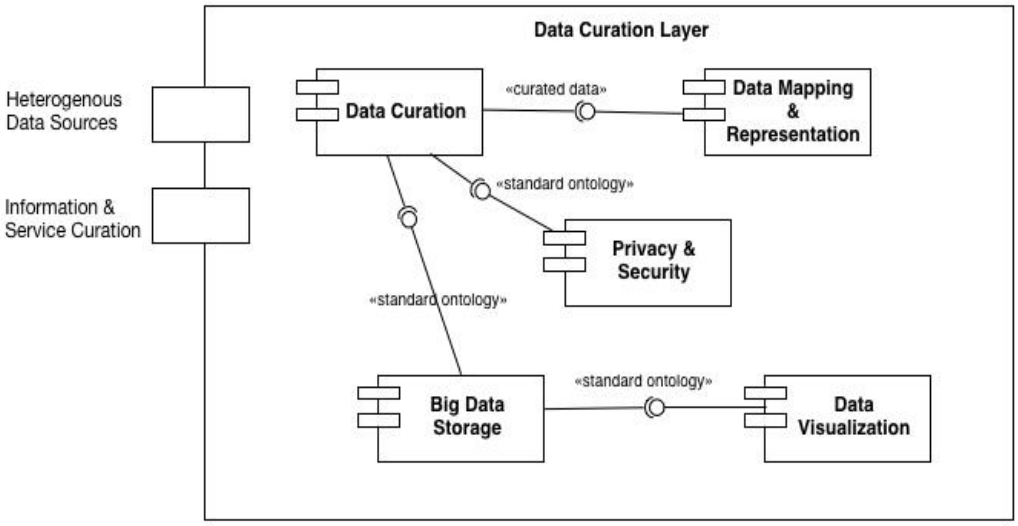


Figure 2.4: Component diagram of data curation layer

2.5. Data Streaming and Communication

2.5.1. Introduction

In this era of exponentially growing data from diverse input sources has given rise to the concept of big data. To handle such big data, Hadoop Distributed File System (HDFS) is used as storage framework and MapReduce as its processing framework. However MapReduce is a batch processing framework. To overcome the problem of batch processing, streaming and communication component provides real-time and online access to the data in HDFS and generate intermediate data for other components use.

2.5.2. Related Work

- Streaming in computing is a process of analyzing data in real time to increase the speed and accuracy. To cover the needs of streaming data generated exponentially, many systems have been proposed to incorporate real-time streaming in big data. Hadoop is one of the big data technologies which supports batch processing model as MapReduce.
- Big data Streaming **[**Zahria2010**]** and Spark **[**Shahrivari2014**]** discuss big data streaming and various systems developed for it. MapReduce can handle stream using a technique called micro-batching. Some systems have been proposed for big data processing without MapReduce and Hadoop, and they provide streaming like Spark, and Twitter Storm.
- Muppet [Lam2012] is a system that provides cloud based platform in which support of MapReduce is established. In this system, they process the stream of incoming data based on some event and the stream is processed in two phases including *map* phase and *update* phase. It is a system which includes MapReduce as part of cloud platform.
- Granules [Pallickara2009] is a cloud platform for stream processing at run time with support of iterative, periodic, and data driven semantics for individual components. Within this cloud environment, it provides the support of MapReduce to process the incoming streams. However it does not provide streaming directly to Hadoop system.
- In **[**Zinn2010**],** the authors have suggested a MapReduce as processing framework for large and numerous XML files as streams. They propose system for data parallelism through MapReduce in which XML documents are processed using XML-structured data. The process implemented in

MapReduce are based on "black box" scientific functions. They have listed their main contributions as development XML pipeline processing strategies in MapReduce and discussion on its pros and cons. Their work is mainly focused on how to streamline XML processing in MapReduce instead of real time processing.

- Streaming overhead in **[**Ding2011**]** discusses the Hadoop default streaming capacity, capability and their consequence and performance. It states that, although the streaming utility in Hadoop provides somewhat flexibility to the users to remove compulsion of using MapReduce programming instead they can write in any standard executable and execute as MapReduce jobs. However, it has some technical difficulties in improving Hadoop performance like it produces extra copies while executing such jobs which causes overhead. In the same way extra effort is made when there is conversion from streaming to standard input and output.

2.5.3. Components Description

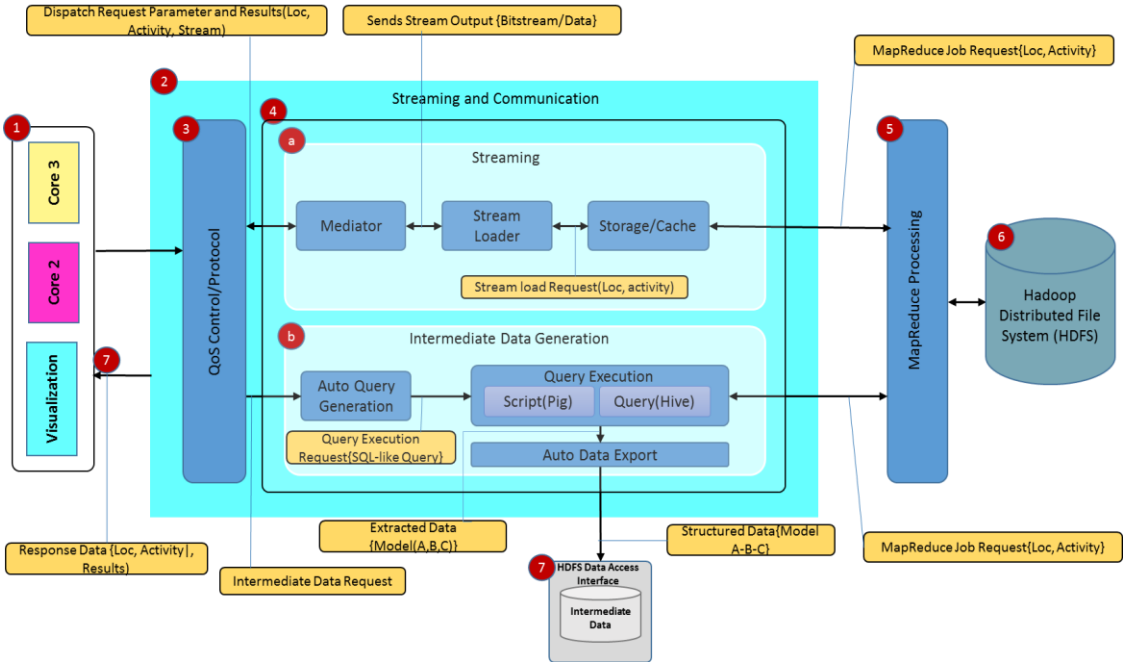


Figure 2.4: Architecture

Streaming

- Input: Data from data curation layer
- Output: Streaming data for visualization and layer 2 components
- Description: In the proposed solution, streaming is provided to support real time data access in big data environment.
- QoS Control and Protcol is a component, each of the Mining Mind components will request, and it is responsible for manipulating and fulfilling the request. It will identify the type of request to be fulfilled e.g., request of intermediate data or streaming data.
- Mediator component works as an agent/mediator that differentiates between types of request and selection of the data. It will be performing some selection on the data based on the incoming request parameters e.g., request can be querying data for visualization or data required by the other sub components. It will be selecting the exact data.

- Stream Loader is invoked once the mediator has performed its selection procedure, it will be loading data from the storage or generate new request to HDFS to load the data into storage. This generated stream will contain the requested data and sent into response.
- Storage/Cache is introduced to provide quick data transfer because MapReduce is a batch access model. It is in-memory storage that will be storing the already fetched results returned using MapReduce from HDFS. And these results will be contained for future reuse and can only be re-executed when some change is occurred in the data in HDFS.

Intermediate data generation

- Input: Intermediate data request and model/schema from other core layers.
- Output: Intermediate structured data
- Description: To provide access to HDFS data, we introduce intermediate data, which contains structured data required by the components in Mining Minds.
- Query Generation: It gets schema/model from each component of Mining Mind that needs data stored in HDFS to be in a structured format. After getting features/schema, it will be generating queries and scripts i.e., NoSQL queries to be executed using Hive and scripts using Apache Pig. The components in other layers like reasoned and predictor etc. would need to provide some features/parameters to extract data from HDFS and store it in intermediate database.
- Data Exporter: The data returned by the above component need to be safely transformed and exported to the database. This component will be performing a direct export function on a remote or local place for the data returned. It will also be responsible for generating schema in the intermediate database because, sometimes there will be new data requests coming, which may not already be in the database. Therefore, its functions include schema generation and data export. Once the data is exported, it can then be used by other components of Mining Minds.

2.5.4. UML Diagrams

- Sequence diagram

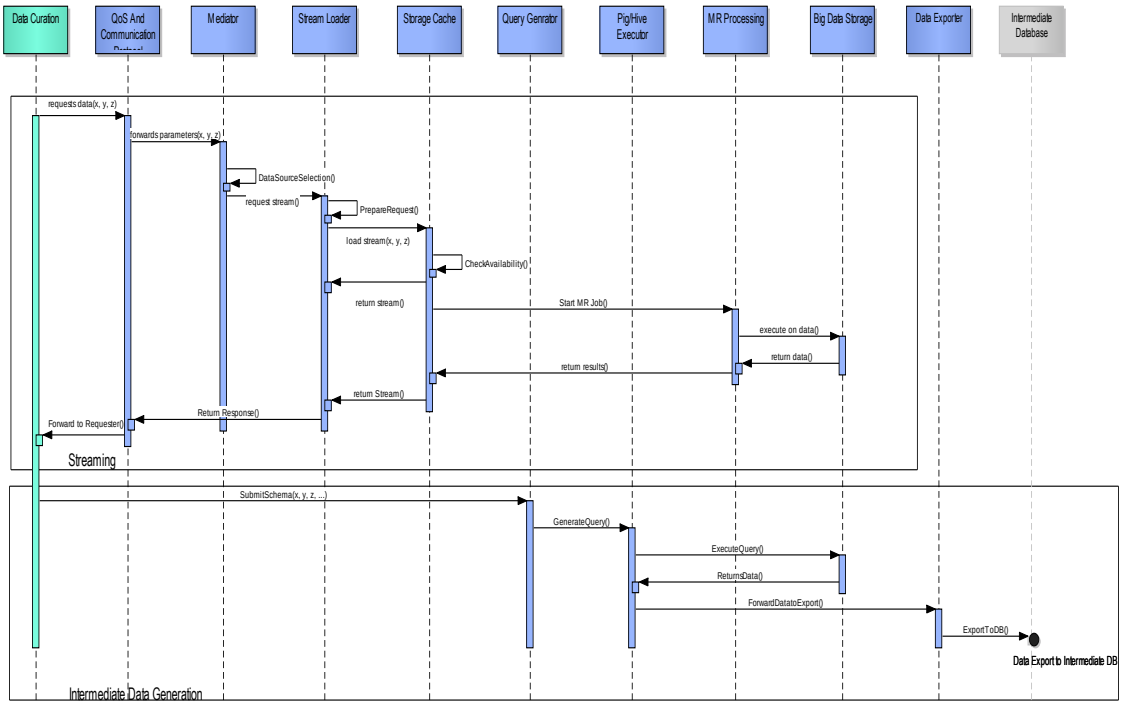


Figure 2.5: Sequence diagram

- Component diagram

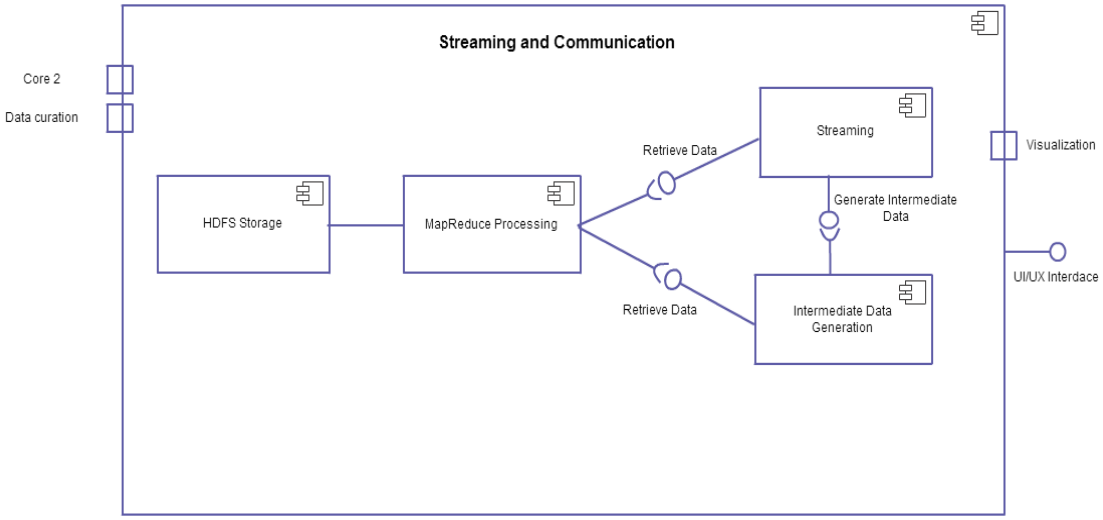
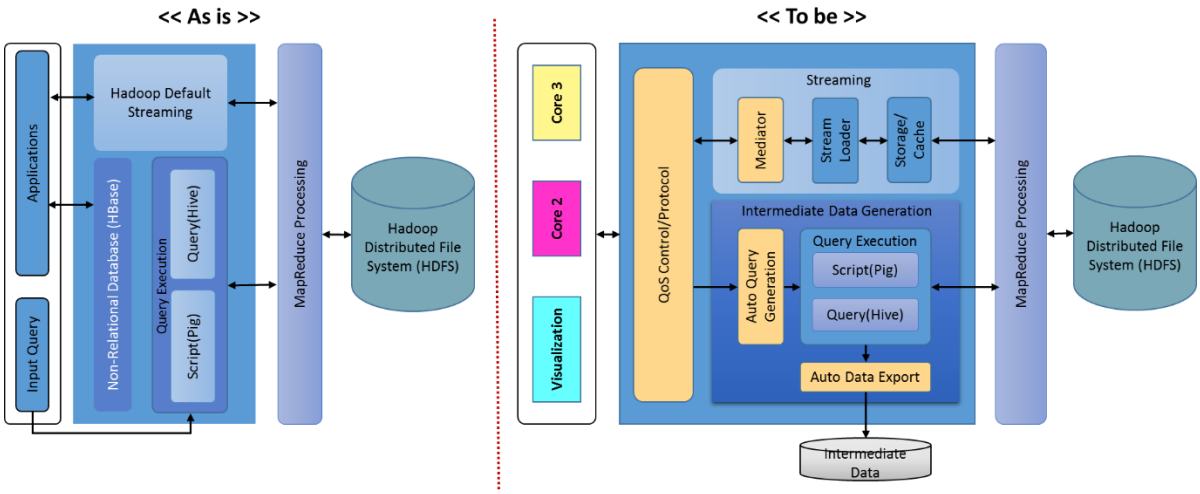


Figure 2.6: Component Diagram

2.5.5. As Is - To Be Systems



2.5.6. Uniqueness

- Development of a streaming and communication utility to provide higher layers subcomponents the real-time access to the data
- Provides adaptive data selection with granularity from integrated variety sources of data
- An automated query generation for generation of intermediate structured data
- Run-time intermediate schema generation and data export from NoSQL to structured intermediate data.

2.5.7. Summary

Streaming and communication module provides streaming facility to access big data in HDFS in real-time and generates intermediate data from HDFS based on the needs of higher layer

subcomponents. Based on the user/component requests, it provides adaptive data selection and intermediate data generation using automated query generation.

2.6. Data Representation and Mapping

2.6.1. Introduction

Data representation and mapping is emerging and integral part of a system after the data acquisition phase. After data acquisition the data passes through different phases of the data curation process to the Data Representation and Mapping module to conformance of data and to validate and verify the data. This module provides unified format of heterogeneous data to the data preservance. That preserves the unified data for further processing in upper layers of data curation. The complete process from data curation to data representation shown in figure 2.2.1, it depicts that this process model is working in sequential manner. First the heterogeneous data is collecting from various sensors, social media, profile, environment and smart TV then Data Representation and Mapping component identify, classify and dynamically select an appropriate representation model and conform, validate the data accordingly.

2.6.2. Related Work

- Data representation and mapping is emerging and integral part of every system after the data acquisition phase. Scalable and interoperable data representation models are playing vital role in context-aware systems, smart homes applications, and mobile environment and in many more systems.
- Different representation models like Ontological, Graphical, Tuple based, Object Oriented, Spatial, Domain Focused, Hierarchical, Hybrid are all usable but in latest systems the ontological representation model have high preferences.
- L. Chen et al. [Chen2012] have presented a generic system architecture using ontology-based recognition process. Semantic information representation of the contextual information provides the formal specifications and is most appropriate method of context modelling [Khattak2011, Chen2012]. Ontologies provide the base for design of the contextual information representation. Ontology-based models of context information exploit the representation and reasoning power of logics [Ertel2011] for multiple purposes.
- M. Hussain [Hussain2013] has used an ontological based knowledge repository to persist and manage different kind of data. The knowledge repository can easily evolve with arriving new activities of a person and social media information. This ontological model is able to collect different sensors and social media data in raw form and able to stores the real time data after recognition that data by activity recognition module.
- [Kofod-Petersen2005] divides knowledge into two types, domain and non-domain knowledge. Domain knowledge taxonomic structure consists of Environment information, Personal information, Social information, Task information, and Spatio-temporal information. The model enables the system to infer relationships between concepts by constructing context-dependent paths between them.
- As ontological model, importing ontologies that defined by others, into a system is a normal and mostly happening task. Usually a system needs only some specific module of the importing ontology but there is no way in OWL to import that specific module instead of whole ontology. In each model a privacy model is needed to prevent different context information like personal information, location information and health information by little dissemination.

2.6.3. Components Description

- Input: Curated Data (Structured, Unstructured, Semi structured)
- Output: Validated Data Model (Ontology Schema)
- Description: Data Representation and Mapping contains three major components Representation Model Selector, Data Mapper and Semantic Validation.
- Representation Model Selector selects an appropriate data representation model dynamically according to heterogeneous sources of data as well as the nature of the data.
- Data Mapper specifies and maps the classes, attributes, relationships and annotations from XML chunks of data provided by Data Curation and creates semantics among input data.
- Ontology Validator validates the semantics as well as syntax of the created information ontology.

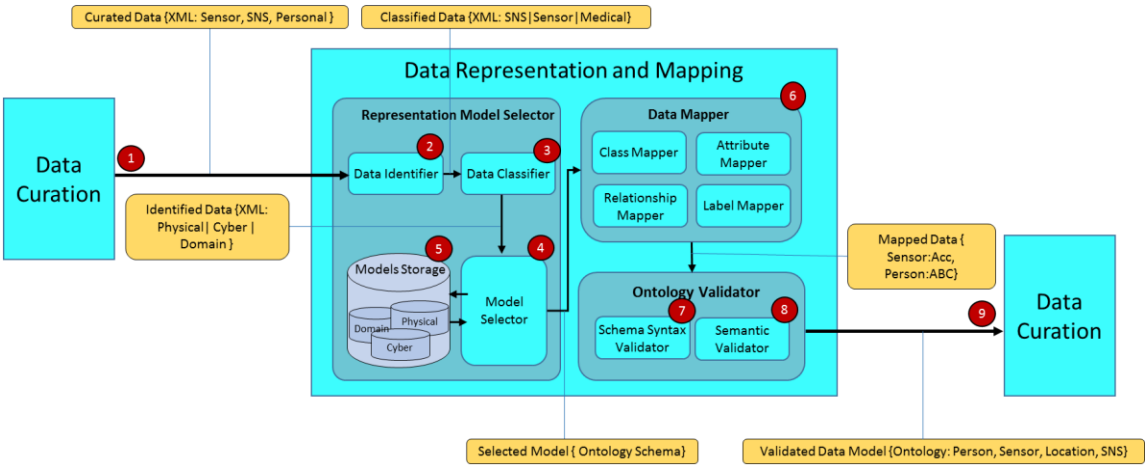


Figure 2.7 Architecture of Data Representation and mapping

Representation Model Selector:

- Input: Curated Data (XML/JSON)
- Output: Selected Representation Model (Ontology Schema)
- Description: This component identifies, classifies the curated data and selects appropriate representation model based on source and nature of data.
- Data Identifier sub component identifies the source of data by using characteristics based categorization.
- Data Classifier is using to classify and split the input XML/JSON data into sub categories based on nature of data like physical, cyber and domain data.
- Sub component of Model Selector dynamically selects the representation model schema for corresponding data category.

Data Mapper:

- Input: Selected Representation Model, Classified Data
- Output: Mapped Data (Classes, attributes, relationships with instances)
- Description: Extracts objects, attributes and relationships from input XML/JSON data as resources.
- Identifies and maps the classes for the extracted resources.
- Maps and finds the attributes, relationships and labels of the classes and puts the instances accordingly.

Ontology Validator:

- Input: Mapped Data (Classes, attributes, relationships with instances)
- Output: Validated Data Model (Ontology)
- Description: This component validates the ontology syntactically as well as semantically.
- The OWL ontologies are using three types of syntax RDF, RDFS and OWL, this component validates these three types of syntax.
- The semantic validator also validates the semantics among the input data. It checks the consistency, constraints and semantics.

2.6.4. UML Diagrams

- Sequence Diagram

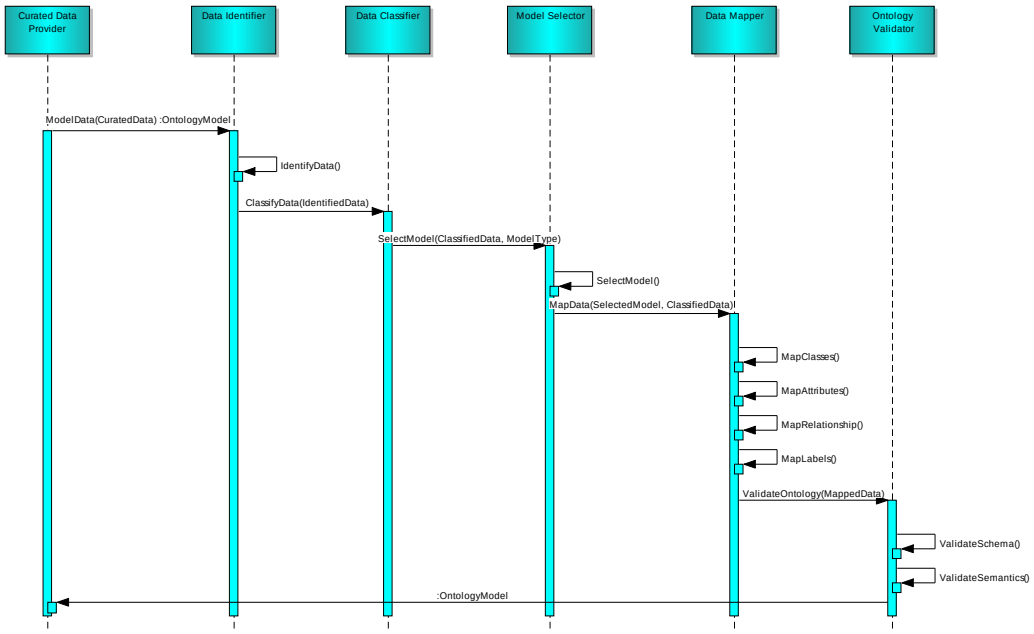


Figure 2.8 Use case diagram of Data Representation and Mapping

- Component Diagram

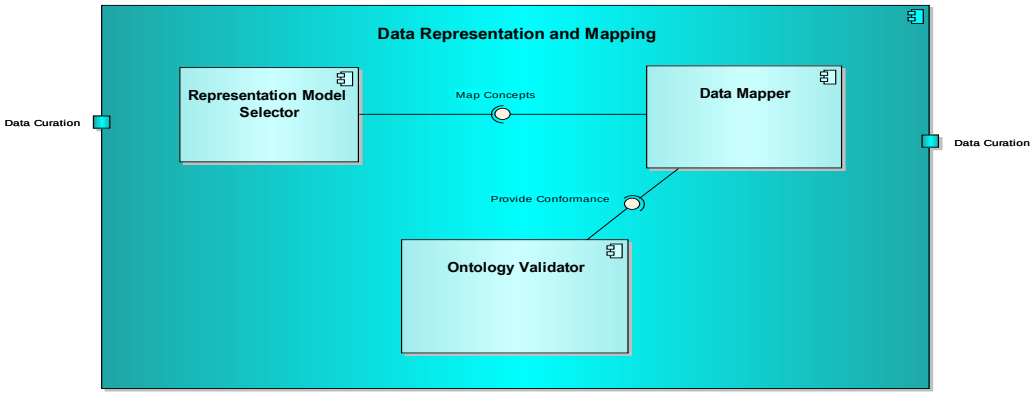


Figure 2.9: Component Diagram of Data Representation and Mapping

2.6.5. As Is – To Be Systems

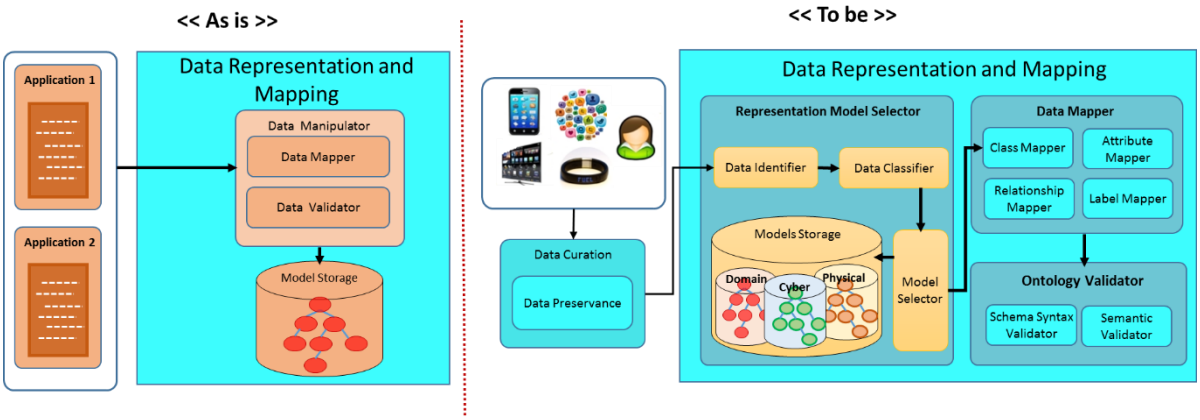


Figure 2.10: As Is-To Be of Data Representation and Mapping

2.6.6. Uniqueness

- Provides source-based categorization to identify the source of data for help in classification.
- Provides characteristic-based classification to help in model selection.
- Provides a dynamic model selection in multiple representation models to achieve loosely coupled model and Evolving the storage model easily.
- Provides unified format representation to validate and manage data of diverse resources and nature

2.6.7. Summary

The Data Representation and Mapping module is providing the unified format to represent the heterogeneous data that collects from diverse sources. It provides conformance of data for further processing and validates the syntax as well as the semantics of the collected data.

2.7. Data Curation

2.7.1. Introduction

In the activity recognition field, the number of subjects is usually limited to a few individuals, due to the difficulty in involving many people for long recordings. In contrast, the largest possible amount of diverse sensors should be used, since human activities can be observed through various channels (vision, sound, movement, etc.) and it is usually a fusion of different sources, which gives the best results. In complement to sensory data, due to comprehensiveness of activity dataset for mining minds we are involving data from SNS (Social Networking Systems also). Due to the diversity of type of data with its heterogeneous origins several steps are required to conform this data for further utilization.

As shown in the figure, Data curation has 3 primary goals to achieve:

- Data Labeling,
- Data Analysis,
- Data Preservance.

2.7.2. Related Work

- Data Curation so far only deals with persisted and stale data for intermediate data processing or service enabling. The opportunity of data curation over realtime data has largely been missed.
- The nature of data that has been under the investigation is so far homogenous i.e., data either belongs to same type or same scope. However, due to the nature of today's personal data, it is obtained from multiple sources and is of various types. Thus traditional data curtion is unfit for today's heterogeneous data.
- Moreover, data curation has also been focused over limited amount of data by type and volume. Thus our proposal of data curation over personal big data with heterogeneity is a novelty.
- Although most of the research under the umbrella of data curation is focused on data quality and provenance, a comprehensive data curation platform with labeling, analysis, and provenance is still an active research issue.
- Data curation has been actively done of biomedical and bioinformatics data, due to its scope homogeneity and its volume.
- This includes works over Protein Data (Research Collaboratory for Structural Bioinformatics, Rutgers University USA), Genetics Data (OpenHelix, University of California, Santa Cruz Genome Bioinformatics Group), and DNA Variations Data (DNAVaxDB, University of Michigan Medical School, Ann Arbor, MI).
- PDB Data Curation [Wang06]
    - Only support Protein Database files in XML-like formats.
    - No support for heterogeneous data sources and real-time data analysis.
- iRODS [Mark07]
    - Intended for Data Preservation
    - No Data Analysis and Labeling Support
    - No support for heterogeneous data sources and real-time data analysis.
- Data Curation for DNA Variation [Sophia11]
    - Only support Gene centered data and databases
    - No support for heterogeneous data sources and real-time data analysis.
- The Green Computing Observatory [Andrew09]
    - Build for collecting usage data of datacenters and clusters.
    - Built for datacenter data interfaces, no support for heterogeneity.
- Effective curation of cancer biomarker research data [Cecile11]
    - Build for U.S. National Cancer Institute's Early Detection Research Network (EDRN).
    - no support for heterogeneity

2.7.3. Components Description

- *Input*: Raw Data from Heterogeneous Data Sources
- *Output*: Labeled Data with context and source id
- In proposed resolution, data curation is evaluated over stream data and binary data over distributed big data platform (HDFS). Stream data is incoming data received at real-time from the heterogeneous sources; however, each source produces its data within its custom format.
- SNS data is usually received in XML or JSON format, which requires DOM-based interpretation.
- Similarly wearable and embedded sensors produce stream of binary data without context. Data labeling component provide such context to the data and tags the incoming stream with the event tag.
- This tag includes context information such as, location, user, timestamp, source, etc. This labeling process enables data fusion during data preservance component.

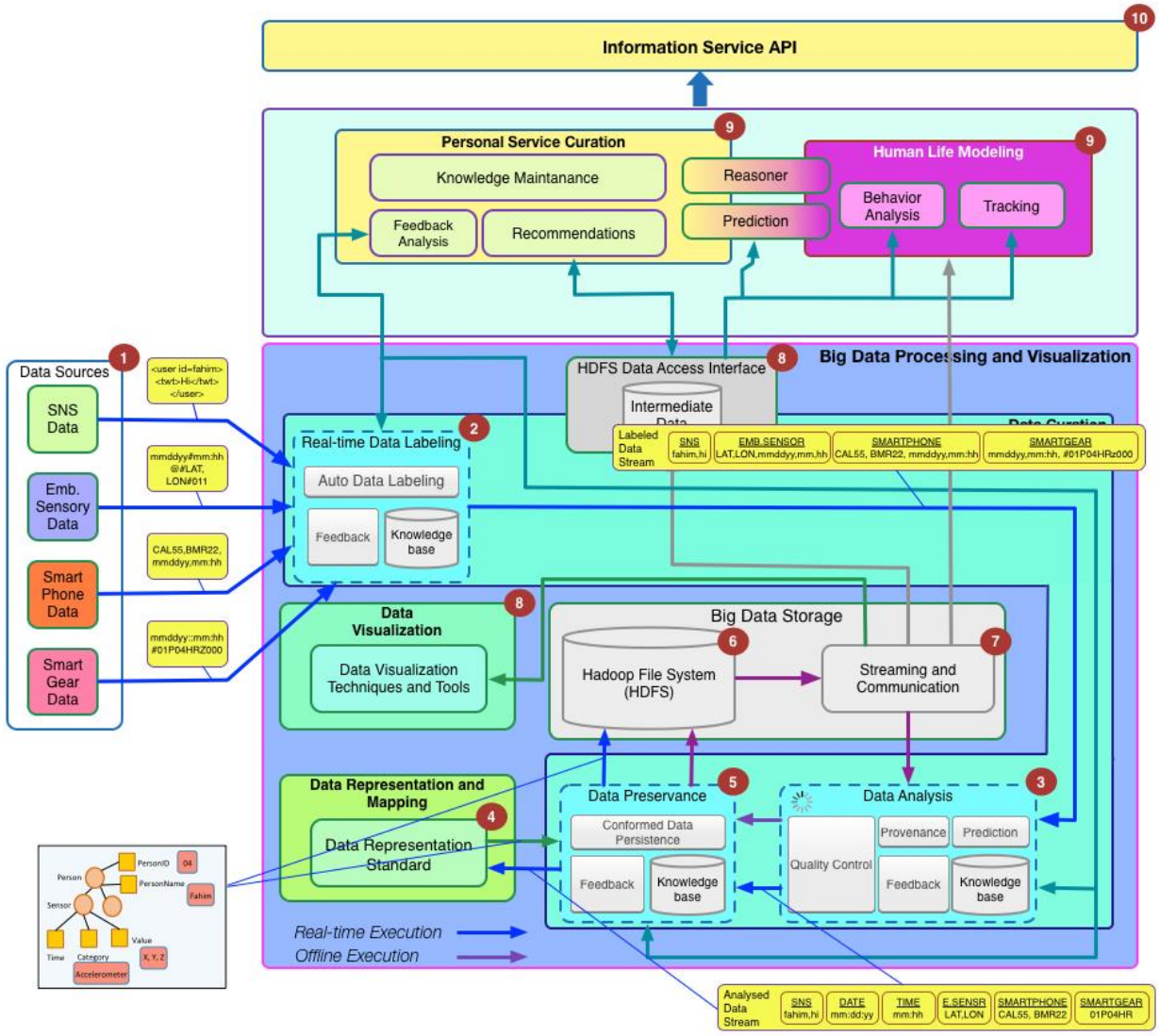


Figure 2.11: Data Curation Layer Architecture

Data Analysis

- *Input*: Labeled Data with context and source id
- *Output*: Analyzed Data with context and source id
- Data analysis is a cleansing process of real-time or streamed labeled data from heterogeneous data sources.
- Filtration process of data analysis is classified into two categories: Quality Control and Data Provenance.
    - Quality control as described in the figure constitutes upon 5 different filter operations. These operations are; Data Error Correction, Redundancy Resolution, Inconsistency Removal, Data Conflict Resolution, Obsolete Data Decommissioning.
    - Set of operations can be executed as a customized set for performance and precision trade-offs. Following table provides details regarding analysis filters.

Filter Name	Time of Executions	Condition	Resolution
Data Error Correction	Online / Offline	• Missing context info. • Label without data • Data format error	• Real-time stream data analysis • Context info with marked headers
Redundancy Resolution	Online / Offline	• Redundant data with multiple context info.	• Priority tables of data sources according to reliability
Inconsistency Removal	Offline	• Inconsistence data with same context info.	• Priority tables and data flagging • Manual amendments for complicated inconsistencies
Data Conflict Resolution	Online / Offline	• Data instances with mutual negation	• Checker module with Machine learning based fuzzy methods
Obsolete Data Decommissioning	Offline	• Stale data	• Removal of un-used an old data by request or time interval.

- As data comes from heterogeneous sources and generated by different methods vary in the degree of real reliability.
- Also persisted data has been transformed multiple times are more likely to have incorrect transformation or lose of context. In order to utilized used data, its evaluation for data reusability is of vital importance.
- In that respect data provenance plays its part with two aspects: Data Versioning and Uncertain Probability Evaluation.

Filter Name	Time of Executions	Condition	Resolution
Data Versioning	Offline	• Same context changed data over usage	• Priority table • Greedy algorithm for up-to-date data with more confidence
Uncertain Probability Evaluation	Online / Offline	• Erroneous data by un-autonomous update	• Measure the uncertainty of data by using fuzzy method

Data Preservance

- *Input*: Analyzed Data with context and source id as a model
- *Output*: Analyzed Data as a valid model
- For data to be used for visualization and more-understandable needs it has to conform to a standard.
- Data Preservance component takes the conformance ontology as standard scheme from Data Representation & Fusion, maps the labeled data from heterogeneous sources after analyses and transforms it into the standard ontology.
- With change in standard scheme, Data preservance can re-iterate over preserved data for transformation.

2.7.4. UML Diagrams

- Sequence Diagram

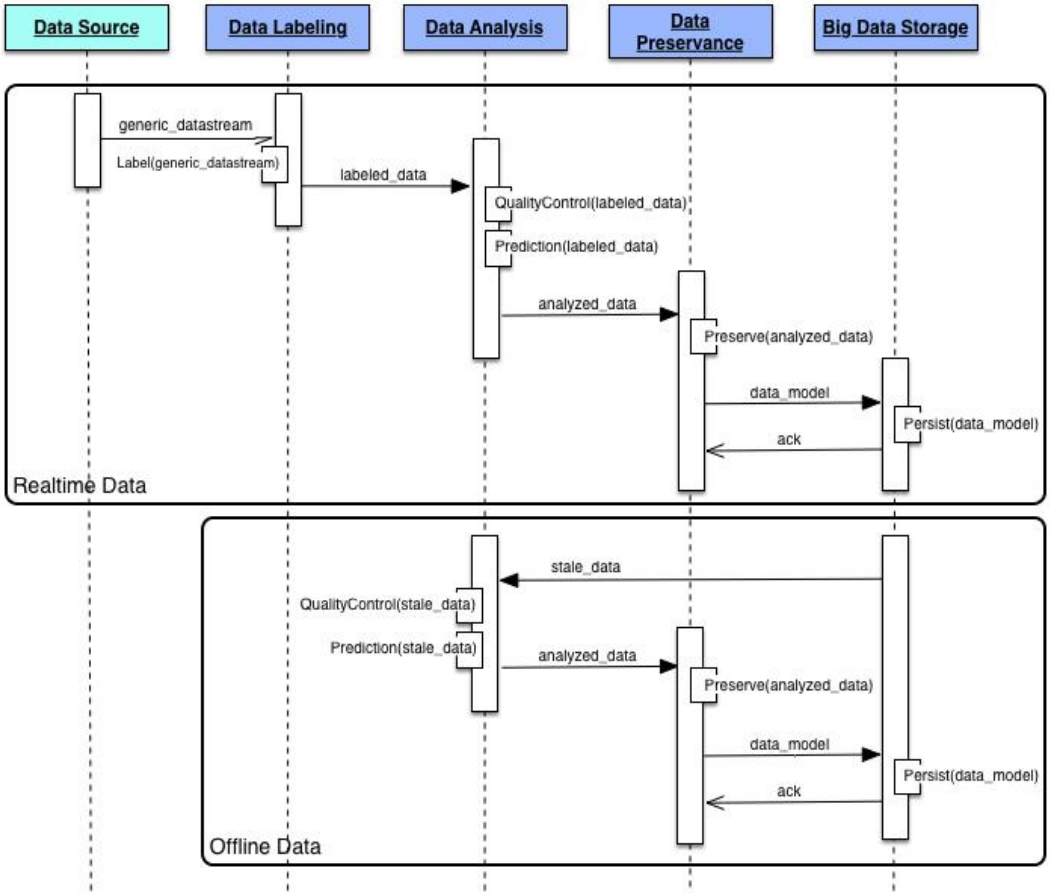


Figure 2.12: Sequence Diagram

- Component Diagram

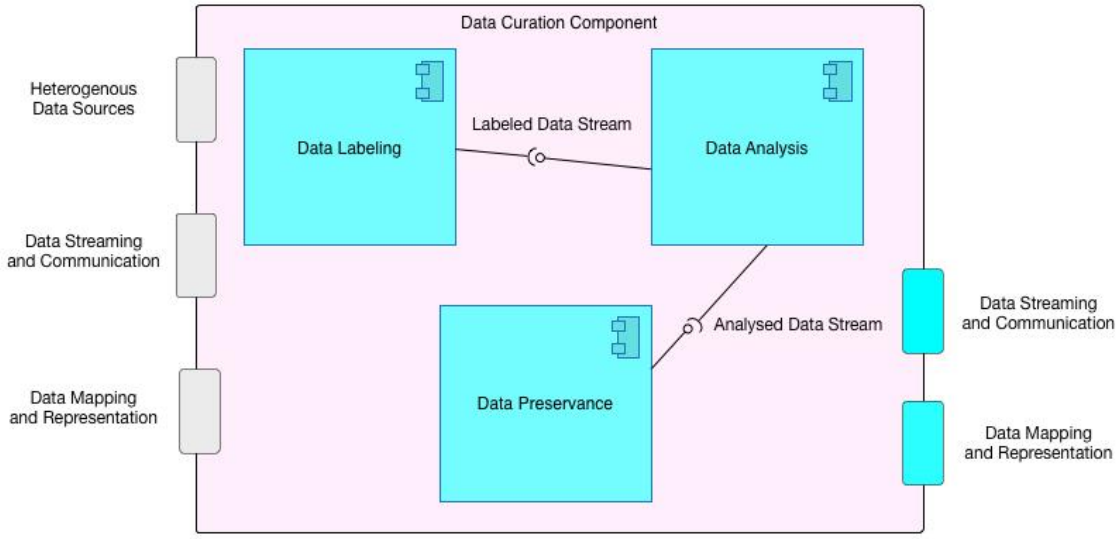


Figure 2.13: Component Diagram

2.7.5. As Is – To Be Systems

<< As is>>

<< To be>>

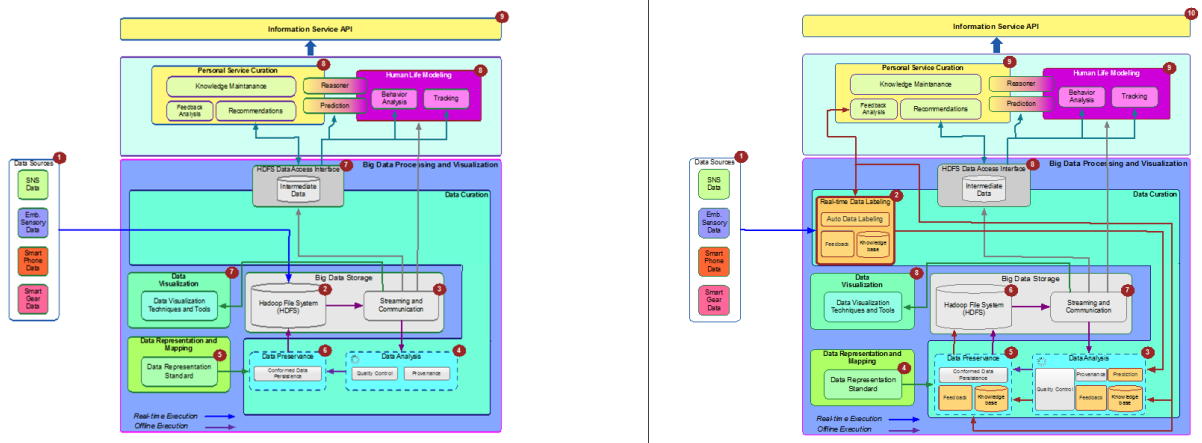


Figure 2.14: As IS - To Be

2.7.6. Uniqueness

- Comprehensive Curation
  - **Customized Data Analysis** filters for **offline and online labeled** data processing
  - **Data confidence by source priority tables** providing redundancy and heterogeneity resolution
  - Checker module for **data flagging** in case of inconsistency
  - Machine learning based checker module for data conflict resolution
- Performance
  - Improved analysis performance **decommissioning stale data**
- Conformance
  - Conformance to Standard **Ontologies for Data Preservation**

2.7.7. Summary

- Data Curation component is built to handle the real-time inflow of heterogeneous data, in larger volume with reliability.
- It enables mining mind system to predict on real-time data and curate it according to the data sources. This data is further streamed and stored in HDFS.

2.8. Visualization

2.8.1. Introduction

Big data is creating many opportunities for different and diverse fields to achieve deeper and faster insights that can enhance the decision making. These field are increasingly turning to visualization based tools to estimate 30 percent growth rate in 2015. We can provide interactive and easily understandable visual formats to improve the overall user experience.

2.8.2. Related Work

- A novel network traffic visualization scheme is proposed based on exploiting frequent item-set mining with the aim to visualize traffic patterns extracted from communication logs using hypergraphs [Glatz2014].

- The new emerging directions are highlighted to create clear, meaningful and integrated visualizations that give biological insight, without being overwhelmed by the intrinsic complexity of the data [Gehlenborg2010].
- A novel Massive Model Visualization via a rendering approach called MMDr using spatial hierarchies, with the goal of a better understanding, to achieve interactive frame rates on extremely large data sets, and help to accomplish more advanced research into Massive Model Visualization [Bennett2009].
- A tuning parameter taxonomy is proposed which consists of behavioral, Numerical statistically and temporal classifications and are further sub divided. This taxonomy help analyze different attributes and dimensions of the data being visualized [Dancy2008].
- An overview of current real-time massive model visualization technology, with the goal of providing readers with a high level understanding of the domain [Dietrich2007]. They have included simulations and different rasterization algorithms for rendering.
- A novel approach to automatic image annotation based on two key components: (a) an adaptive visual feature representation of image contents based on matching pursuit algorithms; and (b) an adaptive two-level segmentation method [Shi2004].
- For parameter tuning McAdams et al uses a subtle, but significant, change in the design: the addition of a tuning parameter in place of an increase in component precision. Statistical models are used to develop a framework for the tuning parameter design method in [McAdams2000].

2.8.3. Components Description

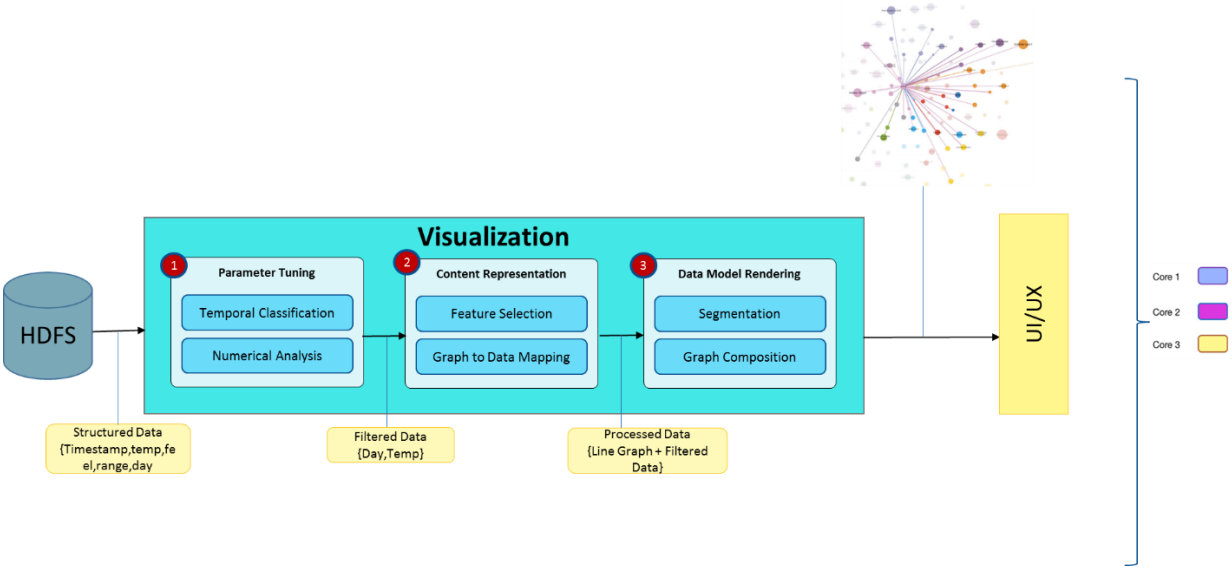


Figure 2.15: Architecture of Visualization

Parameter Tuning

- Input: Structured Data from big data storage.
- Output: Filtered data from Parameter tuning component
- Description: In Parameter Tuning, Numerical Analysis and Statistical Classification are performed. In Numerical Analysis, tuning will be done by considering capacity, counter, size and threshold parameters. While in Statistical Classification, statistical features like Minimum, Maximum, and Average features that are used to tune the parameters. This approach finds the ranks of features as high ranked features produces more accurate models.

Content Representation

- Input: Filtered data from Parameter tuning component.
- Output: Graph Selection according to filtered data
- Description: In Content Representation, quantitative representation and infographics techniques are used. The infographics are graphic visual representations of information, data or knowledge intended to present complex information quickly and clearly. They can improve understanding by utilizing graphics to enhance the human visual system's ability to see patterns and trends.

Data model rendering

- Input: Graph Selection according to filtered data and filtered data
- Output: Graphs for data curation, information curation and service curation
- Description: In data model rendering phase, processed input data is segmented to render the information and composition of related information are performed.

2.8.4. UML Diagrams

- Sequence Diagrams

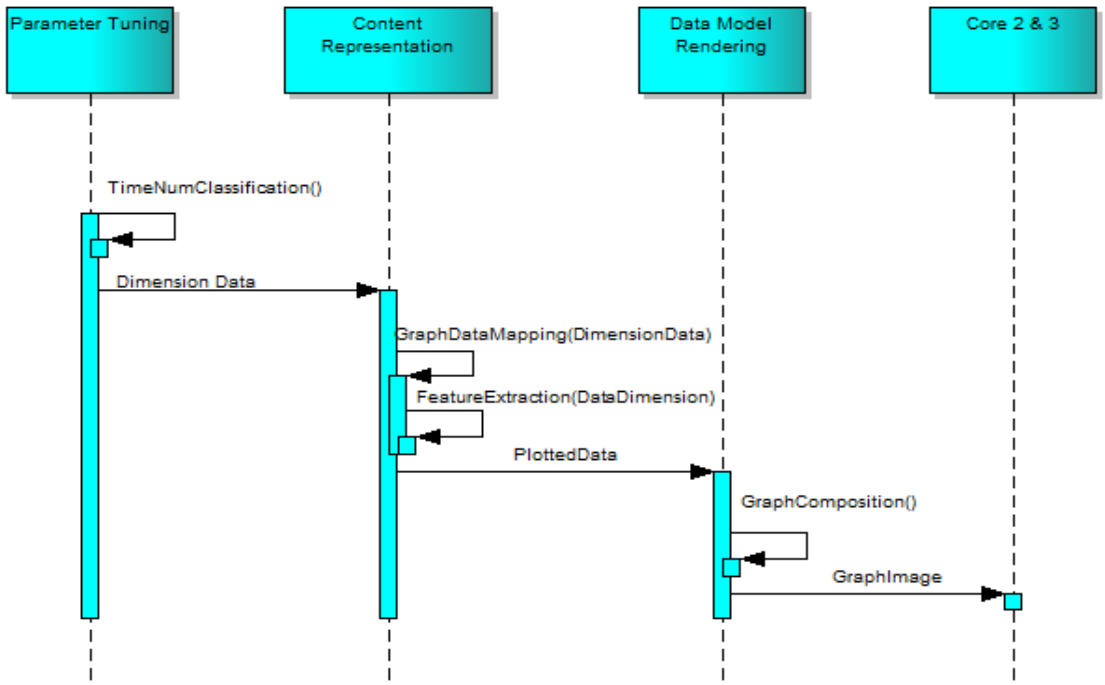


Figure 2.16: Sequence Diagram of Visualization

- Component Diagram

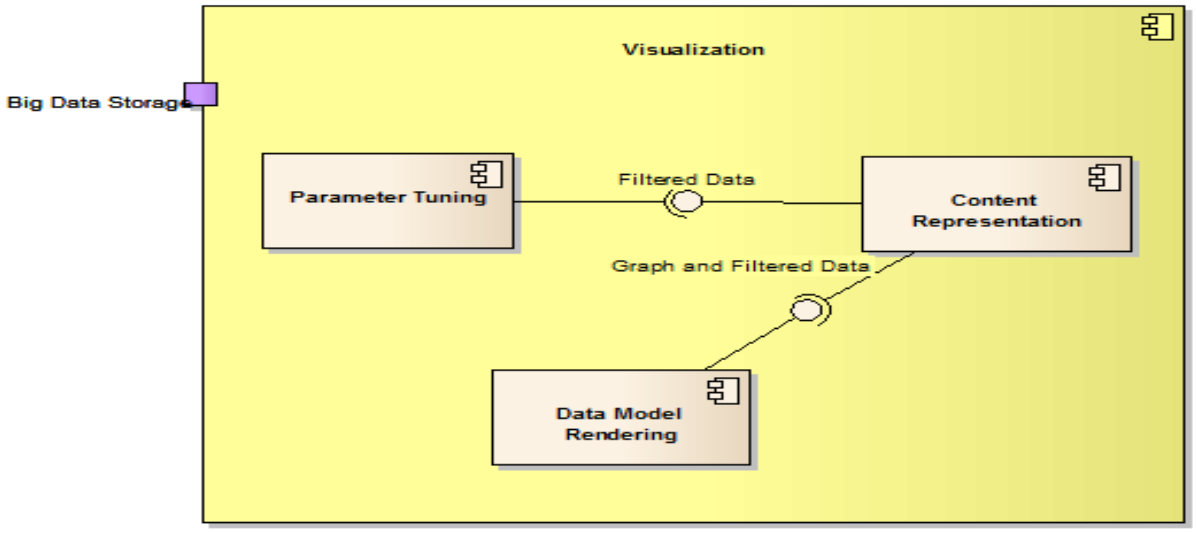


Figure 2.17: Component Diagram of Visualization

2.8.5. As Is – To Be Systems

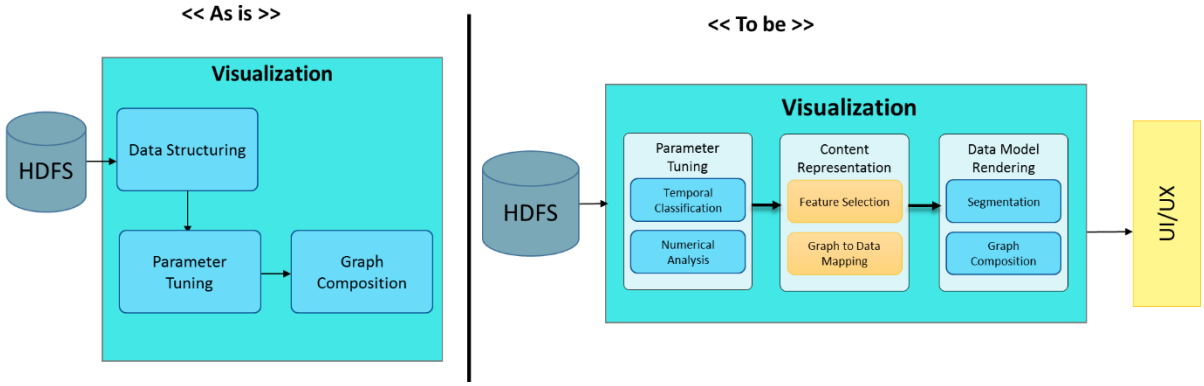


Figure 2.18: As Is – To Be

2.8.6. Uniqueness

We map the graph according to the features selected from the data depending on the temporal attributes. These attributes and graph features are mapped and the graph model is selected.

2.8.7. Summary

Big data analytics and visualization go hand in hand as we need an effective way to display the data intuitively for the users and the developers. Interactive data analysis, infographics and data mashups are the latest trends in displaying big data graphics in social networks. The parameters vary from social attributes, temporal attributes and location based. There are several trends that are occurring today that involve 3D visualization from a number of different perspectives. We will take the data from the big data and show visualization in different layers to different users which vary from service users to domain experts.

2.9. Security and Privacy

2.9.1. Introduction

Privacy and security are the inevitable components of any system dealing with sensitive contents. The sensitive contents mainly include personal information, data related with health or finance and the media files. The core idea behind Mining Mind is to collect data from multiple streams like sensors, wearable devices, cameras and smart devices like mobile phone and television. After collecting data from various streams the curation process is applied on that to discover hidden knowledge out of it. This knowledge which otherwise not possible without using the curation process, is then induced and shared back to its owner to assist in better and healthy living style. The orientation of this data is purely related with an individual; therefore concerns for protecting this data and information are very high.

2.9.2. Related Work

- For secure storage, public key encryption will be used and to protect further concerns of user query tracing or pattern discovery, paradigm of oblivious computing will be used. For oblivious computation, we have already developed oblivious term matching and oblivious access policies while dealing with the public cloud infrastructure [Zeeshan2012] [Zeeshan2013].
- While working with the encrypted storage, the data utility gets compromised and for this reason various searchable encryption schemes [Michel2005], [Mihir2007] have been proposed to deal in this situation. Besides using the oblivious term matching, searchable encryption are also helpful to deal with the encrypted storage yet utilizing the data utility at optimal level.
- Private matching is another way that is useful to protect the condition of any query until it is not satisfied. Recently, we have also proposed a light weight protocol for the private matching. The idea behind this protocol is to randomize the output result even repeated with same values again and again and that too without using the encryption. The motivation behind this approach is to mislead the intrusion or inference process with least instrumentation. If this approach is further wrapped with encryption, the overall system will become more deterrent against the malicious activity.
- With completion of curation and inferred knowledge, it is time to deliver it with the intended user. This step can be done by using the public key cryptography where only authorized user would be able to recover the information. To further enhance knowledge delivery, the output can be tied with some dependency of biometric or physical devices like smart phone or geographical area.
- The idea behind Mining Mind is to refine and alleviate the life style of humans especially the old aged society. If the information is useful for an individual, the same information can be shared as collective knowledge with the research community. For this purpose anonymization [Roberto2005]. [Ashwin2007] can be used that help to preserve the individual identity. At further level of refined access, the secure function [Louis2006] evaluation or private matching [Mahmood2013] are yet other available options.

2.9.3. Components Description

- Data Encryption: Encrypted data storage is a strongly recommended while working with the sensitive data. Therefore, encryption and decryption are relatively frequent operations that will take place within the system data flow. Whenever the information is required to be stored in HDFS or it is required for processing, it is encrypted or decrypted respectively.
- Process Authorization: This component serves in two dimensions. For any processing request that invokes this component, either the data is first decrypted and then reply is sent back or it can

facilitate the processing directly on the encrypted data. Provisioning of later feature makes the application more resistive against additional information leakage.

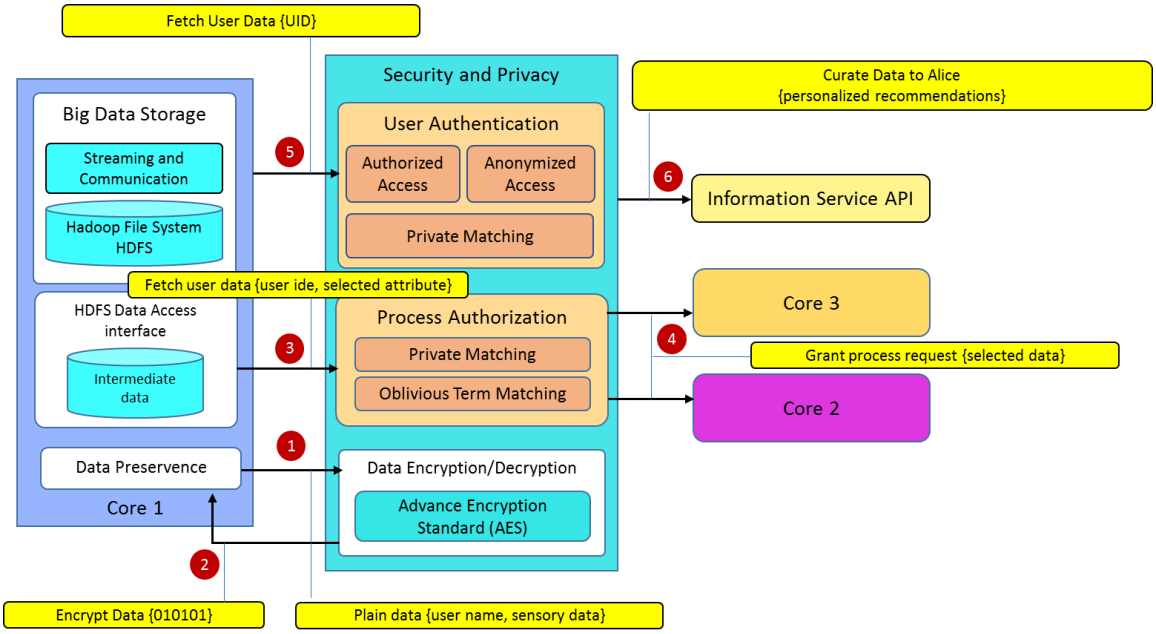


Figure 2.19: Architecture for Security and Privacy in Mining Minds

- Oblivious Term Matching: In this technique, a query is obliviously evaluated in the untrusted domain of cloud computing without disclosing the query word (selection criteria) or the reply. This technique is indigenously developed.
- Searchable Encryption: Searchable encryption (SE) is a standard that is used to work directly with the encrypted data without decrypting it.
- Private Matching: Although there are various standards available for private matching (PM), but for this application we have an additional advantage of our indigenously developed technique. Mainly, PM, is used to find the intersection of two datasets and outputs the common values. Provisioning of this component facilitates such scenarios where a requested resource(s) is considered as a datasets with available dataset stored in the repository.
- Authorized access: The curated services are not just valuable for its user but have higher concerns for its privacy and security. For better safety, this component ensures that only authorized users can have access to the mining minds output services.
- Anonymization: Other than authorized service for individual users, the output of mining minds can be shared for various research purposes. To facilitate this aspect the output cannot be served as it is however; with anonymized services the knowledge can be shared with subscribed users without disclosing the individual identity.
- Private Matching: This component mainly possesses the same characteristics as explained before. Here, its purpose is to further restrict public users or subscribed users with their queries only.

2.9.4. UML Diagrams

- Sequence diagram

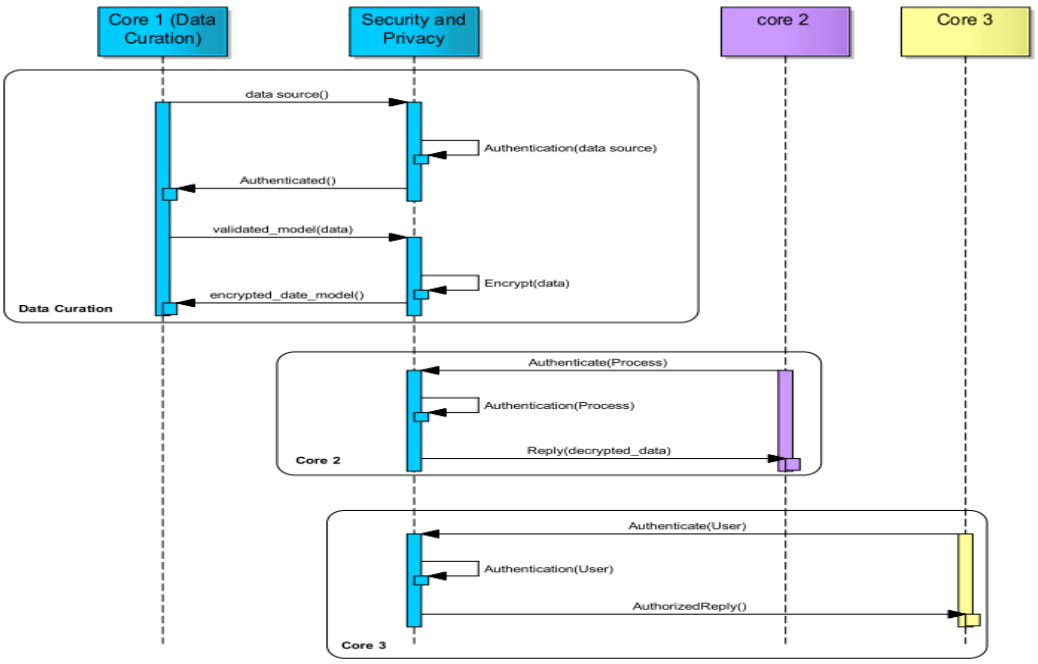


Figure 2.20: Sequence Diagram (Security and Privacy) for Mining Minds

- Component Diagram

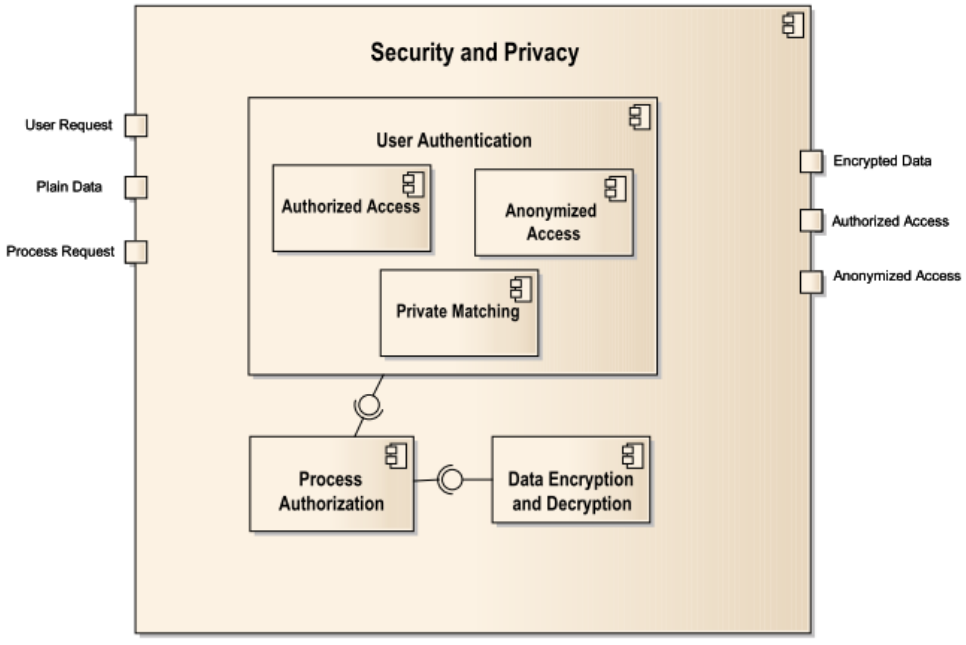


Figure 2.21: Component Diagram (Security and Privacy) for Mining Minds

2.9.5. As Is – To Be Systems

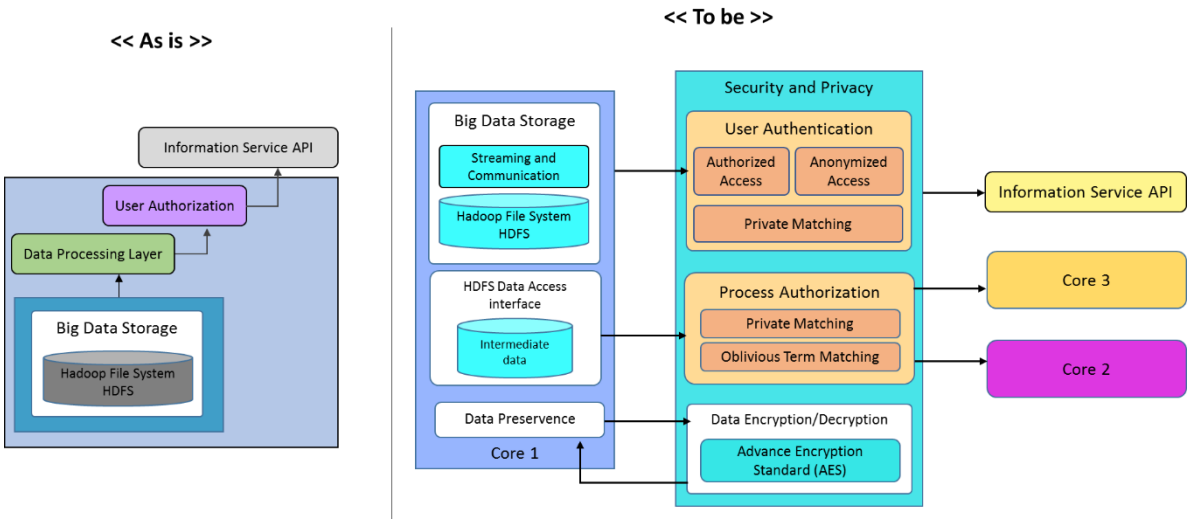


Figure 2.22: As Is – To Be

2.9.6. Uniqueness

- Mining Mind contains personalized trove of information and ensuring user privacy on that data is mandatory. While working in the untrusted domain of public cloud, encrypted outsourcing is first option, however; within high volume of information the processing speed might be slow due to frequent operations of encryption and decryption. To work for this situation, we have developed a high entropy solution with efficient execution time named "Reflection". It ensures high entropy to avoid side channel attacks yet preserving the underlying information of processed information.
- Working with encrypted data is a serious challenge and to meet this challenge we have our own design developed in this circumstances. It works obliviously while working with the encrypted storage by utilizing the cloud resources optimally.

2.9.7. Summary

Privacy has two fold aspect to safeguard it. First is directly related with data and second is related with its usage. This usage is again subdivided into two categories. The usage within the system and second is by the user. While availing the services of cloud infrastructure it is desired that a cloud service provider must not lean anything else beyond expected. The expected is the user request and its reply. Within cloud infrastructure the output result can be made independent from the fact that either the quires have been met with the given condition or not. Considering the second perspective with respect to user, her identity as well as access pattern must remain hidden as an additional leakage of information.

For secure storage, public key encryption will be used and to protect further concerns of user query tracing or pattern discovery, paradigm of oblivious computing will be used. For oblivious computation, we have already developed oblivious term matching and oblivious access policies while dealing with the public cloud infrastructure

2.10. Overall Data Curation Layer Uniqueness

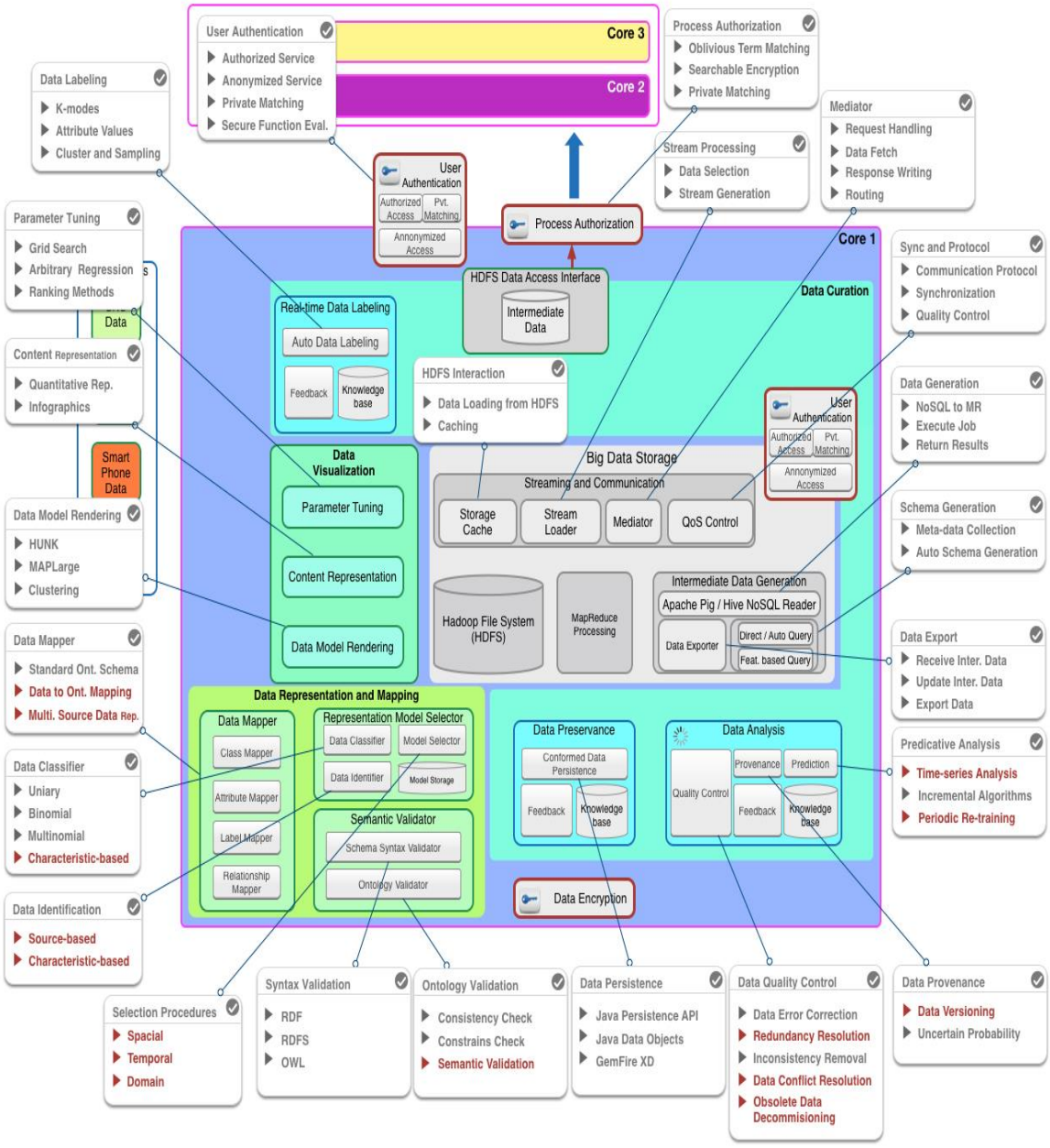


Figure 2.23: Data Curation Layer Uniqueness

Chapter 3
Information Curation Layer

3. Introduction to Information Curation Layer

The information curation layer curates the information to extract the low-level and high-level information. Furthermore, it also analyze the long-term and short-term user behavior to provide quality of services. The details of the key methods and our unique component (i.e. highlighted red) are mention in the following figure and details are provided in the sub-sequent sections.

3.1. Information Curation Layer Functional View

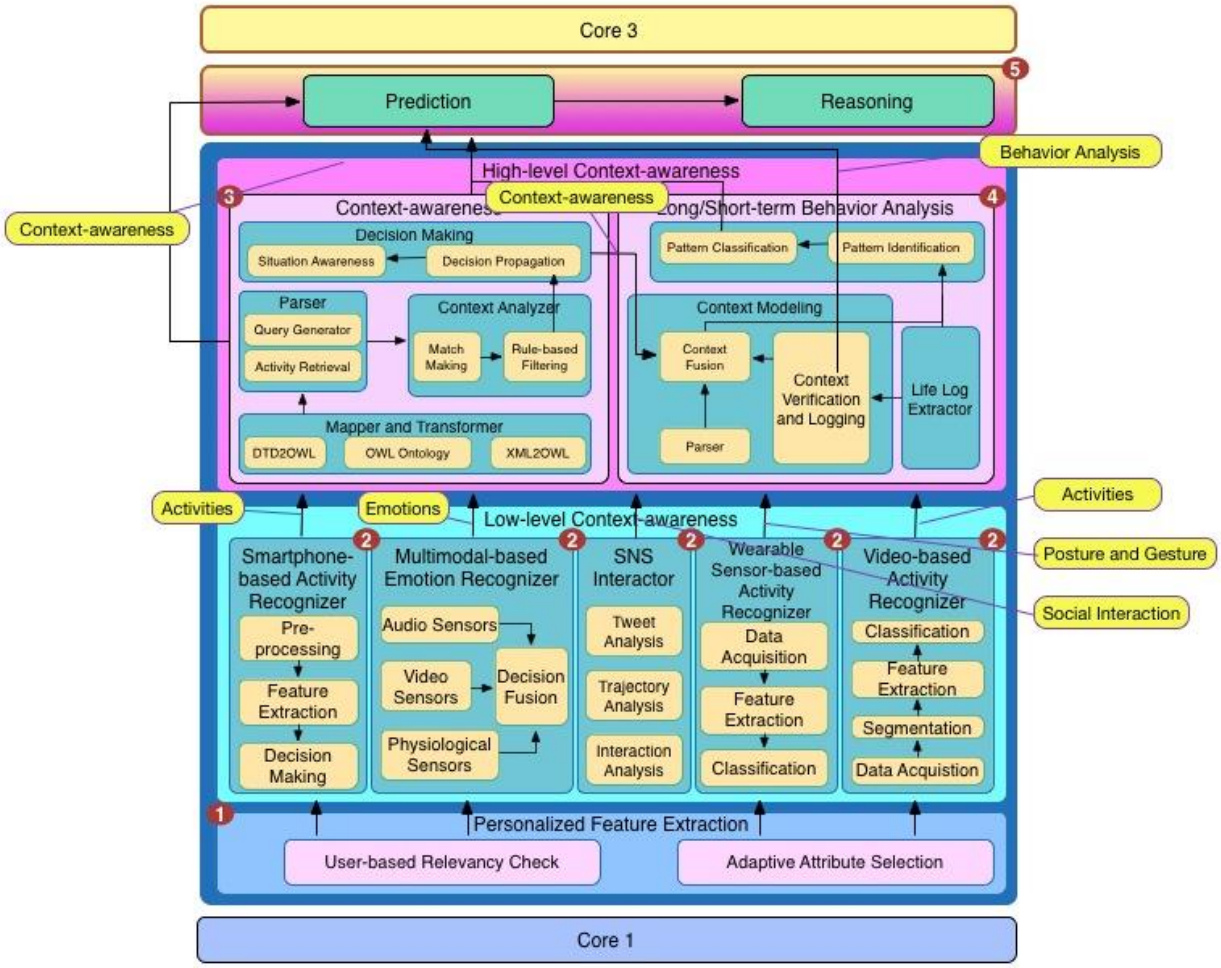


Figure 3.1: Information Curation Layer Functional View

3.2. Information Curation Layer Use Case Diagram

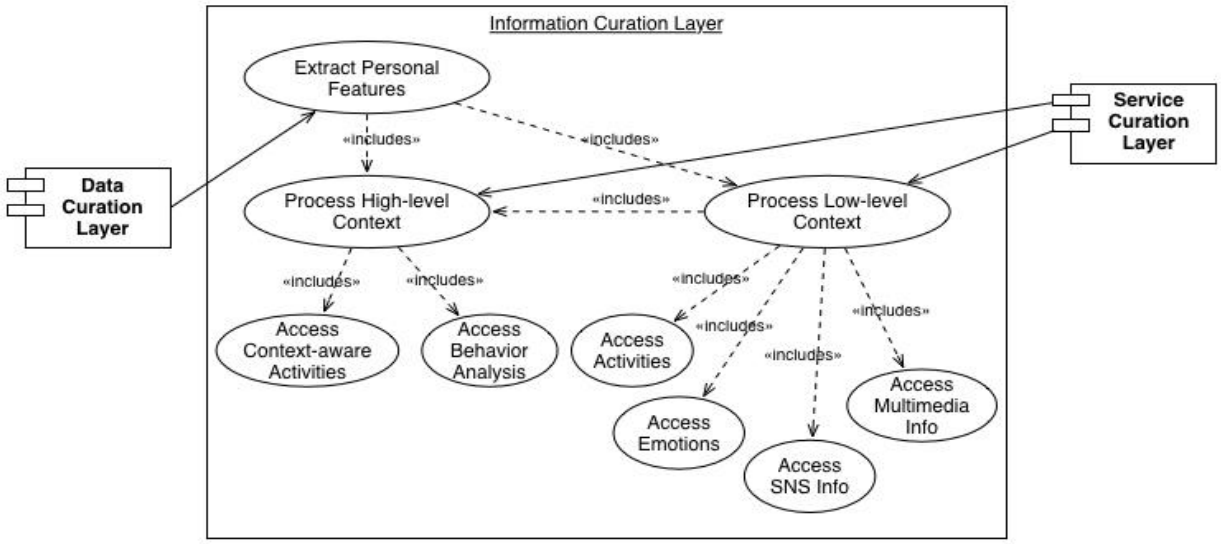


Figure 3.2: Information Curation Layer Use Case Diagram

3.3. Information Curation Layer Sequence Diagram

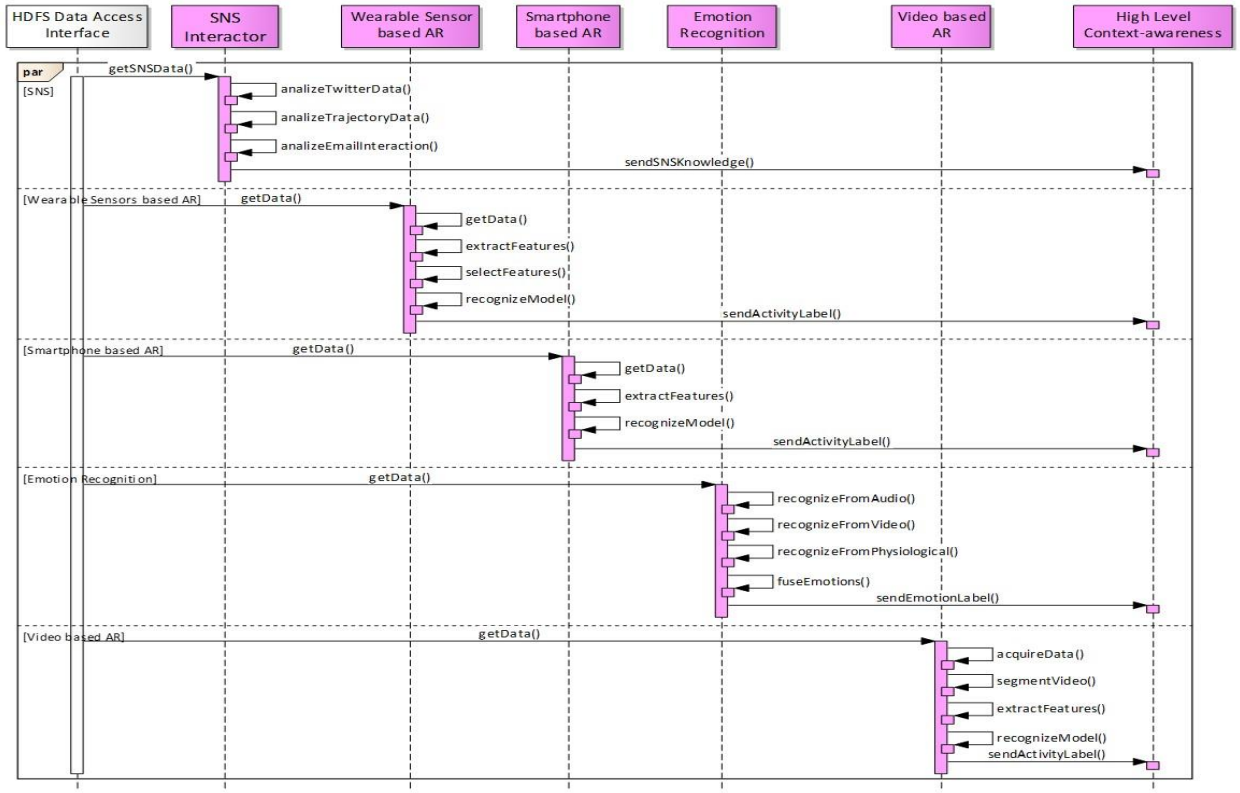


Figure 3.3: Information Curation Layer Sequence Diagram

3.4. Information Curation Layer Component Diagram

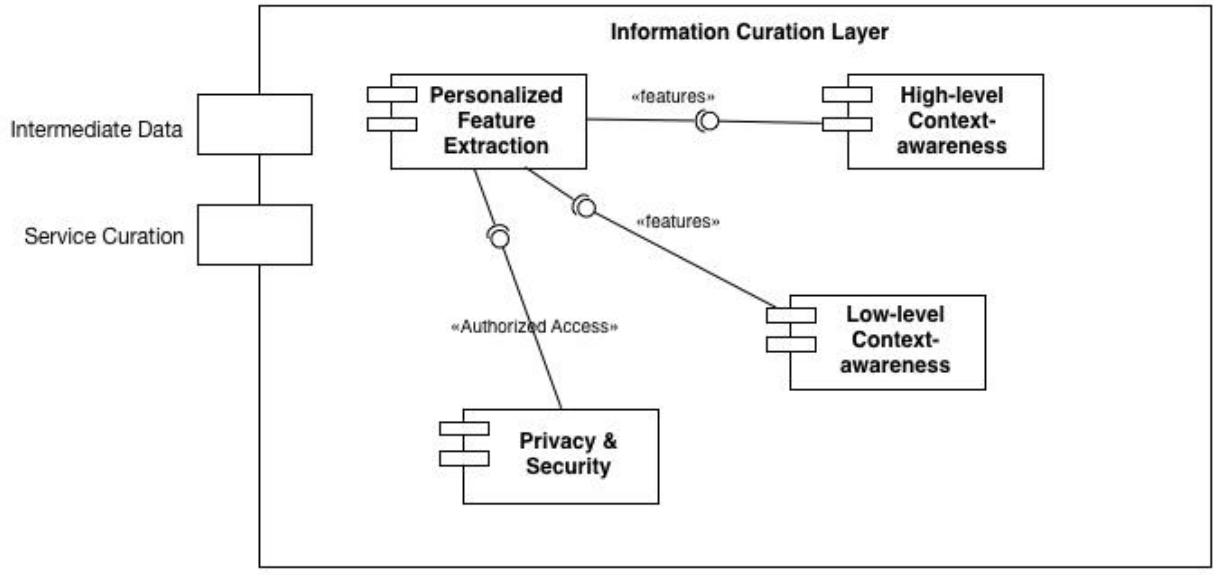


Figure 3.4: Information Curation Layer Component Diagram

3.5. Personalized Feature Extraction

3.5.1. Introduction

Personalized feature extraction module extract the required user's attribute from the intermediate data and apply relevancy check procedures and adaptive selection mechanism based on the request of the low-level and high-level context recognizer modules.

3.5.2. Components Description

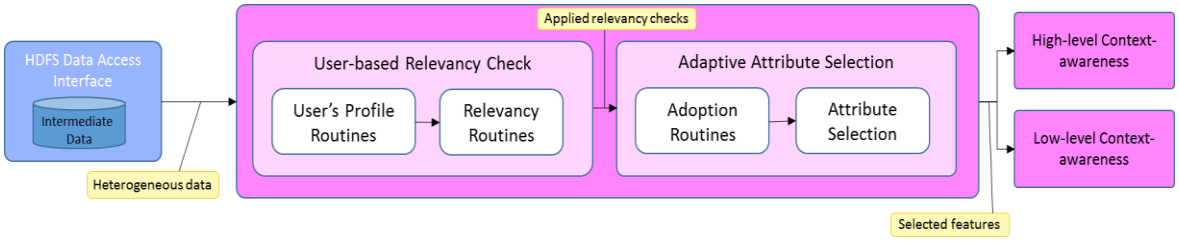


Figure 3.5: Personalized Feature Extraction

The personalized feature extraction module consists of the following two sub-components.

User-based Relevancy Check

- Input**:** Structured data

- Output**:** Relevant data for information curation
- Description: We grab the sensory data from heterogeneous sources and perform the relevancy check according to the low-level and high-level contexts.
  Adaptive Attribute Section
- Input**:** Structured data
- Output**:** Attribute selection for information curation
- Description: We select the attribute for particular process or method and passed to the required component.

3.5.3. UML Diagrams

- Sequence Diagram

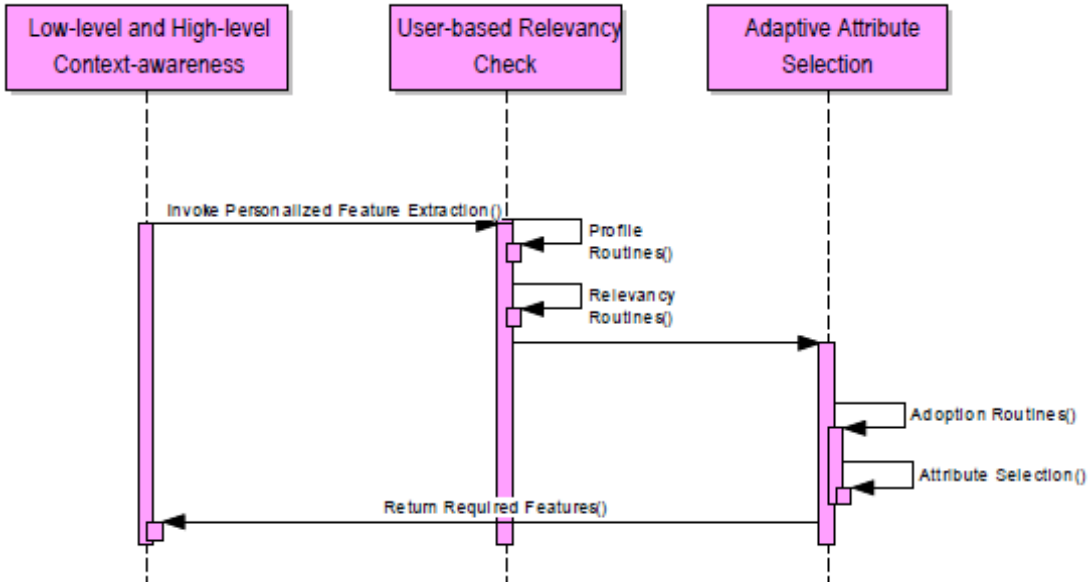


Figure 3.6: Sequence Diagram of Personalized Feature Extraction

- Component Diagram

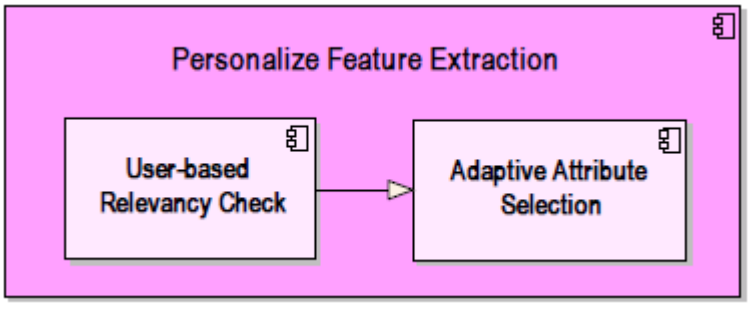


Figure 3.7: Component Diagram of Personalized Feature Extraction

3.5.4. Uniqueness

- In order to provide the required features from the curated data to the internal components of the information curation layer.
- We are providing the flexible mechanism to meet the requirement of the real-time information extraction and curation.

3.5.5. Summary

Personalized feature extraction module provides the required feature to the low-level and high-level context-awareness to convert the curated data to information. In this module we perform the relevancy check and adaptive attribute selection according the required components in this layer.

3.6. Low Level Context Awareness

3.6.1. Introduction

To provide rich information sources to life-care services, we can extract knowledge from different aspects that people are involving in daily life. With explosion of social networking and mobile technologies, we can acquire many different types of information of human life. In this report we consider to analyze social media interaction, activity recognition and emotion recognition.

We aim to improve the users' health by utilizing his social interaction in order to suggest them appropriate lifestyle patterns. For instance, after observing a user's daily routines, our proposed Social Media Interaction Engine (SMIE) is able to finds some complications with his lifestyle. The proposed SMIE is integrated through life-log that will take the information and integrate it with patient demographic to facilitate the behavioral analysis and suggest changes in unhealthy life patterns through better way. To achieve above goals, we design SMIE with several sub-components. Firstly, tweet analysis extracts user interest, health conditions and sentiment from user tweets. Secondly, trajectory in terms of outdoor movement of the patient is tracked using GPS enabled location aware mobile devices, such as smart phones. Finally, email interaction analyzes the users' actions to identify significant behavior and life threatening complications in daily routines to gain knowledge about their habits and preferences.

Physical activity recognizer method is introduced to solve the challenging problem of intrusiveness of sensor devices and server side processing. We presents a novel method for real-time physical activity recognition, position independent, and light weight classifier inside the smartphone environment. It is based on embedded accelerometer, ambient light and proximity sensors of the smartphone. Furthermore, our model introduces a device independent data acquisition method because data frequency is dependent on the sensor manufacturer and provided sensor manager API.

Emotion is a mental state that arises spontaneously. In daily life, emotion is not only an effective way to convey our intention in communication but also a good indicator of our mental health. That is the reason why automatic detection of human emotions is an important factor to enhance the quality of the service provided by the computer such as human-computer interaction. In this module, we are taking care of audio, video and physiological sensors to detect emotions of users. We also propose a fusion technique that is used to combine results from different sub-components and provide the final decision.

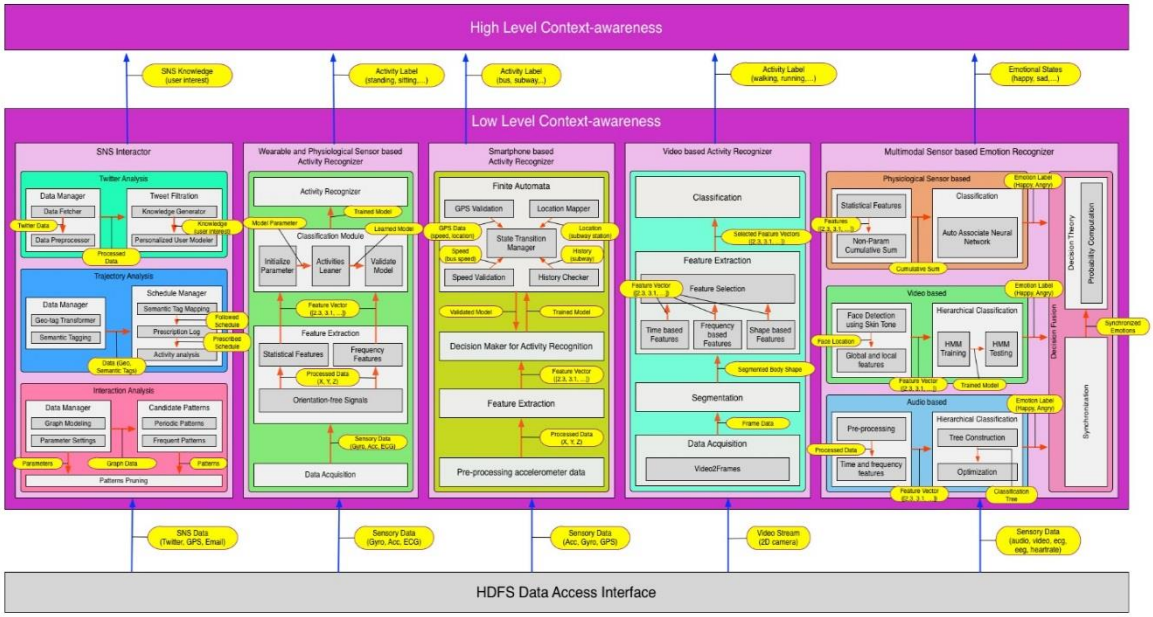


Figure 3.8: Architecture view of Low Level Context Awareness for Mining Mind

3.6.2. Related Work

- Social Media Interactor

Twitter Analysis

Much research work has been done to analyze the tweets, trajectory, interaction and other social media resources for different application domains [Abel2011-3] [Juyoung2010] [Christopher2003].

Chen et al. [Chen2010] analyzed URL recommendations on Twitter using data stream technique. System working is based on content sources, topic interest models and social voting to design URL recommender.

Celik et al. [Celik2011] studied semantic relationship between entities in Twitter to provide a medium where users can easily access relevant content for what they are interested in.

Trajectory Analysis

For trajectory analysis, mostly work is done for finding effective and efficient path tracking based on movement patterns. Yang et al. [Yang2009] used GPS for finding people preferences regarding attractive areas and movement patterns, which can lead to instructive insight to transport management, urban planning, and Location-Based Services (LBS).

Zhu et al. [Zhu2011] proposed Automatic Identification System (AIS) that uses trajectory mining techniques for finding the ship movement paths. Its purpose is self-navigation and collision avoidance.

Braga et al. [Braga2011] designed a trajectory based tracking system named 'Captain'. This system is designed for tracking of short, yacht trajectories. The focus of this system is to record the movement path of the person by using the parameters of the pictures, temperature, and coordinates of the locations.

Email Interaction Analysis

The focus on analysis of email network to identify importance of individuals on the basis of their communication patterns in network. The interaction analysis is also used to analyze the huge amount of data such as e-mail habits [Lahiri2010], mobile phone usage patterns [Amit2006], and dominance behavior [Hayley2011].

Christopher et al. [Christopher2003] analyzed the email contents to discover experts on particular topic. They proposed two approaches (a) content based approach consider emails text and (b) graph based approach that consider both text and communication network.

Yingjie et al. [Yingjie2010] examined email data by applying value patterns to cluster a social network. They applied statistical analyses, including hierarchical clustering, overlapping clustering, and correspondence analysis, to identify the value profiles of the employees.

Paweł et al. [Pawel2012] studied the email network to discover the importance of individuals according to their communication capacity. They find implicit ranking about the importance of users and by measuring the procrastination in answering of messages.

- Activity Recognition

  Wearable based Activity Recognition

  Tapia et al. [Tapia2007] coupled heart rate sensor with five accelerometers attached to the human body for recognizing the daily life activities. They analyzed the performance of Naïve Bayes Decision Tree (DT) and C4.5 tree family algorithm over the physical activities.

  Ayua et al. [Ayu2012] compared and evaluated different classifier algorithms for mobile's phone accelerometer-based activity recognition. They utilized the smartphone as an obtuse agent and their pre constrain is fixed location.

  In classification algorithms, nearest neighbor is one the most used algorithm in many applications domain and rated in the list of top ten best algorithms in data mining [Garcia 2012]. However, the nearest neighbor suffers from several drawbacks such as high storage requirement, low efficiency in classification response, and low noise tolerance [Xindong2007].

  Smartphone based Activity Recognition

  Bao et al. [Bao2004] worked with 2-axis accelerometer. They attached sensors to forearm, wrist, pelvis, knee, and calf fixed, and recognized walking, jogging, stay, etc. They calculated average, energy, frequency domain entropy, and correlation on each sensor, and classified with different classifier.

  There were some experiments using audio to recognize vehicles. Lee et al. [Lee2008] recognized environmental sound using significant feature vector automatically extracted when 3GPP2 Seletable Mode Vocoder is coded.

  Activity recognition research was still actively undergoing even after the emergence of smartphone. In [Wang2010], they used 3-axis accelerometer of Nokia's N82 mobile phone and recognized 6 activities, such as walking, subway, bus, car, cycling, and stay.

  [Han2012] tried to overcome the limitation of accelerometer based activity recognition, and used audio, GPS, wifi, etc. He used accelerometer which has big differences such as walking, running,

and stay, and used audio, GPS and wifi to recognize bus and subway which has small differences of signal.

Video based Activity Recognition

There lots of segmentation works such as graph-cut based segmentation by [Pourjam2013], and mean-shift algorithm by [Atefian2013] have been proposed for human body segmentation. However, graph cuts do not provide any uncertainty measure associated with the solution they produce. Similarly, in mean-shift algorithm, the window size (bandwidth selection) is not trivial; means inappropriate window size can cause modes to be merged.

Some recent well-known methods such as motion history image by [Takur2013], spatio-temporal interest points by [Everts2014], and dense motion trajectories by [Wang2011] have been proposed for feature extraction in activity recognition systems.

However, the scalability is one of the major limitations of motion history image-based methods because it analyze the lateral motion of the gesture.

Likewise, good segmented silhouettes are required for spatio-temporal interest points features and also these methods are very sensitive to viewpoint and occlusion.

Similarly, dense motion trajectories-based methods typically lost the underlying sequential information provided by the ordering of the words, when the activities are represented as bags of words

- Multimodal sensors based Emotion Recognition

  Audio based Emotion Recognition

  Prosodic speech features such as pitch and energy are extracted from each frame and called local features. On the other hand, global features are calculated as statistics of all speech features extracted from an utterance [Ayadi2011].

  Various types of classifiers have been used for the task of speech emotion recognition HMM, GMM, SVM, artificial neural networks (ANN), k-NN and many others. In fact, there has been no agreement on which classifier is the most suitable for emotion classification. It seems also that each classifier has its own advantages and limitations [Ayadi2011].

  Several emotion research works tried to separate the original complex multiple emotion classification problem by applying hierarchical approach with combination of different classifiers [Xiao2007, Lee2011, Mao2010].

  Video based Emotion Recognition

  Some recent methods such as feature-based methods [Azami2013], skin tone-based methods [Ghimire2013], and appearance-based methods [Beham2012] have been proposed for automatic human face detection.

  However, feature-based methods have trouble in automatic feature detection and prior knowledge should be required for these methods.

  Similarly, under certain lighting conditions, color is orientation invariant; therefore, skin tone-based methods have difficulties in such environments.

Likewise, the performance of appearance-based methods degrade with the environmental change.

Some well-known holistic methods such as local features [Mistry2013], eigenvectors [Kalita2013], and local binary pattern (LBP) [Ahsan2013] have been employed for facial feature extraction.

However, all the holistic methods do not know what exact facial features are the most important for facial expression recognition systems. Similarly, the performance of LBP degrades in non-monotonic illumination change, noise variation, change in pose, and expression conditions

Physiological sensor based Emotion Recognition

A number of prior approaches related to the present work can be mentioned. Kim et al. suggested the utilization of a support vector machine (SVM) trained on three physiological signals, i.e. ECG, skin conductance (SC) and temperature (T), to classify four emotions: sadness, anger, stress and surprise [Kim2001].

In [Nasoz2003] a study was conducted to detect sadness, amusement, fear, anger, frustration, and surprise using three physiological measurements, i.e. electrodermal activity (EDA), T, and heart rate (HR), k-nearest neighbor (KNN) and Discriminant Function Analysis (DFA).

Haag et al. on the other hand employed a neural network (NN) to estimate valence and arousal values from HR, blood volume pressure (BVP), electromyogram (EMG), SC, respiration amplitude and rate (RESP) data [Haag2004].

More recently Setz et al. used a fused classification combining Linear and Quadratic Discriminant Analysis to identify sadness, amusement, anger, contentment and the neutral emotional state. The physiological measures used on this occasion included ECG, the vertical component of the electrooculogram (EOG), EMG, the tonic and phasic element of EDA, RESP and T [Setz2009].

Leon et al. attempted to compensate for some of the real-time accuracy and reliability disadvantages of earlier methods by using sequential probability ratio test (SPRT) and auto-associative neural networks (AANNs) to provide real-time indication of emotional valence using information from HR, SC and BVP [Leon2007].

3.6.3. Components Description

- SNS Interactor

Twitter Analysis

- Input: Twitter data fetched by using Twitter API
- Output: Extracted knowledge such as user interest, health conditions
- Tweet analysis extracts user interest, health conditions and sentiment from user's tweets.
- Process tweets using natural language processing techniques with machine learning algorithms and returns entities, sentiments of the user about specific health condition to be used as knowledge for clinicians.

Trajectory Analysis

- Input: GPS location, prescribed schedule from practitioner
- Output: Recommendation such as daily exercise, timely medication
- Trajectory analysis in terms of outdoor movement of the patient is tracked using GPS enabled location aware devices, such as smart phones.

- Usually a patient is prescribed to follow a particular schedule from practitioner based on ailment e.g., it may contain suggestions of daily exercise, avoidance of alcohol, and timely medication.
- Trajectory analysis identifies the focused activities after considering imperative location and semantic tags.

Email Interaction Analysis

- Input: Email Interaction (sender, receiver, content)
- Output: Knowledge about habits and preferences
- Email interaction analyzes the patient's actions to identify significant behavior and communication trends in daily routines.
- It mines the frequent and periodic interaction patterns that change over time to gain knowledge about their habits and preferences.

➢ Activity Recognition

Wearable based Activity Recognition

- *Input:* Sensory data such as accelerometer, proximity, physiological information
- *Output:* Activity label (walking, running, sitting, standing, …)
- Current approaches require server side processing and predefined location of the devices that limits their applicability in real-world applications.
- To solve these problems, we proposed a novel PAR model to recognize the physical activities inside the smartphone environment.
- Our model utilized ambient light, proximity and accelerometer sensor to solve the issues of location and low classification of nearest neighbor algorithm.
- In addition, we introduce smartphone independent accelerometer sensing method that can be work well with android OS regardless of its manufacture sensor frequency rate.

Smartphone based Activity Recognition

- *Input:* Accelerometer, Gyroscope, GPS
- *Output:* Activity labels: walking, running, standing, bus, subway
- We recognize walking, jogging, stay, bus and subway independent to smartphone's position and orientation.
- Accelerometer data is revised with gyroscope data, and acquire fixed signal vector from front and rear, left and right, up and down independent to smartphone's orientation.
- Features is extracted from this signal to get similar signals which have identical activity.
- In addition, data is collected from different positions that we may carry on real life such as in top and bottom cloth, bag, hand, etc.

Video based Activity Recognition

- Input: Video stream from 2D or 3D camera
- Output: Activity labels: walking, running, hand waving …
- Video-based human activity recognition refers to an algorithm that a computer system uses to automatically recognize what human activity is being or was performed, given a sequence of video frames.

- The proposed human activity recognition system automatically segments the human bodies from the activity frames, extracts and selects the most informative and prominent features, and then accurately classifies the activities.

➢ Multimodal sensors based Emotion Recognition

Audio based Emotion Recognition

- *Input:* Audio stream
- *Output:* Emotion labels: happy, sad, normal, boring …
- Provided flexible engine for the task of emotion recognition from speech signal that can be applied easily in different system such as personal computer, smartphone, robot or home appliances.
- It can be combined with other emotion recognition method using different signal such as facial expression or physiological signal to provide a more general framework for classification emotion.
- Our engine can be extended more valuable and efficient by taking into account emotional speech recognition with independent language and in natural communication.

Video based Emotion Recognition

- *Input:* Video stream from 2D or 3D camera
- *Output:* Emotion labels: Happy, sad, angry …
- An accurate and robust facial expression recognition, whose accuracy is not effected by noise and by the race of gender of subjects in a given image.
- Our facial expression recognition system easily detects the human face from an expression frame, extracts the facial movement features, selects the most prominent features, and finally classifies the expressions accurately.

Physiological sensor based Emotion Recognition

- *Input:* Physiological sensor heart rate, blood pressure, ECG …
- *Output:* Emotion labels: happy, angry, sad …
- The present work is an initial step towards a real-time affective self-regulation system that combines affective and pervasive computing.
- It should be noted that although our main goal is to devise a system that is user-independent and that works across different age groups, there are users whose physiological or cognitive characteristics make emotion detection more challenging, i.e. elders or disable people.
- Along these lines, it should be noted that the results presented here have been done using data from young adults.
- It remains to be seen whether, the characteristics of our algorithm are preserved when data from a different population are used.

Decision Level Fusion

- *Input:* Different emotion labels from audio, video, physiological sensor module
- *Output:* Fused emotion label that has the largest confidence
- To make the final consensus, decision fusion method is introduced that harnesses the Bayesian decision theory due to its powerful mathematical description and capabilities to accurately handle the information delivered by heterogeneous sources.
- In proposed system, video and physiological sources have made the observations, and then roughly estimate the emotion state of the subject.

- For gaining the confidence about the emotional state, we refine our beliefs more precisely by multiplying the prior knowledge and maximal likelihood for a particular instance.

3.6.4. UML Diagrams

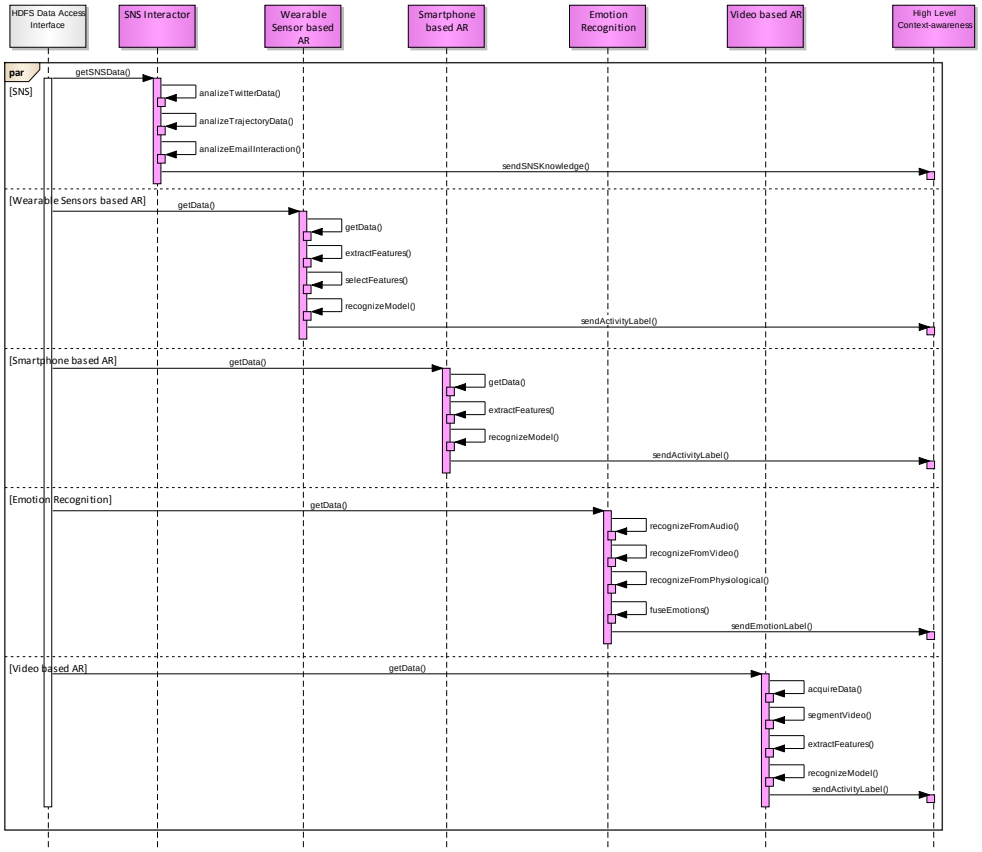


Figure 3.9: Sequence Diagram of Low Level Context Awareness for Mining Mind

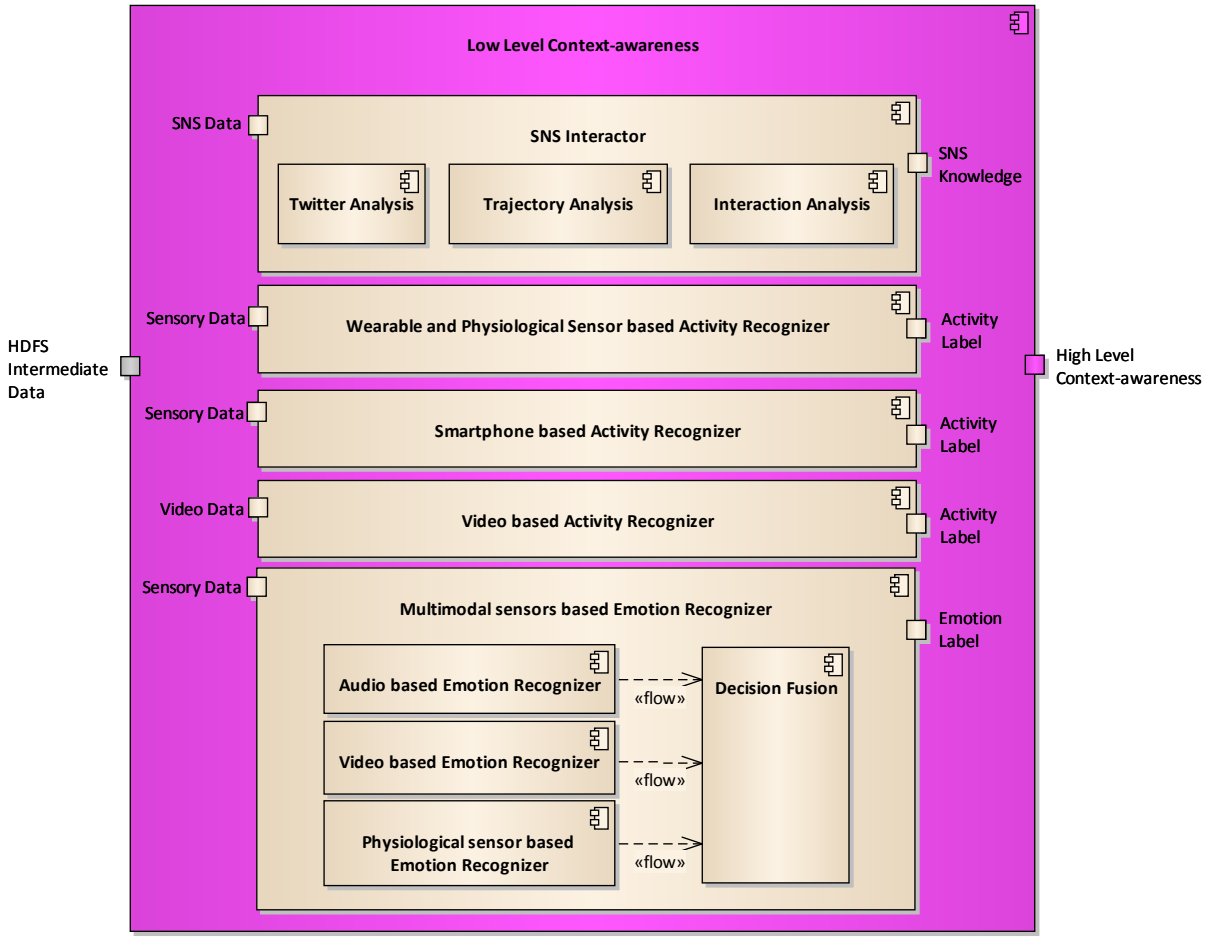


Figure 3.10: Component Diagram of Low Level Context Awareness for Mining Mind

3.6.5. As Is – To Be Systems

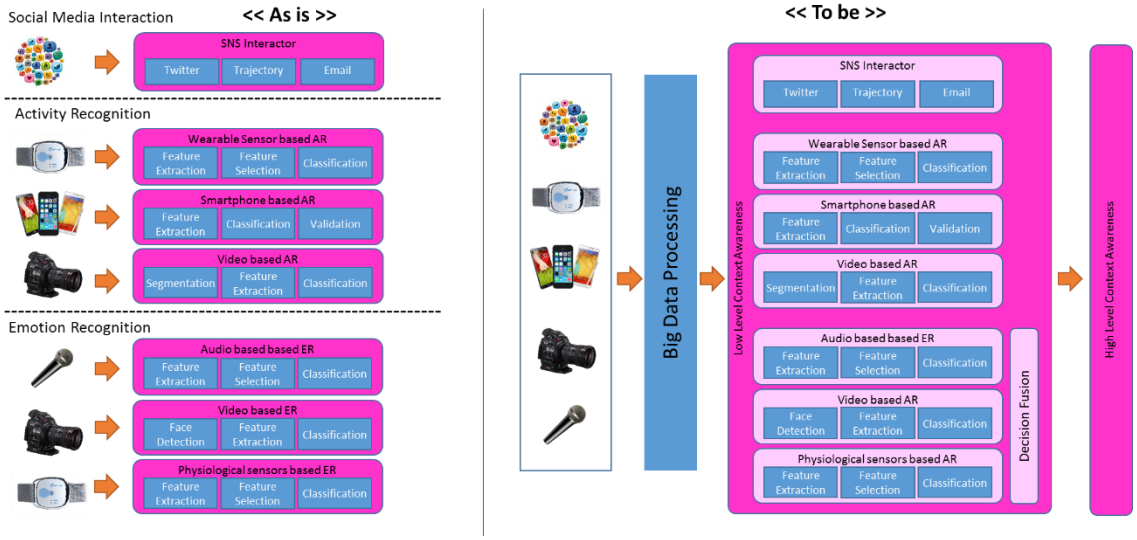


Figure 3.11: As Is – To Be

3.6.6. Uniqueness

- SNS Interactor

  Twitter Analysis

- Semantically processing of Natural language from Twitter to extract user interest and sentiments to build user profile. This profile would be used to provide personalized services.
- Classification of user interest into different categorizes to enhance system efficacy in providing domain specific services.

  Trajectory Analysis

- GPS enabled smart phone data with the availability of internet at each imperative location and fetch all the required information.
- Processed data over the Google API for conversion of GPS coordinates to Geo tags and further processing.

  Email Interaction Analysis

- In this module we mine the patients' frequent and periodic interaction patterns that change over time. The purpose is to gain knowledge about the preferences, needs and habits of the user.
- Users can act in two different roles: senders and receivers. These two roles are not interchangeable while mining the patterns of interest from his daily interaction routine.

- Activity Recognition

  Wearable based Activity Recognition

- No existing method of physical activity recognition that gives the maximum flexibility to carry the phone according to one's own choice either in pocket or hand.
- Processing sensory data inside the mobile environment to save the communication cost. Therefore it is an objective of the present recognizer to introduce a novel physical activity method to utilize the embedded sensors of smartphone.

  Smartphone based Activity Recognition

- We recognized bus, subway, and stay extracting unique vibration from fixed axis signal by accelerometer independent from smartphone's orientation. With this, we could recognize walking, jogging, stay, bus and subway by only using accelerometer with high accuracy.
- A revision algorithm is proposed to show high accuracy in real-field, and confirmed it by experiment.
- The usage of battery is minimized by only using accelerometer and GPS among all other sensors.

  Video based Activity Recognition

- A preprocessing method has been introduced which reduces the environmental parameters such as noise and lighting effects in order to smooth the activity frames.
- An automatic human body segmentation method has been presented to accurately segment the human bodies from the activity frames, and that is robust to illumination and clothing changes, typical issues in practical activity recognition systems.
- In the proposed activity recognition system, a robust feature extraction technique has been described that extracts and selects the informative features before fed it to a classifier.

- Multimodal sensors based Emotion Recognition

  Audio based Emotion Recognition

- Proposed hierarchical approach is flexible and efficient in term of classification emotional states from speech signal in different stages.
- Hierarchical binary classification tree is constructed automatically based on different criteria.
- The imbalance dataset is taken into account.
- Each classifier is optimized by feature selection and parameter settings.

Video based Emotion Recognition

- An unsupervised face detection and extraction technique has been introduced in order to accurately detect the faces.
- A facial movement feature extraction method and non-linear feature selection method have been proposed to select the most informative features.
- A hierarchical recognition scheme has been proposed in order to accurately classify the expressions.

Physiological sensor based Emotion Recognition

- NPCUSUM is a reliable mechanism to perform real-time detection of physiological changes associated with non-neutral emotional states.
- Although the results presented stem from simulated and not actual real-time detection, we have initiated experiments to perform online detection.
- The main advantage of this method is that a simple yet effective algorithm is employed which does not require any previous knowledge regarding the attributes of the signal being analyzed such as the power density spectrum or probability density function.

Decision Level Fusion

- The likelihood functions is exploited to fuse the data and maximize the confidence about the current emotional state.
- To prevent probabilities zero-frequency problem, we apply smoothing methods that discount the probability of the observed data and assign the extra probability mass to the unobserved events.

3.6.7. Summary

- We provide low level context awareness module that is able to provide useful information from heterogeneous data sources including of social networking, wearable sensors, physiological sensors and smartphone based sensors.
- Extracted information covers many aspects of daily life such as the behavior, activities and emotions of users.
- Output from this module is exploited to predict short-term and long-term behaviors.

3.7. High Level Context Awareness

3.7.1. Introduction

As the human lifespan increases, people are becoming more interested in living a healthy life, which results in high-cost healthcare systems and services. Maintaining good quality and widely available healthcare services at a minimal cost is challenging. Home healthcare systems are becoming a more important form of healthcare service delivery. The management, maintenance, and coordination of healthcare services, educating users, and empowerment of individuals to manage their own health are the main focus. To support this, a powerful, flexible, and cost-effective infrastructure is required for healthcare services that can fulfill the vision of ubiquitous healthcare (u-healthcare).

3.7.2. Related Work

- Over the last two decades, the concept of context awareness was first introduced in computer science domain. It is clear that the context awareness is considered as a very important initiative in ubiquitous computing.
- The trend of the early context aware systems like Cyberguide [Abowd1997] was limited to domain specific applications that considered only the location of the users like The Active Badge System [Want1992].
- While, later context aware systems like Gaia [Ranganathan2003] and CoCA [Ejigu2007] broadened their range by providing support to context aware applications in many domains, due to which they became able to process diverse types of context information received not only from sensors but also other external sources.
- However, Gaia does not focus on scalability and privacy protection. Similarly, the current ontology engines in CoCA system cannot process computationally intensive high level inferences.
- A recent context-aware system was introduced by [Yanwei2011] that came up with most of the existing systems limitations. However, this system was computational bound (i.e., time to complete a computation, systems keeps the user waiting), graphic bound (i.e., Limitations to displaying information), and storage bound (i.e., Limitations to the amount of data that can be stored).
- In some research projects such as [Guillemin2009] the context awareness has been identified as an important IoT research need by the Cluster of European Research Projects on the IoT (CERP-IoT) funded by the European Union. The EU has allocated a time frame for research and development into context-aware computing focused on the IoT to be carried out during 2015-2020.
- Similarly, some research institutes like Carnot Institute [Carnot2011] and Georgia Institute of Technology [Dey2000] are working on high level context-awareness in order to facilitate the human in their daily life routine.
- There lots of works have been done for high level context-awareness; however, most of them have their own limitations.
- Rule-base system [Nalepa2014] has been proposed for context-awareness. In their system, they have presented an overview of common approaches for modeling context, and we compared them according to certain critical factors.
- A fuzzy logic-based system [Saeedi2014] has been introduced that allow more natural representation, simple to define, easy to extend, less resource, (for example processing, storage) intensive, and can handle uncertainty.
- An unsupervised learning method [Banaee2013] has been proposed for context-aware system, for which no training data was required, and there were no need to know the possible outcomes.
- Likewise, a probabilistic model-based method [Tenorth2013] has been described for context-awareness, which allowed to combine the evidences, can handle the unseen situations and uncertainty, and provided meaningful results.
- An ontology-based models [Kayes2013, Attard2013] have been proposed for context-awareness, which allow complex reasoning and representation, have more meaningful results, also have validation and quality checking, and can reason both numerical and textual data.
- However, rule-based systems suffer from the difficulties originating from ambiguity, boundary conditions, and undefined models.
- Similarly, fuzzy logic-based methods should define manually (i.e., it is a kind of heuristic), and there is no validation or quality checking [Charith2013].
- Likewise, unsupervised learning models can be complex in hug data, are less meaningful because of having less semantic. Also these methods are difficult to validate.

- However, it is important for the probabilistic methods to know the probabilities for the numerical values.
- However, for ontology-based models, the data should be in a compatible format such as OWL or RDF, limited in numerical reasoning, and require more computations power and time [Charith2013].

3.7.3. Components Description

- Input: It takes input data from low level context-awareness and HDFS Data Interface modules
- Output: It provides high level contextual output to reasoner and prediction, recommendation manager, and short/long term behavior analysis modules for further processing.

➢ Mapper and Transformer

- It is a component that is responsible for recognizing data inputs from diverse sources. Video-based, sensor-based, motion-based, and location-based activity recognition engines will provide output in different formats like XML and simple text. So there is need of transformer in the engine to properly parse all type of output produced by these sources. All the data are converted into ontological format and create the ontology models for each data. Then it can extract the data from the file and provide these to the next module i.e. Context Analyzer.

➢ Context Analyzer

- In this module, the context of the data has been checked. Two kinds of methods have been utilized for data checking in context analyzer: Match matching, and rule-based checking. When the context analyzer received data from mapper and transformer, then the pattern has been checked either by match making algorithm or by rule-based filtering algorithm. If the data consists ontological data model, then the match matching algorithm requests to the query generator module of the parser which generates the query according to the needs of the user, and then retrieved the data with the specific matched pattern. While on other side, if the data consists of some rules, then another method such as rule-based filtering algorithm has been utilized in order to check the pattern of the received data by rules. The rule-based method requests to the rule-base repository via rules and checks the data based on rules. Once the pattern of the data has been matched, then the data is given to decision making in order to make the higher level decision based on the collected information.

➢ Parser

- For any type of information manipulation from the Knowledgebase, parser is responsible to properly handle all the operation regarding that matter. The parser normally communicates with activity representation to properly represent the activity, it also parse the knowledgebase for the inference engine for verity of different reasons like verification of activity and decision making, To populate the knowledgebase for newly recognized activity, the parser is also used in that case.

➢ Decision Making

- After the process of parsing and context analysis, the system can take decisions or give suggestion against different activities. So this module is responsible for performing some actions against the suggestions based on the collected contexts. This module also visualizes the activities and the knowledgebase for proper understanding of the activities. This module have two modules: situation analyzer, and decision propagation. The situation analyzer reports and monitors the contexts of the human in order to take higher level decision; while based on the situational analysis, the decision propagation module activates the devices, stored high level alerts. For example, if professor enters

a class room then the situation analyzer module using the information from the sensors and from the knowledgebase infers that its lecture time and it issue commands for turning off class room lights and turn on multimedia (projector) in class room.

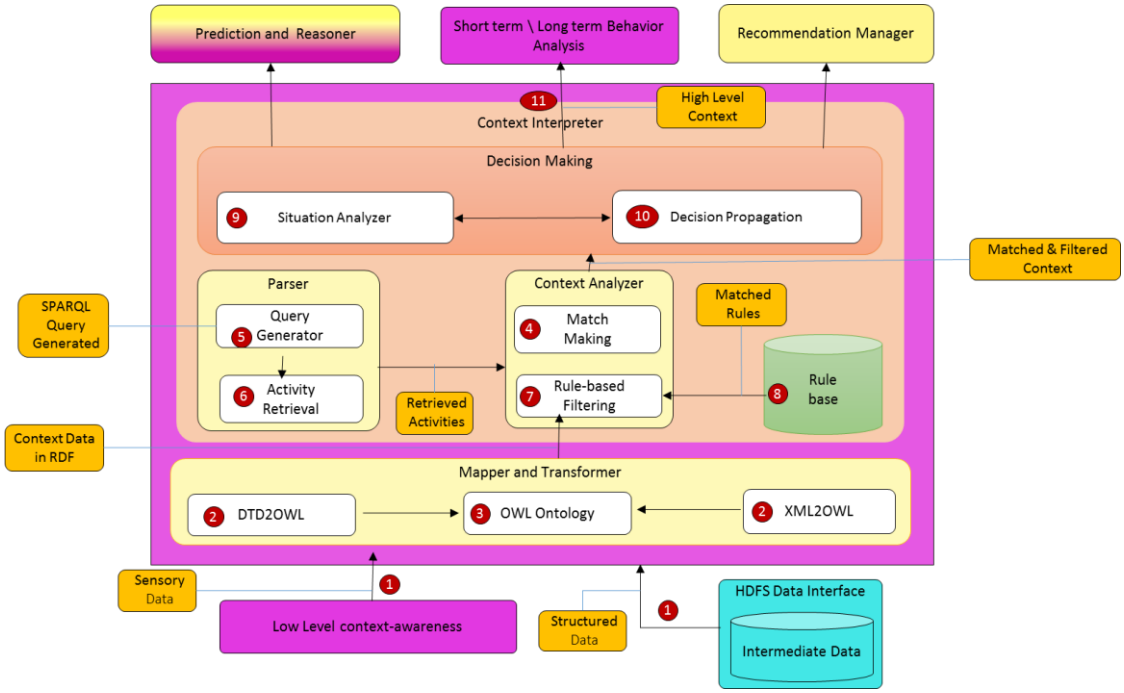


Figure 3.12: Architecture of High Level Context-awareness for Mining Mind.

3.7.4. UML Diagrams

- Sequence Diagram

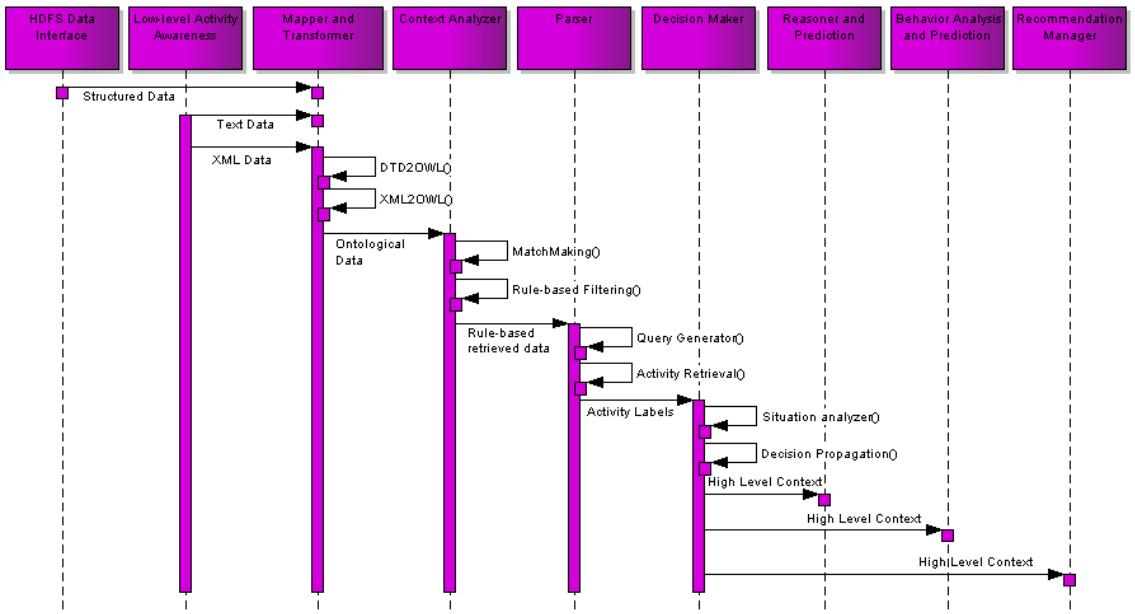


Figure 3.13: Sequence Diagram of High Level Context-awareness for Mining Mind.

- Component Diagram

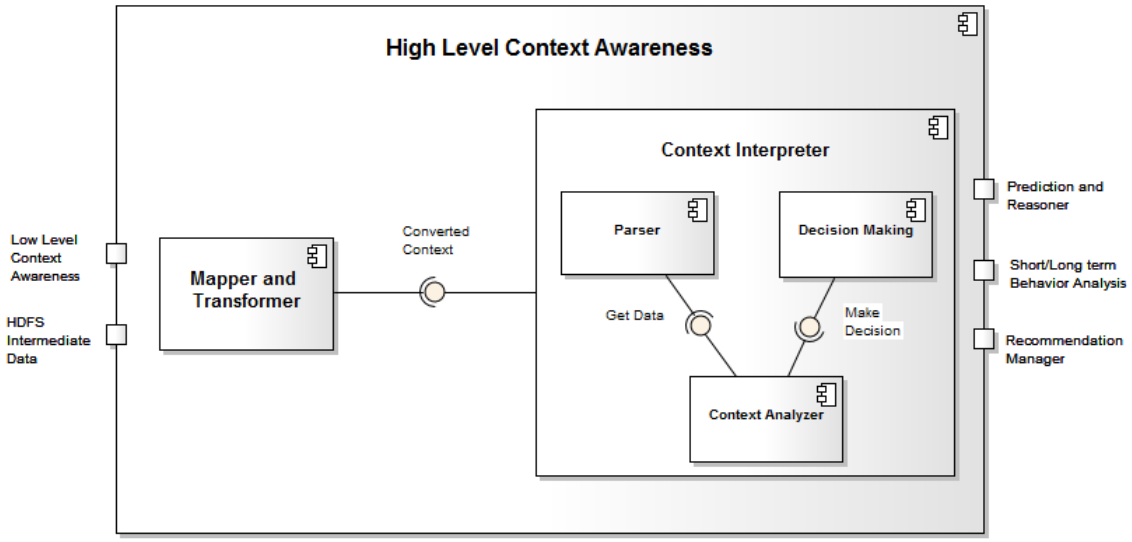


Figure 3.14: Component Diagram of High Level Context-awareness for Mining Mind.

3.7.5. As Is – To Be Systems

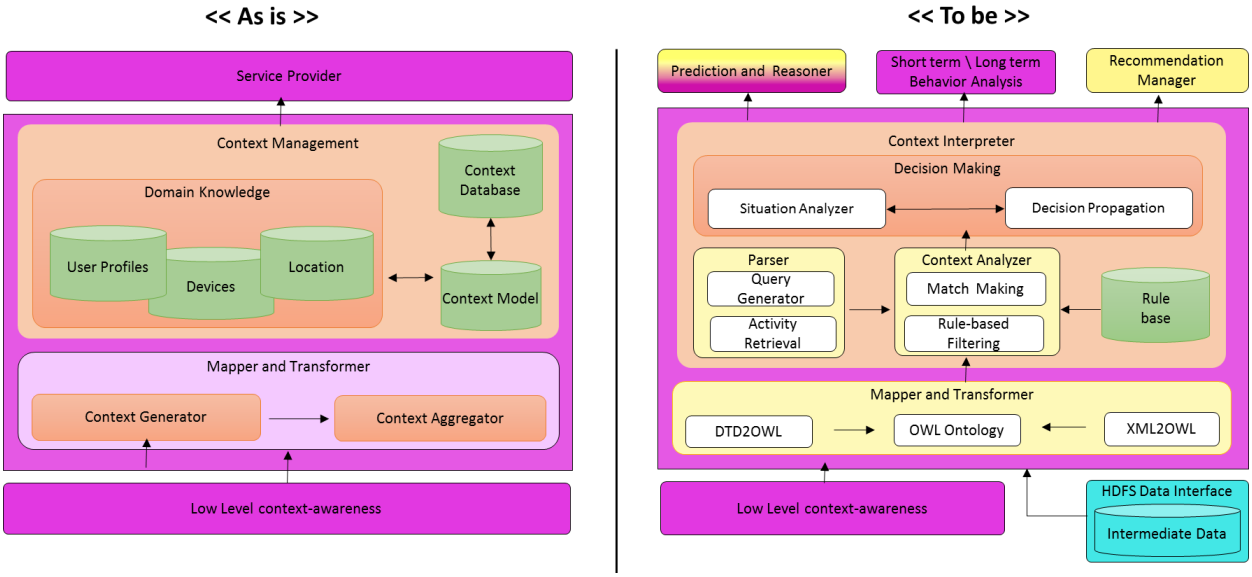


Figure 3.15: As Is – To Be

3.7.6. Uniqueness

- Using ontology for modeling the context, user profile information, and representation of information.
- Knowledge driven approach using ontology is used to make decisions after detail context analysis.
- Using ontology to infer higher level activities using forward chaining for detection of human behavior.
- Using profile information (Life style) of a person for making personalize decision.

- Two phase filtering procedure is used to make proper decisions and to minimize error rate.
- Match making using H-Match and Falcon. Description Logic Rules based on expert (Doctors) knowledge for decision making.

3.7.7. Summary

The high level context-awareness has been presented using a knowledge-driven approach to recognize Activities of Daily Livings (ADLs). The objective of this system is to infer high-level activities from low level real-time ADLs detected by sensors and to facilitate the provision of better healthcare services. The nature and characteristics of ADLs were analyzed. Based on the analyses, ontology was used to model the ADLs (including activity, location, time, profile, and environmental information), domain knowledge, and expert knowledge. Using ontology with the knowledge engineering practice, a context model for personalized service provisioning and intelligent healthcare facilities has been developed. Ontological modeling of the context and using it for recommendation is the compelling feature of the proposed system. An integrated framework architecture has been developed in addition to the modeled knowledge in order to use the sensed activity information for generating reminders, alerts, and emergency situation analysis for decision making. To achieve better results and provide caregivers an interface for rich interaction, description logic rules have been incorporated. The description logic rules filter out the unnecessary information during decision making.

3.8. Long/ Short Term Behavior Analysis

3.8.1. Introduction

The proposed research Ontology Based Context Fusion for Profiling and Behavior Prediction (OCF-PBP) use the new technologies of information manipulation and new means of communication for information transportation to achieve the desired objectives. The proposed research is to use the technology to collect the context information surrounding a user (patient) as well information of the patient from multiple sources and of diverse nature and later use the collected context information for profiling and behavior prediction. The context information collected are: 1) patient profile information, 2) patient social media interactions, 3) patient diet information, 4) patient daily life activities information, and 5) patient environment information. To collect and fuse/integrate the currently available context information from different devices and creating semantic relationships among the context information will be a practical starting point of realizing the Life Log.

3.8.2. Related Work

- The scientific demonstration of events that cause behavior is called human behavior modeling. There are three types of modeling methods, such as movement models, which Concentrate on the simulation of occupant movement without a behavioral component; partial behavior models that primarily calculate occupant movement, but also simulate behavior to some degree; and the last one is behavior model, which incorporate occupants performing actions in addition to movement to a specified goal.
- Very soon after behavioral ecosystem emerged as a paradigm in the late 1960s and early 1970s, a tradition of applying behavioral models were developed to apply on human behaviors. This tradition quickly became an important voice in the human-related sciences.
- From the 1970s to the present, behavior analysis has become recognized as an empirically supported approach to improving the lives of humans, as well as understanding fundamental principles of human behavior (American Psychological Association [American2010]).

- Some research institutes such as Carnegie Mellon [Brian2009], Virtual Institute on Human Behavior Representation [Institute2010], and Western Michigan University [Michigan] are establishing some behavior modeling systems to help in facilitating the humans for their future predictions.
- There lots of works have been done for human behavior modeling; however, most of them have their own limitations.
- A probabilistic approach [Helaoui2013] that provides great flexibility when controlling different alternatives in the performance of behaviors and may be easily adapted to different environments.
- Similarly, the instance-learning model [Gonzalez2013] has been utilized for the human behavior modeling, where they have shown robust explanations of behavior across multiple tasks.
- Another analysis system has been proposed by [Rowe2013] in which they have presented a behavior ontology that captures user behavior within a given context (i.e. time period and community) and a semantic-rule based methodology to infer the role that a user has within a community based on his/her exhibited behavior.
- Similarly, an automatic annotation of human nonverbal behavior system has been presented by [Knyazev2013]. In this system, various media annotation solutions, computer vision and knowledge representation methods were examined.
- However, probabilistic approaches suffer from the first order assumption and a separate model is required for each, if a behavior is performed in a multiple ways [Rodríguez2014].
- Similarly, Instance-learning model does not has the capability to inference with the command line environment [Dandekar2013].
- Likewise, the main drawback of [Rowe2013] is that: either they use a very reduced set of behavioral features and represent each role with a simplistic combination of these features and their intensity level (high, medium, low) or, if the aim is to cover a broader set of features they need to limit the set of roles they aim to identify (otherwise the combinatorial options may increase significantly) [Zhu2011].
- However, the architectures presented in the nonverbal behavior systems [Knyazev2013] are an adaptive routine for selecting behaviors that depend on characters' emotions and attitudes, non-verbal behaviors were manually encoded within the behavior specification. This limits the design as authors still need to hand-code all non-verbal behavior patterns and vary them based on variations in the character models [Á lvarez2012].

3.8.3. Components Description

- Input: It takes input from the low level context-awareness, high level context-awareness, and HDFS Data Interface modules.
- Output: It provides output reasoner and prediction, and recommendation manager modules for further processing.

Context Receiver

- Context receiver is a component that is responsible for receiving data inputs from diverse sources such as tweets data, trajectory data, and social data etc. These engines will provide output in different formats like XML and simple text. So there is need of receiver in the engine to properly parse all type of output produced by these different engines. Once all the data has been received, then it can extract the data from the file and provide these to the next module i.e. Activity Representation.

Context Representation

- It is important to convert the incoming context into the required format defined using Ontology Based Life Log. As we consider multiple input modalities and with diverse information, so we will develop a novel context conversion and representation algorithm that will represent the emerging context information from diverse sources in a uniform representation.

Context Fusion

- Context received at different time intervals and of particular domain can be of vital importance and might need to be fused together (as mentioned above) which will provide a more sophisticated view of patient situation and behavior. To achieve the objective of context fusion, we will design and develop horizontal and vertical context fusion algorithms to fuse the related and dependent context for better understanding of patient situation and behavior.

Context Verification and Logging

- As the context information coming from social media, diet, sensors, and environment are independent, so it is hard to associate context information from multiple independent sources with a particular patient. To handle the situation, we will develop an algorithm to associate and authenticate such information with particular patient. Afterword's, the represented context information also needs to be verified for its structural, semantic, and existence against the Life Log. This research will contribute to the development of such algorithms for verification of context for its structural, semantic, and existence against the Life Log.

Parser

- For any type of information manipulation from the Knowledgebase, parser is responsible to properly handle all the operation regarding that matter. The parser normally communicates with activity representation to properly represent the activity, it also parse the knowledgebase for the inference engine for verity of different reasons like verification of activity and decision making, To populate the knowledgebase for newly recognized activity, the parser is also used in that case.

Life Log Extractor

- The development of Life Log is not a new idea; however, the work discussed above have limited expressivity and have less comprehension as they focus on limited aspects of user daily life context. Some of the logs only focused on text data, some on audio and/or video, some on activity information coming from different sensors, and some just on user online and computer usage behavior. On the other side, we not only include all the aspects; moreover, we also include patient profile information, patient travel history, patient diet information, and patient environment information. So to mainly mention the originality, the existing systems mainly work with one input limited input modalities whereas, we are focusing on multiple input modalities and with diverse nature of information which is totally preserved with its semantics using ontology.

Behavior Analysis

- To use the context information logged in the Life Log repository, a need is to build data mining algorithms that can process the information and uncover the hidden information. The existing systems worked on identifying abnormal behavior based on only single type of information and report those to caregiver. No one has worked on analyzing user (patient) behavior and lifestyle based on patient daily life behavior. So our contributions are to develop algorithms for analyzing patient behavior/lifestyle for short-term and long-term, monitoring patient behavior for daily life activities, and predict patient behavior based on current activities profile of patient against the

logged behaviors of previous patients that can facilitate in providing batter picture of patient overall behavior.

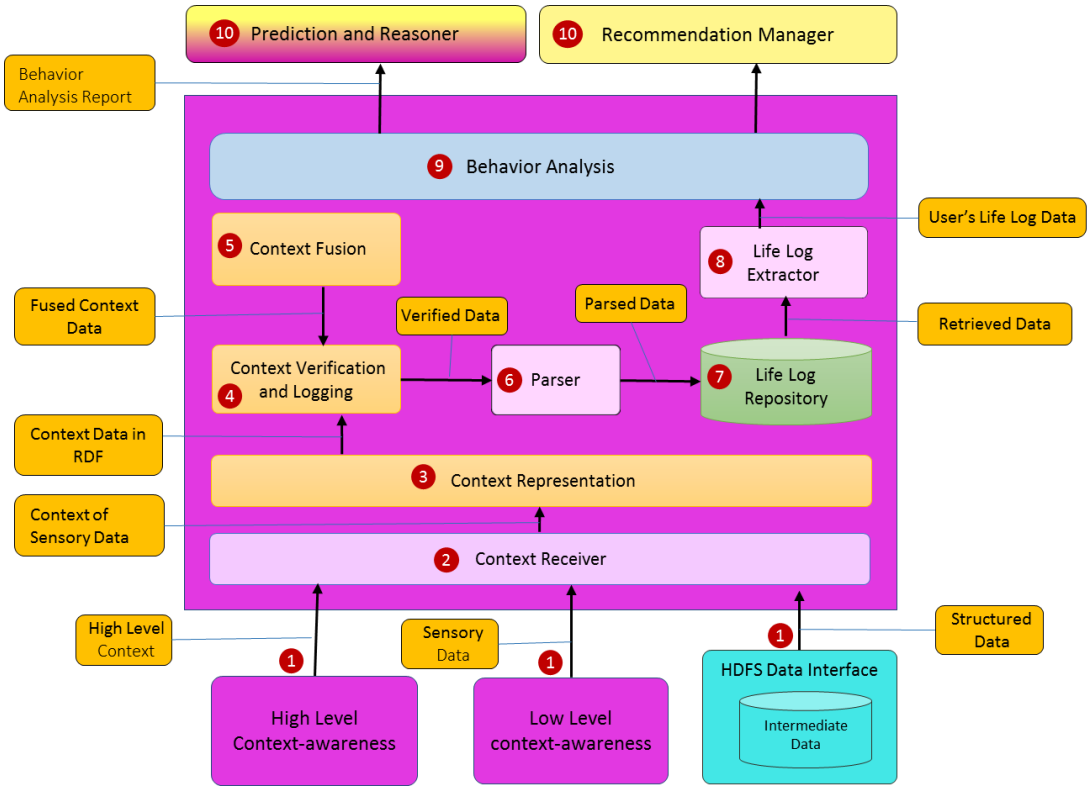


Figure 3.16 Architecture of Long/Short Term Behavior Analysis for Mining Mind.

3.8.4. UML Diagrams

- Sequence Diagram

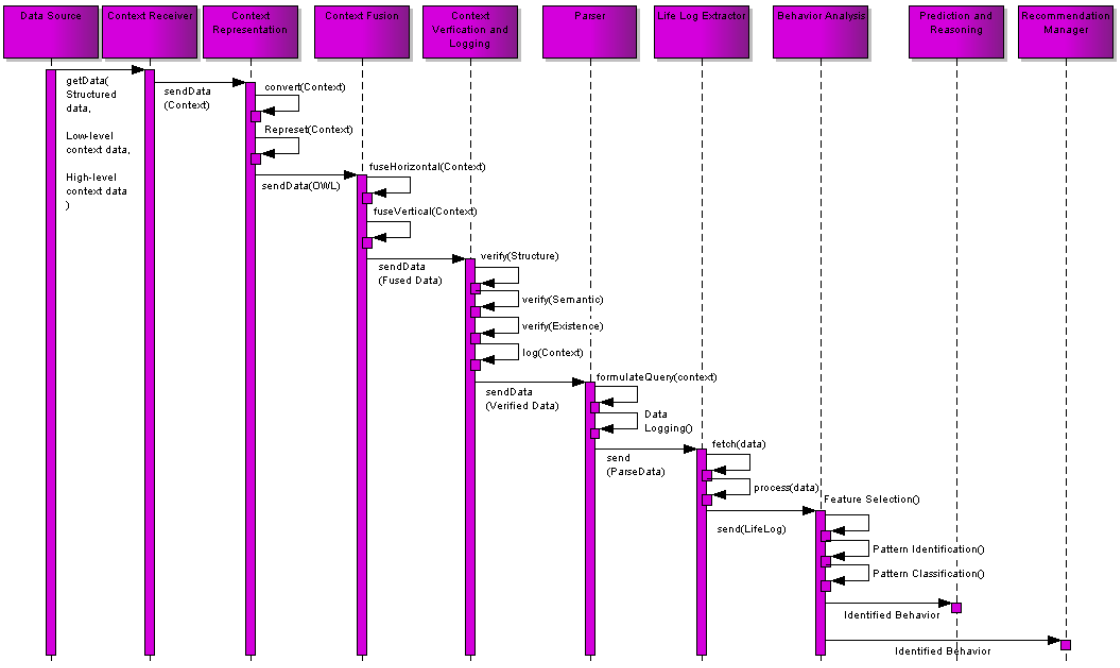


Figure 3.17: Sequence Diagram of Long/Short Term Behavior Analysis for Mining Mind.

- Component Diagram

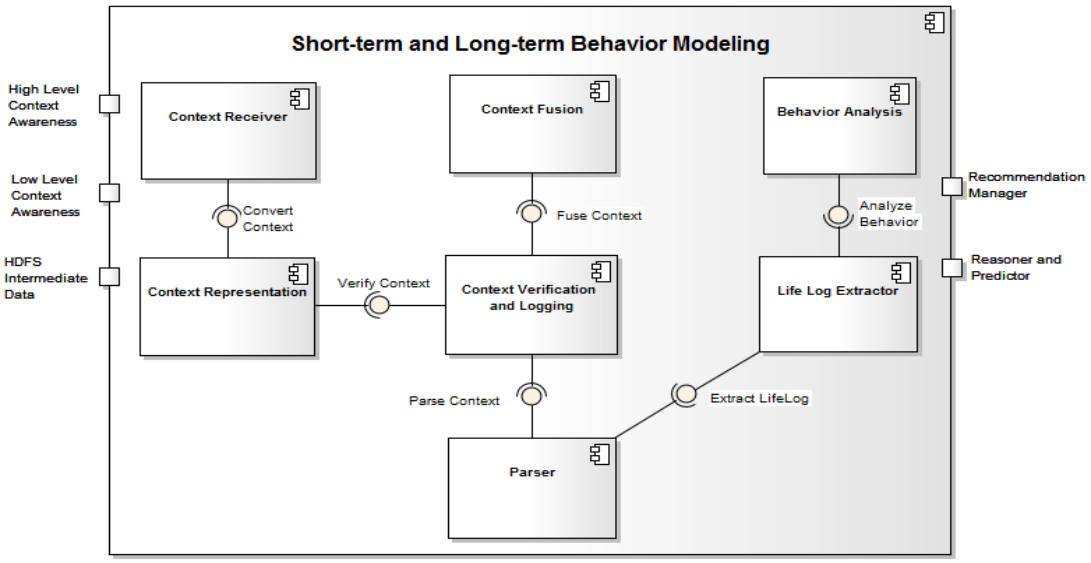


Figure 3.18: Component Diagram of High Level Context-awareness for Mining Mind.

3.8.5. As Is – To Be Systems

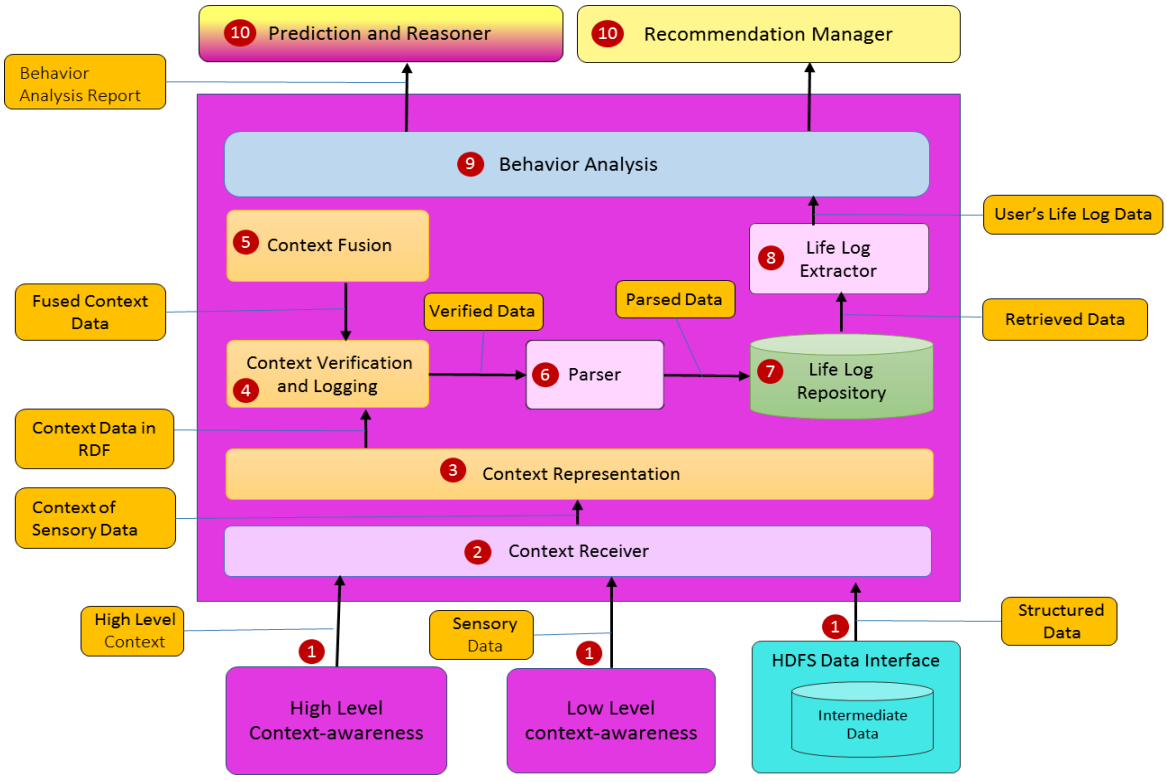


Figure 3.19: As Is – To Be

3.8.6. Uniqueness

- Development of Semantically Rich and Logically Structured Life Log Repository for logging the context information and patient profiles.
- Context conversion and representation algorithm development
- Fusion of context information from multiple information sources and with diverse formats and at different time intervals.
- Development of Horizontal and Vertical Context Fusion Algorithms.
- Patient context authentication and context verification algorithms
- Use of context information from Life Log for Patient Profiling.
- Design and development of Patient Behavior Prediction, Monitoring, and Analysis services for better understanding of his/her behavior in daily life.
- The overall system will provide patients and caregivers such as Understand patient situation and behaviour with right information at right time and at right location.
- Provide patient with self-confidence for performing certain daily life tasks
- Manage the physical, emotional and social Impact of the conditions on patient's lifestyle and behaviour.
- Prediction of future patient behavior to facilitate in living healthy Lifestyle

3.8.7. Summary

The purpose of this research is to fuse the daily life context information of patient coming from various sources and build patient profile based on the emerging context information. The profile of a patient is then used for patient behavior evaluation, analysis and prediction based on existing behavior models of similar disease patients. For the purpose of context fusion, ontology based horizontal and vertical context fusion mechanism is proposed that will fuse patient context information (i.e., physical activities, social media interactions, diet information, and environment information) emerging from diverse sources and at different time intervals of a day. The fused context is then logged in Life Log maintaining the overall profile of a patient. A semantically rich and logically structured ontology based Life Log repository is proposed and designed in this research. It is used for logging patients' context information coming from diverse input sources, patients' profiles and past patients behavior models. A behavior prediction algorithm is proposed that predict patient behavior based on the patient's profile, patient current context (activity) and existing patients behavior models stored in the Life Log.

3.9. Information Curation Layer Overall Uniqueness

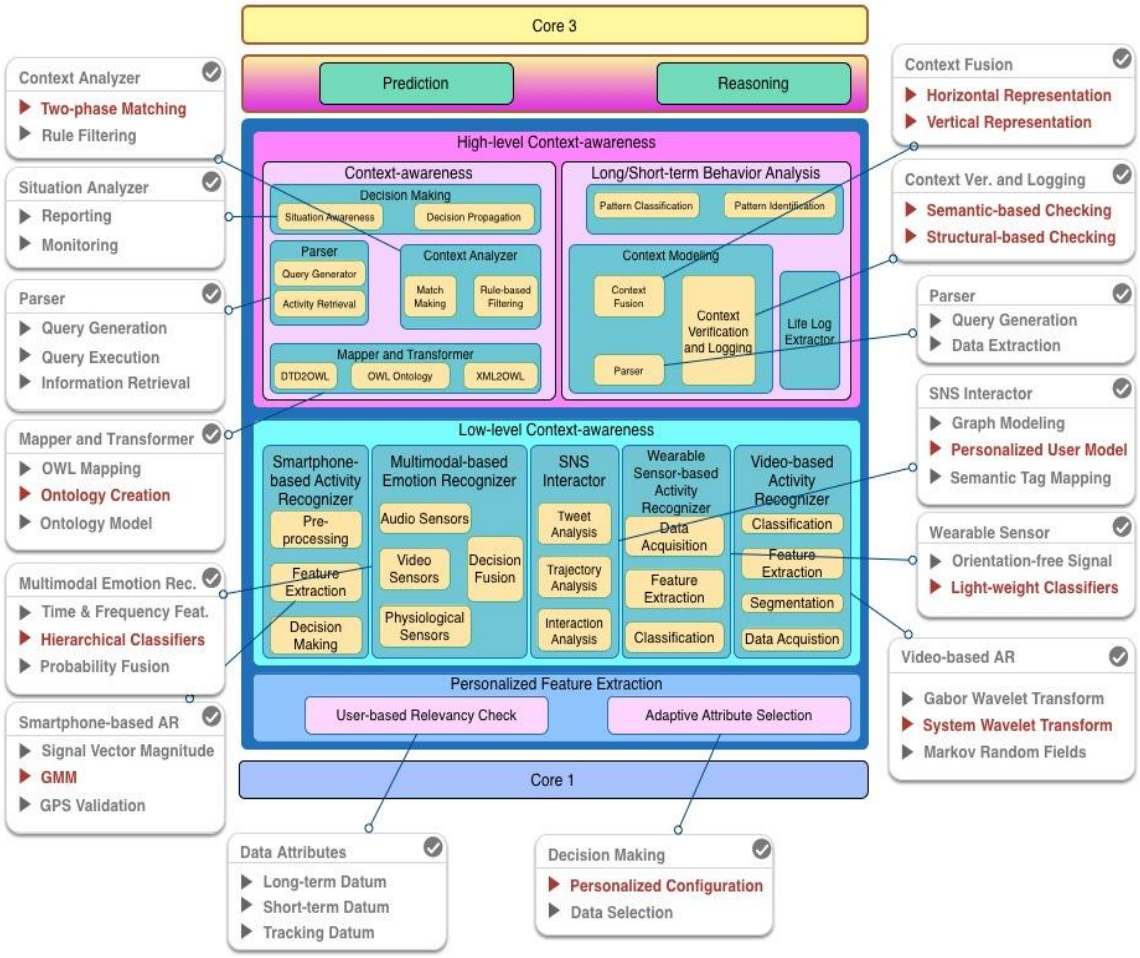


Figure 3.20: Information Curation Layer Uniqueness

Chapter 4
Service Curation Layer

4. Introduction to Service Curation Layer

The service curation layer curates the services based on the user's curated information. It consists of the reasoning and prediction module to create the personalized services. Our unique knowledge maintenance engine and feedback also resides in this layer to maintain the knowledge in the proposed platform. The details of the key methods and our unique component (i.e. highlighted red) are mention in the following figure and details are provided in the sub-sequent sections.

4.1. Service Curation Layer Functional Diagram

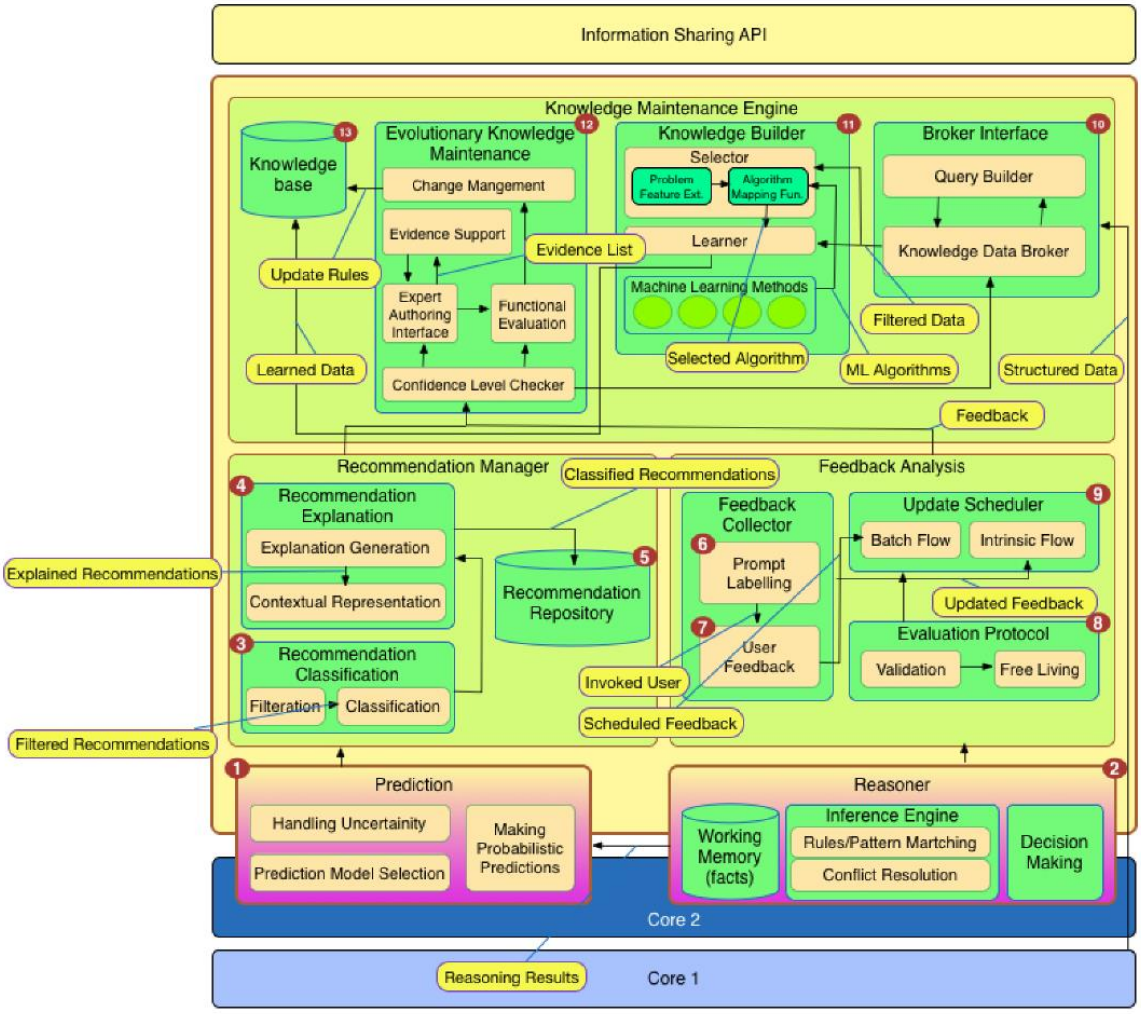


Figure 4.1: Service Curation Layer Functional View

4.2. Service Curation Layer Use Case Diagram

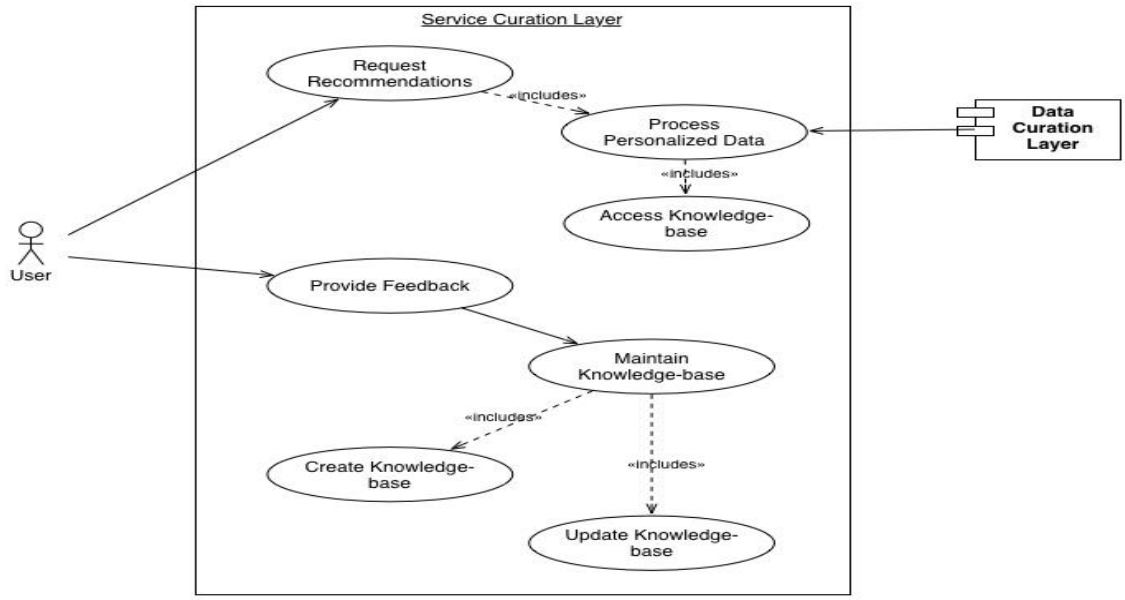


Figure 4.2: Service Curation Layer Use Case Diagram

4.3. Service Curation Layer Sequence Diagram

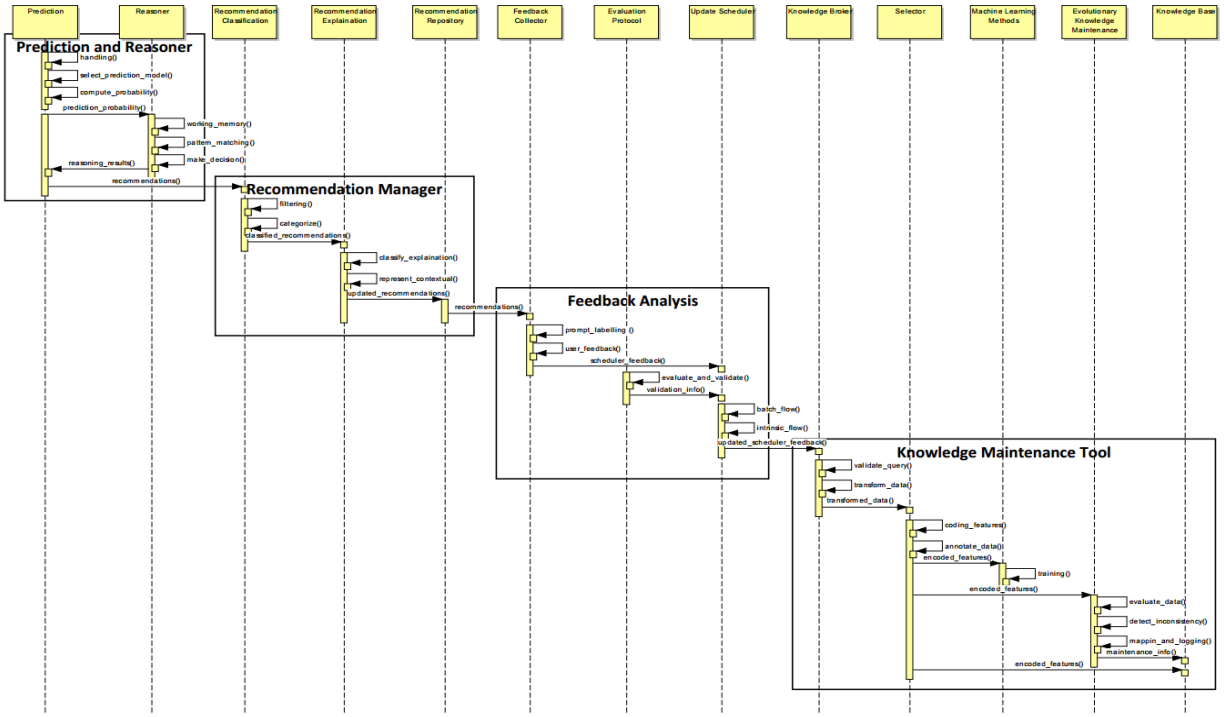


Figure 4.3: Service Curation Layer Sequence Diagram

4.4. Service Curation Layer Component Diagram

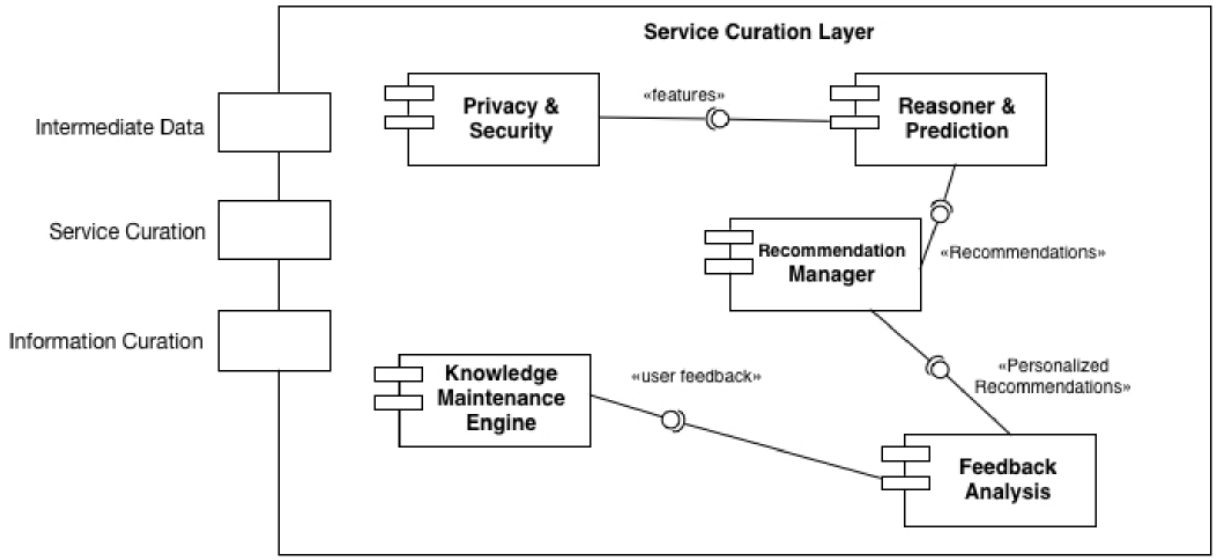


Figure 4.4: Service Curation Layer Component Diagram

4.5. Reasoner and Predictor

4.5.1. Introduction

In the present age, the healthcare data increase drastically to gigabytes and terabytes with hundreds of attributes. The human brain cannot examines this much variables at a time to profile a healthcare problem. To overcome this problem, machine learning methods are used to learn this data and then provide reasoning and predictive modeling techniques for making diagnosis and prediction tasks. The purpose of this module is to automatically generate recommendations to the users based on their requests.

4.5.2. Related Work

- Reasoning and prediction systems have widely been studied and used in domain, such as medical (e.g. PIP, INTERNIST, CASNET, and MYCIN), business markets for predicting stock prices, weather forecasting, fault proneness in softwares systems, lightning detection and prediction (e.g. SkyScan, THOR GUARD) and control system engineering.
- In medical, the Pharmacogenomics project (2006-2008), the viral genomics data is integrated with clinical data to predict responses to anti-HIV treatment to the clinicians [Zazzi2010]. Several prediction engines are designed to predict the efficiency of possible drug combinations and to recommend an optimal treatment by combining the results from all together.
- Hybrid reasoning systems, such as "a personalized wellness recommendation system" [Husain2010], for predicting therapy and Context-Aware Real-time Assistant (CARA) [Yuan2014] for personalized healthcare services of the elderly people have been studied using case-based and rule-based approaches. Similarly, probabilistic approaches (e.g., Bayesian and Apriori) have been used for predictions of hepatitis infection [Drăgulescu2007], and diabetes [Pandey2012].

- In existing reasoning and prediction systems, the predictions are based on knowledge base, learned in its first phase, with minimum support for data abstraction. These systems either have reasoner or predictor as the main engine and the both. Likewise, they use a single learning and conflict resolution method for learning rules and resolving conflict during the inference process. We propose an integrated system with a high-level data abstraction, reasoning and prediction components using a hybrid approach for providing accurate and precise recommendations.

4.5.3. Components Description

- Input: Structured Data
- Output: Reasoning results and Predictions (Recommendations)
- Description: The prediction and reasoning engine provides a set of sequential steps to offer the users a complete and straight-forward solution to first learn data from the data and then predict different recommendations for the users.
- High-level Data Abstraction & Learning module is to preprocess and abstract the intermediate data.
- The knowledge base is the repository which stores the learned model in locally for quick access and use during the reasoning and prediction processes.
- The reasoning module uses the learned model stored in the knowledge base to derive new knowledge/inference results.
- For making future predictions over the inferred results or the learned model, the prediction module gets activated.

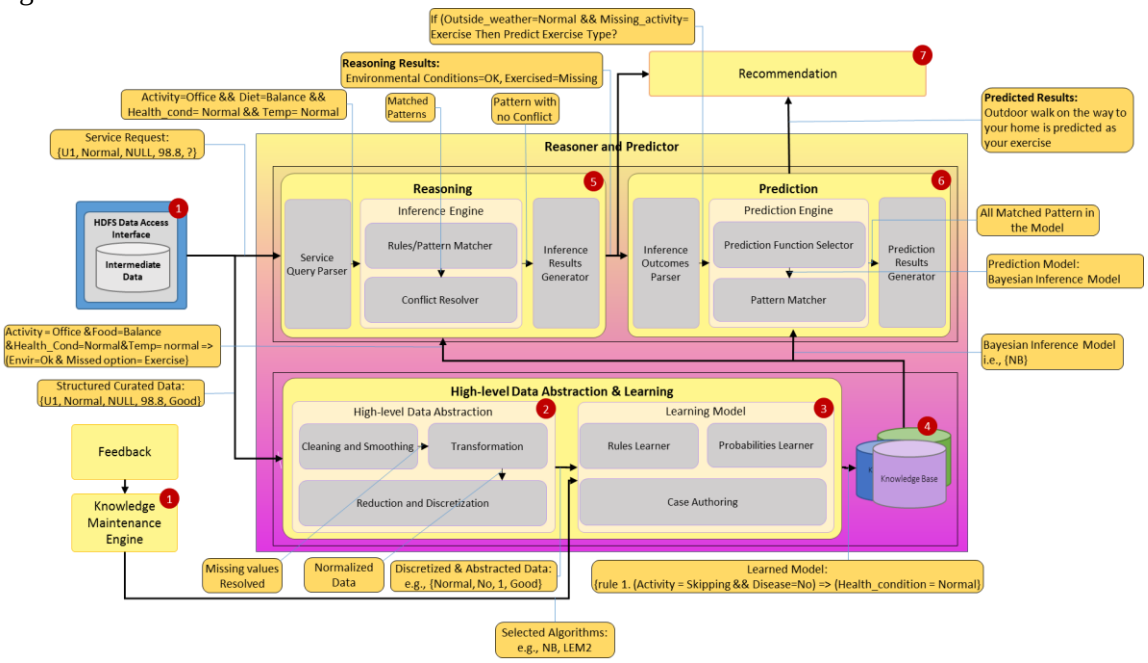


Figure 4.5: Architecture of Reasoning and Prediction for Mining Mind

High-level Data Abstraction & Learning

- Input: Structured Data
- Output: Highly abstracted data
- Description: The aim of this module is to preprocess, normalize and reduce the structured data and then forward to different machine learning methods for learning.
- The Cleaning and smoothing component fills missing values, remove noise and inconsistencies

- The transformation component normalization, attributes construction, aggregation and generalization
- The attributes are reduced and the values are discretized by using the data reduction and discretization component.
- The rules learning components learns if-else rules from the abstracted data using learning from example method (i.e., LEM2).
- The probabilities learning components learns prior and conditional probabilities from the abstracted data using probabilistic approaches (i.e., NB).
- The case authoring components authors user's past cases from the abstracted data using case-based approaches (i.e., CBR).

➢ Knowledge Base

- The knowledge base component is the repository which stores the learned model in the locally for quick access and use during the reasoning and prediction processes.

Reasoning

- Input: Service Query and Learned Model
- Output: Reasoning Results (Recommendations)
- Description: To automate the process of inferring output/decision for the curated service, the reasoning module of the proposed system uses the learned model stored in the knowledge base. This module, derives new knowledge/inference results via the application of multiple steps in specific sequence.
- The service query parser component of the reasoner initializes the reasoning process. It parse the query and make it computable with the learned model to be executed.
- Using search techniques and pattern matching, the inference engine starts with all the known facts of the service query and looks for them in the learned model loaded from the knowledge base. A data driven or forward chaining mechanism is used to draw reasoning results.
- When patterns of the working memory match against multiple rules of the knowledge base, a conflict occurs, in which case, the conflict resolver takes action and resolve the conflict.
- The inference results generator generates the final decision/reasoning results for the required curated service and presents to the user directly or forwards to the prediction module for further prediction service.

Prediction

- Input: Reasoned Results and Learned Model
- Output: Predictions (Recommendations)
- Description: For making future predictions over the inferred results or the learned model, the prediction module gets activated and gives future prediction for the curated service.
- The inference outcome parser of the predictor initializes the prediction process. It parse the outcome received from the reasoned and analyze them so that to make accurate and appropriate predictions over them.
- The prediction function selector component uses different methods, such as rule-based selection and meta-level selection to select the appropriate prediction model for generating prediction results.
- The mining mind pattern matcher component takes the reasoned results as the query and using the selected prediction model to match their pattern and predict the new results.
- To generate the final prediction results, some probabilistic techniques, such as NB and logistic regression are used by the prediction result generator to generate the final results and present to the user.

4.5.4. UML Diagrams

- Sequence diagram

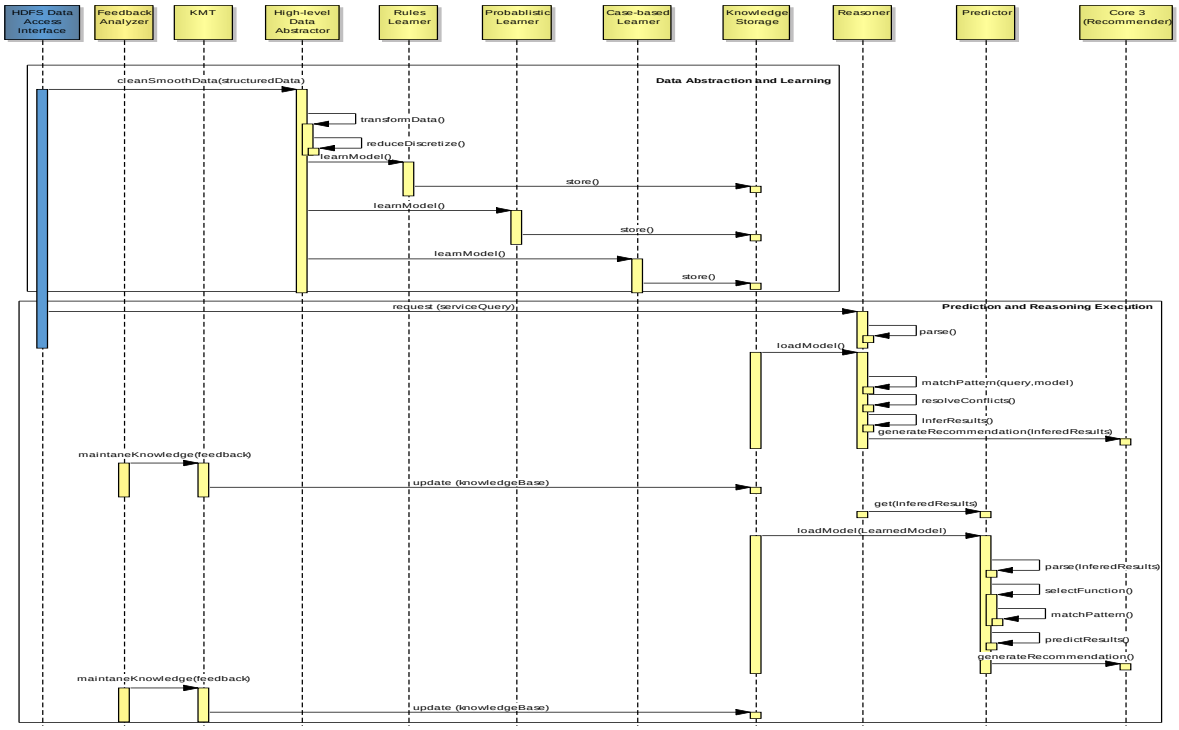


Figure 4.6: Sequence Diagram of Reasoning and Prediction for Mining Mind

- Component diagram

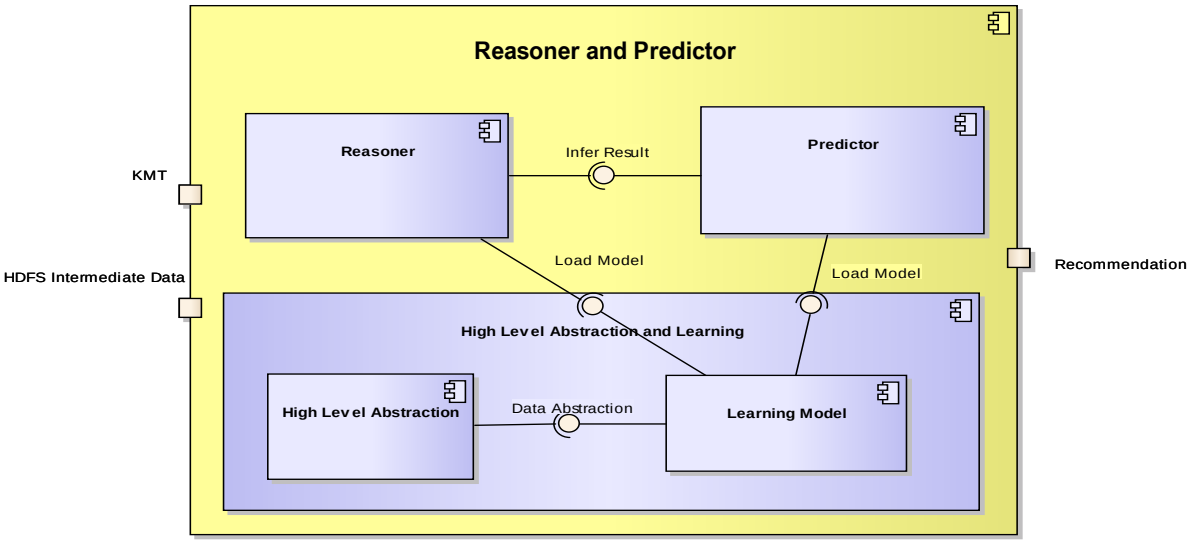


Figure 4.7: Sequence Diagram of Reasoning and Prediction for Mining Mind

4.5.5. As Is – To Be Systems

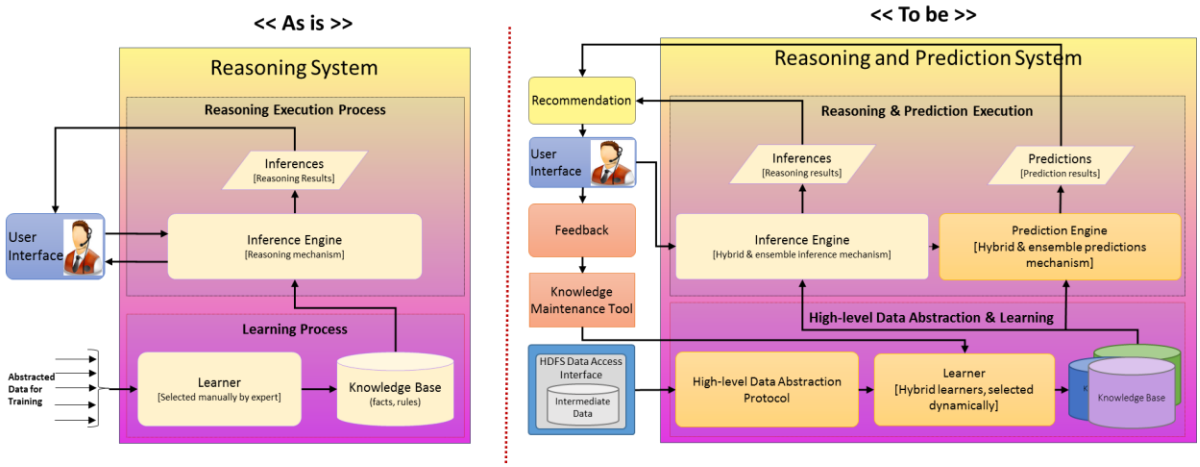


Figure 4.8: As Is – To Be

4.5.6. Uniqueness

- The development of an integrated environment for high-level data abstraction, learning, reasoning and prediction.
- Provides a high-level data abstraction protocol for abstracting the intermediate data to facilitate the learning process.
- Hybrid and ensemble approaches for reasoning and prediction by the use of multiple base learners either in parallel or sequential way for increased accuracy of predictions.
- Provides not only inferred results (output of reasoner) to the users but also utilizes them in higher level future perdition.
- Handles the uncertainty of the inferred results and predicted outcome by the use of hybrid techniques for more precise predictions

4.5.7. Summary

Reasoning and prediction module uses hybrid and ensemble learning methods to learn data stored in the intermediate data, after curation by the data curation module of the proposed mining mind system. The generate knowledge is stored in the knowledge bases. In the service execution scenario, the reasoning and prediction components generate recommendations to the users for their queries

4.6. Knowledge Maintenance Engine

4.6.1. Introduction

- For effective knowledge base it is critical that knowledge base should not go stale and old. For this problem different machine learning and evolutionary techniques are required to update and maintain the knowledge base.
- There is a need for a system which uses machine learning methods to continuously update its knowledge base in response to changes in the application domain or in the requirements of the application. Evolutionary management and maintenance of the knowledge base is a necessity for a user recommendation system as well as a challenge.
- Dynamic construction of the knowledge base through our knowledge service tool and evolutionary knowledge maintenance using both automatic and expert involvement makes the platform unique and distinguishable.

4.6.2. Related Work

- An offline prediction system to select and rank algorithms based on performance is proposed by Lars et al, [Lars2013] through classification and Meta learning. The prediction output is ranking of the algorithms. They proposed 3 levels to categorize the predictive output of a model with respect to what ranking may be obtained from it.
- An approximate ranking algorithm is proposed in [Quan2013] which is based on large data sets and improves and their technique improve the overall performance of meta-learning for algorithm ranking and they optimize the parameters of base learner for specific datasets.
- Hutton et.al described a model-based approach for identifying a small set of both algorithm parameters and instance features that suffices for predicting empirical algorithm performance well. The authors [Hutton 2013] show that for parameter configurations sampled uniformly at random—very good performance predictions can be obtained based on just two key parameters, and that similarly, few instance features and algorithm parameters suffice to predict the most salient algorithm performance characteristics in the combined configuration/feature space.
- Debenham et.al all proposed an analysis of knowledge base maintenance and managed it by constructing a formal model. In this model the representation of each chunk of knowledge encapsulates the knowledge in a set of declarative rules, each of which in turn encapsulates the knowledge in a set of imperative programs [Debenham2005].
- In one other approach [Debenham2003] proposed that knowledge base is maintained by modifying its conceptual model and by using those modifications to specify changes to its implementation. The maintenance problem is to determine which parts of that model should be checked for correctness in response a change in the application.
- A "knowledge library" model is proposed for the distributed maintenance of the integrated knowledge base that drives the clinical decision-support, physician order entry, and notes capture system [Geissbuhler1999]. They minimize dependencies between knowledge sources, the integrated knowledge base, the individuals maintaining various components, and the systems that use the knowledge. Standardized vocabularies and terminologies are employed to minimize the inconsistencies and discrepancies.
- According to [Regier2009], the main challenge is to keep the knowledge bases up to date and decrease the turnaround time for logical changes in the existing rules of knowledge base in every expert system.
- More expressive and semantically as well as syntactically complex knowledge rules need to be edit with rule editing systems that help in maintaining the knowledge base [Kaljurand2008]
- The success of any contemporary CDSS is based on KB upon which it is built [Hulse2005]. The crucial aspect for KB is the richness with respect to domain knowledge which is mainly dependent on domain experts.
- A sheer amount of literature is available online in the medical domain. Only MEDLINE/PubMed Baseline yearly citations totals 22,376,811 from 2014 reported in statistical reports on MEDLINE/PubMed Baseline data by U.S. National Library of Medicine [MEDLINE2014]. With proper utilization of the literature, domain expert can help in knowledge rules creation and modification.

4.6.3. Components Description

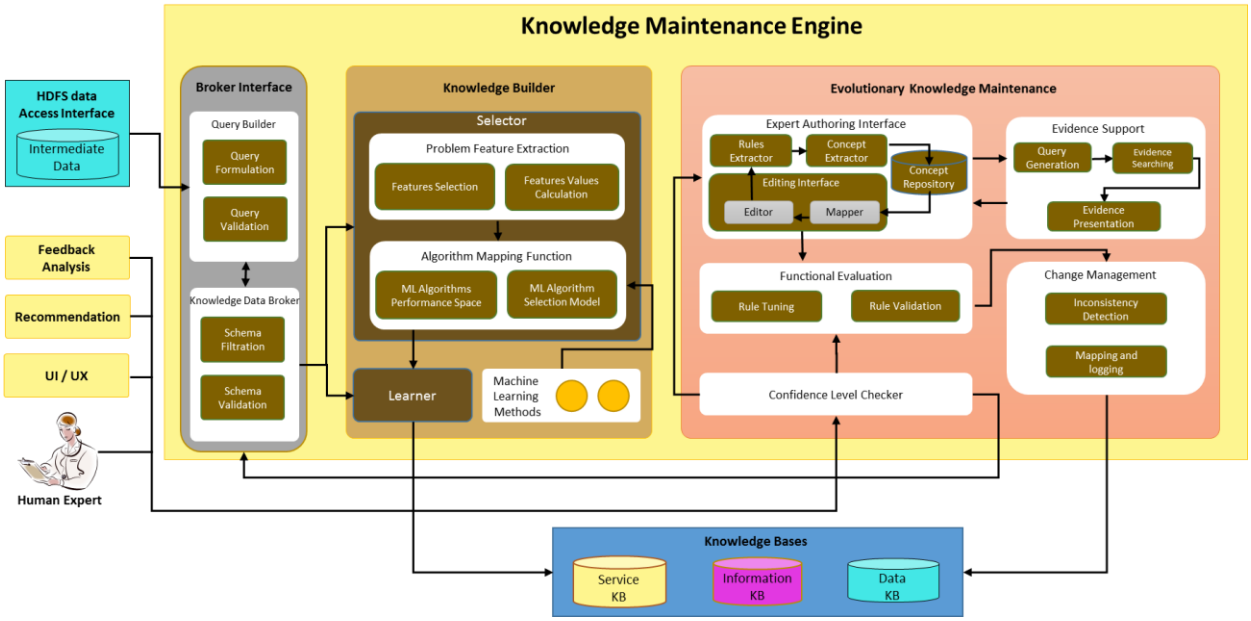


Figure 4.9: Architecture of Knowledge Maintenance Engine

Query Builder

- Input: Structured Data from HDFS intermediate data
- Output: Queried data from intermediate data
- Description: The query builder retrieves the data from the intermediate data. Query Formulator builds the query with relevant operators and parameters, and queries the intermediate database. The Query Validation validates the query before execution of the query.

Knowledge Data Broker

- Input: Structured Data from Query Builder, User Feedback, Recommendation, UI/UX, and Human Expert.
- Output: Filtered data for the Selector module
- Description: The knowledge data broker further filters the queried data to remove unnecessary schema for improving the accuracy.

Problem Feature Extraction

- Input: Filtered data from the Knowledge Data Broker.
- Output: Problem Features with their values

  Description: This component selects various features and then calculates the values of that features. This component determines the nature of data.

Algorithm Mapping Function

- Input: Problem Features with their values from Problem Feature Extraction.
- Output: Machine Learning Method
- Description: The Algorithm Mapping Function selects various machine learning methods and calculates the performance of all features against each selected machine learning method. Based on high performance value, this function chooses the machine learning method and pass this information to Learner component.

Learner

- Input: Machine Learning Method and filtered data from the Knowledge Data Broker.

- Output: Learned data
- Description: The Learner learns the data by using machine learning method. After learning, the created data are stored into the knowledge base.

Confidence Level Checker

- Input: Rule and its confidence level from User Feedback, Recommendation, UI/UX, and Human Expert.
- Output: Rule and Maintenance type in term of Automatic / Manual
- Description: In maintenance module, automatic and manual updation are done. The maintenance depends upon the value of confidence level. If the confidence level is high, then maintenance is done automatically using Functional Evaluation and Change Management functions. In case of low confidence level, maintenance is performed either automatic or manual. In case of automatic with low confidence value, the maintenance is done using Knowledge Data Broker, Selector, and Learner functionalities. While for the case of manual with low confidence value, maintenance is done using Expert Authoring Interface, Evidence Support, Functional Evaluation, and Change Management functionality.

Functional Evaluation

- Input: User Feedback, Recommendation, UI/UX, and Human Expert.
- Output: Tuned Knowledge Rule

  Description: In Functional Evaluation, the feasibility of the update is checked that whether there is need to tune the rule or not. Finally the rule is validated before it is stored into the KB.

Change Management

- Input: Tuned Knowledge Rule
- Output: Updated Knowledge Rule

  Description: The Change Management creates a history of changes and logs it. There can be many inconsistencies while updating the KB so inconsistency is checked for both semantic errors and syntactic errors.

Expert Authoring Interface

- Input: Knowledge Rule needs to edit
- Output: Updated Knowledge Rule

  Description: Expert Authoring Interface is a component to provide a user friendly interface to knowledge experts to edit the existing rules or create new rules in the knowledge base. This component provides the Intelli-sense functionality that helps in creating and editing the rules in a rapid and easy manner. This functionality increases the recall of concepts for the experts and decreases the ratio of errors during the editing of rules. The experts authoring environment is very important to enhance the user's confidence on the rules that created by the machine learning techniques.

Evidence Support

- Input: Concepts from the Knowledge Rules
- Output: List of relevant evidences from online credible sources
- Description: The Evidence Support accepts the concepts from the rules under consideration for a given subject. Query is generated automatically from the concepts and put on search function of target source(s). The relevant evidences are compiled and presented to the user helping in authoring of rules.

4.6.4. UML Diagrams

- Sequence Diagram

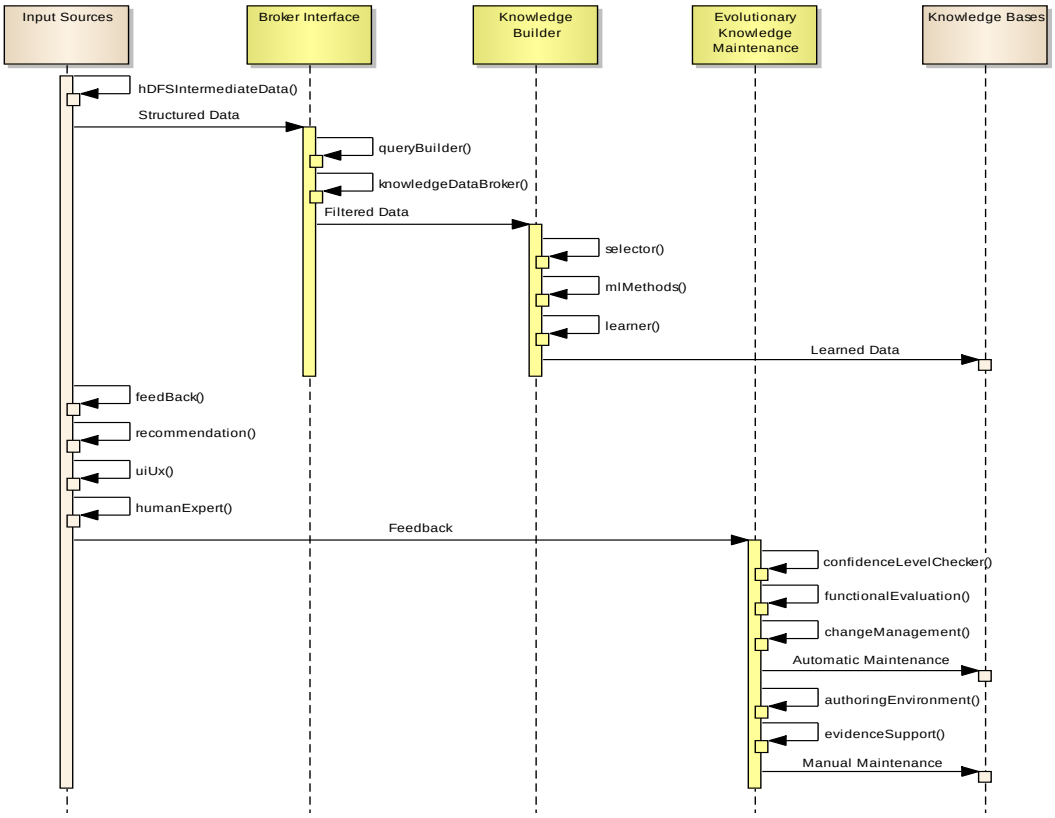


Figure 4.10: Sequence Diagram of Knowledge Maintenance Engine

- Component Diagram

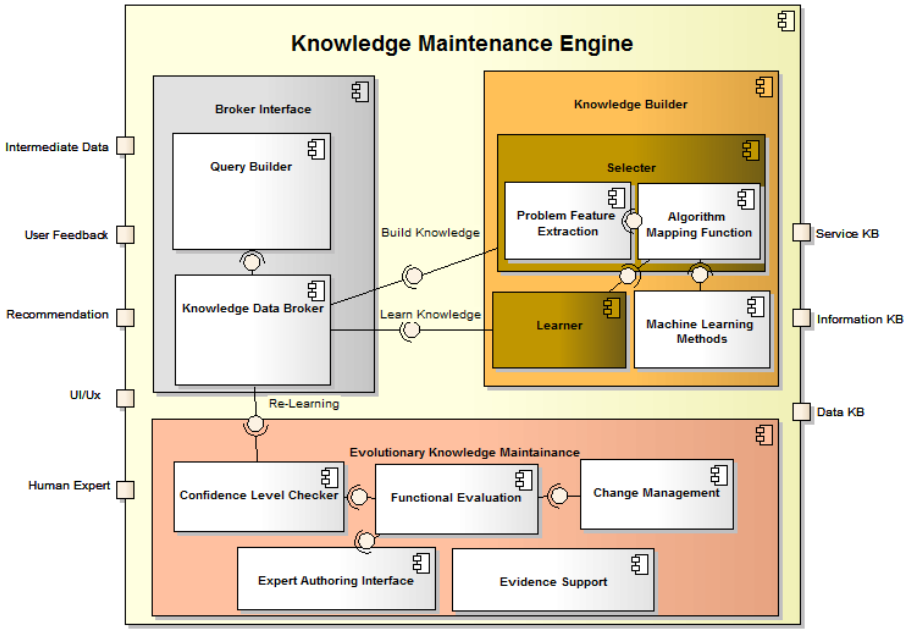


Figure 4.11: KME Component Diagram

4.6.5. As Is – To Be Systems

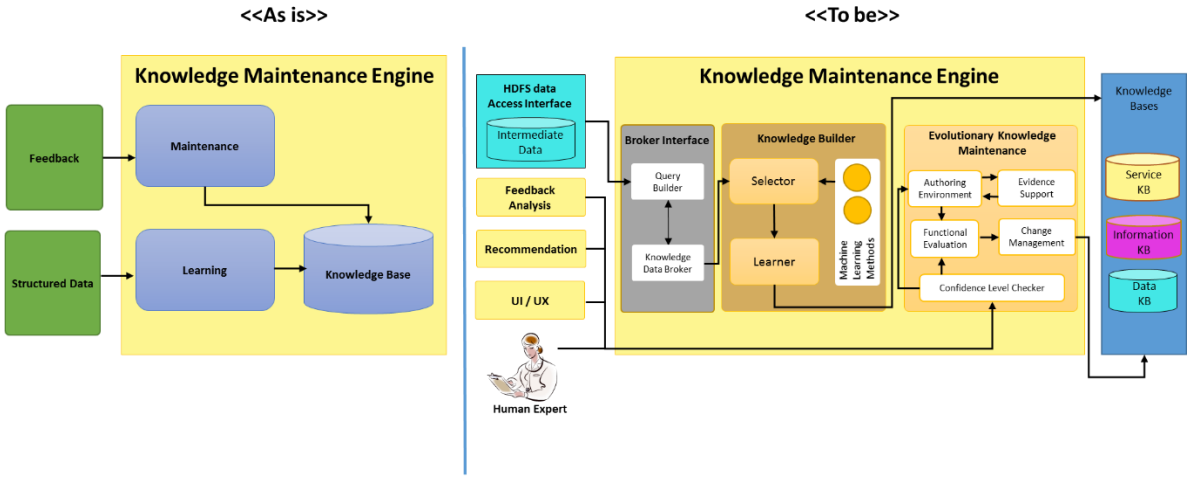


Figure 4.12: KME As Is – To Be

4.6.6. Uniqueness

- Better Quality of service
- Selection model (Mapping Function)
- Inconsistency checking and Rule Validation
- Automatic and Domain Expert Maintenance
- Multi-level Knowledge Base Maintenance
- Dynamic Selection of Machine Learning Methods
- Evidence Support

4.6.7. Summary

Knowledge maintenance engine is nowadays a hot topic in the machine learning field specifically for health related systems and **clinical decision support systems**. This makes it a challenge as many inconsistencies could arise during the updation and maintenance phase. Knowledge base maintenance is a very costly operation as different logical error can come so **inconsistency checking** is nowadays one of the focal points of research in latest research. The new approach for the knowledge maintenance is the **distributed approach** to introduce resilience to the system and make it more error free. Moreover for selection of machine learning techniques for the required purpose is being done through Meta learning and different **ranking** systems. Not always but most of the times automatic generation rules creates anomalies in the rules which need to be corrected from the domain expert. We provide expert authoring interface to author rules in easy to use editor with the support of evidence available from credible online resources.

4.7. Feedback

4.7.1. Introduction

Feedback is one of the more important process to maintain the systems and bring robustness. From the gaming industry, operating systems and developed software applications. The maintenance services is cost effective solution to provide services. A huge cost is spent to maintain the systems and increase the user's usability's. Existing system get user's feedback in a very straight forward manner and requires novel methods to incorporate the diverse sources. One of the most important modern method is social media where users can share their minds to use the services. Process and

methods are required to incorporate such feedback. Secondly, feedback mechanism should be invoke for the user in user's friendly manner to get the correct response.

4.7.2. Related Work

- [xbox2014] Xbox addressed the community feedback to make the system robust and increase its performance. When Xbox One is in the low power Instant On mode, it will automatically download the update when available, and then install it the next time you turn the system on. It's just the update rolling itself out in the background. Give it some time. Go have a coffee. Take a walk. Read a book, even. Use that time to reflect on the state of the world and your place in it.
- [Zhang2011] In e-commerce, how to provide accurate and valuable information for customers is the main problem that the personalized recommendation system needs to solve. They developed a personalized recommendation system model based on customer feedback and focuses on the recommendation algorithm.
- [Vashisth2011] They proposed a design framework for a personalized multi-agent IBRS. The IBRS is an agent- based recommender system that takes into account user's preferences to generate recommendations. The system is based on the agents having Belief-Desire-Intention (BDI) architecture. These BDI user agents are empowered with cognitive capabilities and interact with recommender and other user agents using argumentation. The explanation process uses argumentation so that the recommender can look deeper into the reasons behind user's likes and dislikes. The IBRS considers user's feedback for recommendation repair action. This results in improvement of the personalization process. The experimental study is conducted for Travel Recommender System to show that personalized interest-based recommendations improve quality.

4.7.3. Components Description

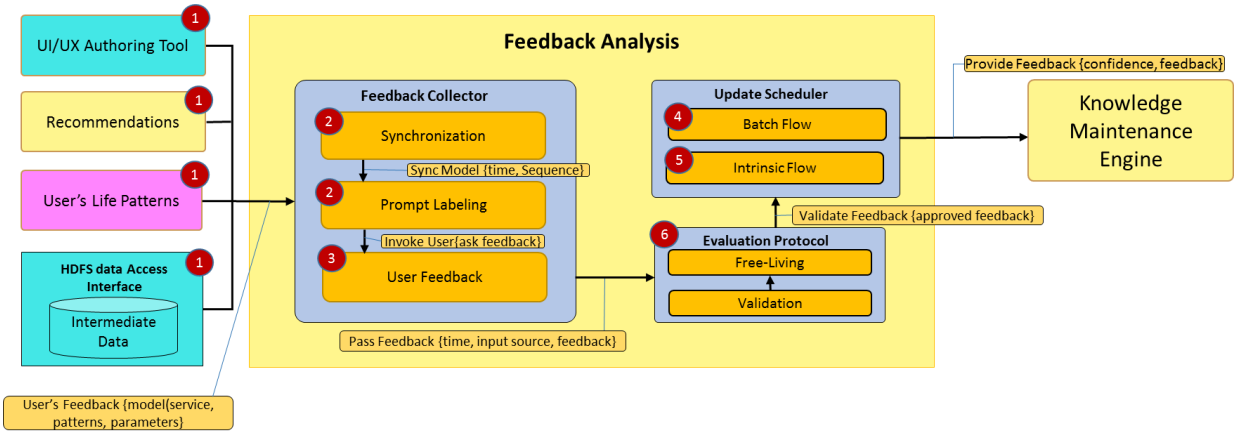


Figure 4.13: The Proposed Architecture of Feedback Analysis

Feedback Collector

- Input: User's life patterns, SNS interaction and provided Recommendation are the input to feedback analysis
- Output: Provide the maintenance parameters to the knowledge maintenance engine to update the system.
- Description: Our platform will provide the curated services to the user then it ask to provide the feedback about the provided services.

- We will develop the procedures for synchronization of multiple input sources as a feedback.
- System automatically understand the transition of contexts and pop-up the prompts.
- Next, it prompt the user's friendly interfaces for getting feedback.

➢ Evaluation Protocol

- Input: User's feedback after synchronization and collector
- Output: Evaluated feedback
- Description: It will evaluates the received feedback by our cross check routines and defined procedures.
- User's ease is very important to get feedback so we also define the number of prompts that user will receive per day.
- Evaluation protocol will also differentiate the feedback about the contexts to update the machine learning algorithms parameters and about curated services to make them more personalized.

Update Scheduler

- Input: Evaluated feedback
- Output: Reflect into system with the help of Knowledge Maintenance Engine
- Description: Our evaluation protocol will validate the user's feedback automatically and ready to reflect in the system with our two defined scheduler methods.
- If it's related to update the parameters of the machine learning methods then it will schedule according to the data management jobs.
- If it is related to the provided curated service then reflect immediately after evaluation process to give robust services for the next time.

4.7.4. UML Diagrams

- Sequence diagram

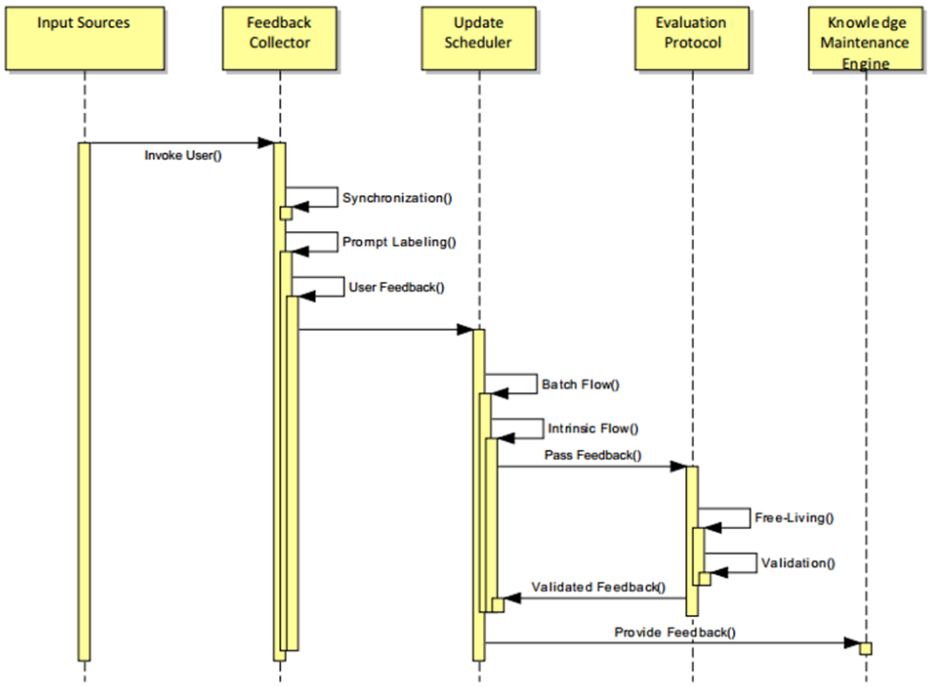


Figure 4.14: Sequence diagram of Feedback analysis

- Component diagram

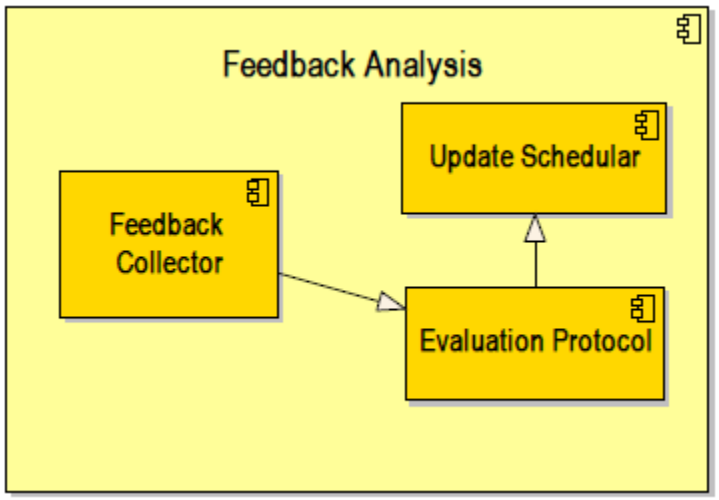


Figure 4.15: Component diagram of Feedback analysis

4.7.5. As Is – To Be Systems

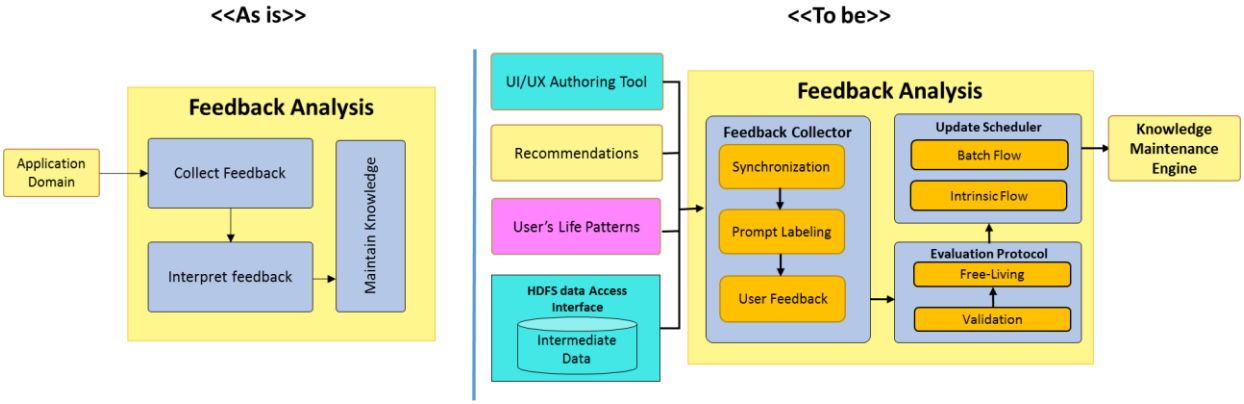


Figure 4.16: Feedback Analysis As Is – To Be

4.7.6. Uniqueness

- We are introducing the prompt labeling mechanism to get the user's feedback.
- We are introducing logs management to keep the track of the situation.
- We are introducing evaluation protocol for validating feedback.
- We are introducing with two scheduler methods for reflecting the user feedback.

4.7.7. Summary

Our designed mechanism can provide user's friendly and more interactive way to receive the feedback. We are also defining the procedures and protocol to validate the feedback and maintain the system in real-time manner.

4.8. Recommendation

4.8.1. Introduction

Recommendation Manager composed of two major components; Recommendation Classification and Recommendation Explanation. Recommendation Classification component filters the information based on user characteristics and preferences as well as categorizes the recommendations as clinical, non-clinical or combination of them. While Recommendation Explanation component generates explanations upon recommendations by exploiting the internal details of how the decision is made by the system. The explanations are represented for the subject user according to his preferences and context.

4.8.2. Related Work

- Researchers have indicated that integrated medical information systems are becoming an essential part of the modern healthcare systems [Duan2011].
- There is increasing awareness in recommender systems research of the need to make the recommendation process more transparent to users. Explanation provides answers to "How" and "Why" questions. How you come up with this recommendation? And why should I accept it? Explanations in recommender systems can be generally understood as a form of communication between the recommender system and the user [Jannach2010].
- Authors of [Tintarev2007] and [Masthoff2007] discussed the set of characteristics that can be associated with explanations such as transparency, comprehensibility, validity, trustworthiness,

persuasiveness, effectiveness and education. The explanation helps the users not only to comprehend about the recommended task, item or service but also get educated.

- In reasoning model proposed by Friedrich, Gerhard, and Markus Zanker, the authors explained two models; white box and black box. A White-box explanation describes the underlying conceptual model of the recommendation engine, while black-box explanations do not disclose the functioning of the system to the user [Friedrich2011].
- The multi-dimensional knowledgeable explanation is required to generate alongside the recommendation in order to increase the level of satisfaction and belief of users on the system.

4.8.3. Components Description

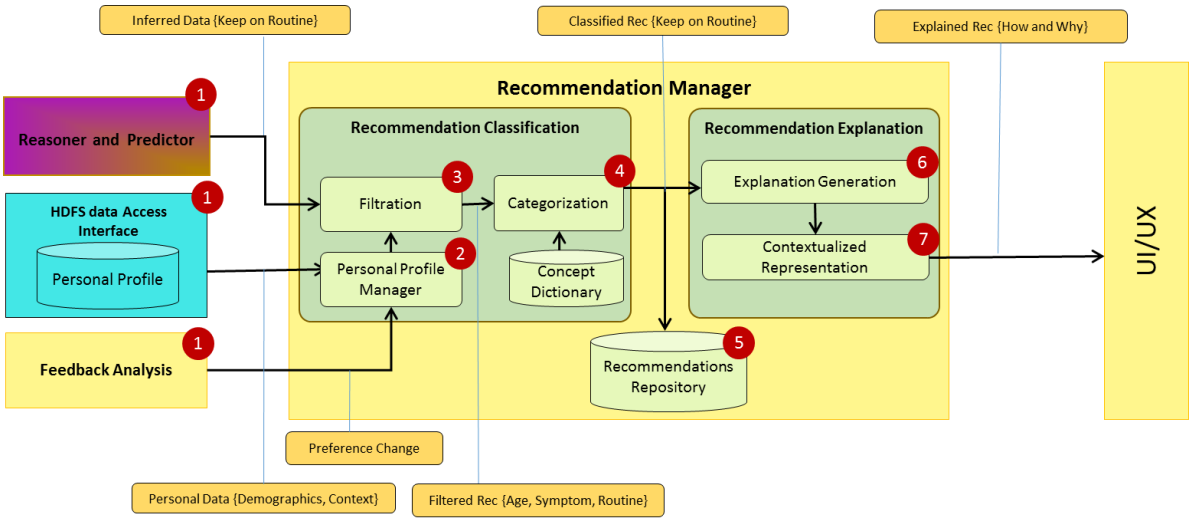


Figure 4.17: Recommendation Manager Component description

- Input: Inferred Data, Personal Data and User Feedback
- Output: Explained Recommendation
- Description: Recommendation Manager composed of two major components; Recommendation Classification and Recommendation Explanation.
- Recommendation Classification component filters the information based on user characteristics and preferences as well as categorization of the recommendation as clinical, non-clinical or combination of them.
- While Recommendation Explanation component generates explanations upon recommendation by exploiting the internal details of how the decision is made by the system. The explanations are represented along with the context customized according to user preferences.

Recommendation Classification

- Input: Inferred Data, Personal Data and User Feedback
- Output: Classified Recommendation
- Description: Recommendations are filtered based on the knowledge of subject (user) and object (recommended item) in order to categorize them into clinical and non-clinical recommendation.
- The Filtration component filters the generated recommendations based on user preferences and context.

- Personal Profile Manager provides input to the filtration and keep updated the personal preferences and interests in personal profile.
- The Categorization component categorizes the generated recommendations as clinical, non-clinical or combination of them using concept dictionary maintained locally.

Recommendation Explanation

- Input: Classified Recommendation
- Output: Explained Recommendation
- Description: Explanations are generated by exploiting the rules in the knowledge base that participated in the decision in order to provide answer to the How and Why questions.
- Explanation Generation generates the explanation as white box and black box explanation in order to answer the questions; "How these recommendations are generated?" and "Why to accept these recommendations?"
- Finally, Contextualized Representation represents the explanations along with the context in graphical form making for users to provide feedback reasonably in easy manner. At the same time, the recommendations are stored for future analysis such as what kinds of recommendations are produced in fast one year.

4.8.4. UML Diagrams

- Sequence Diagram

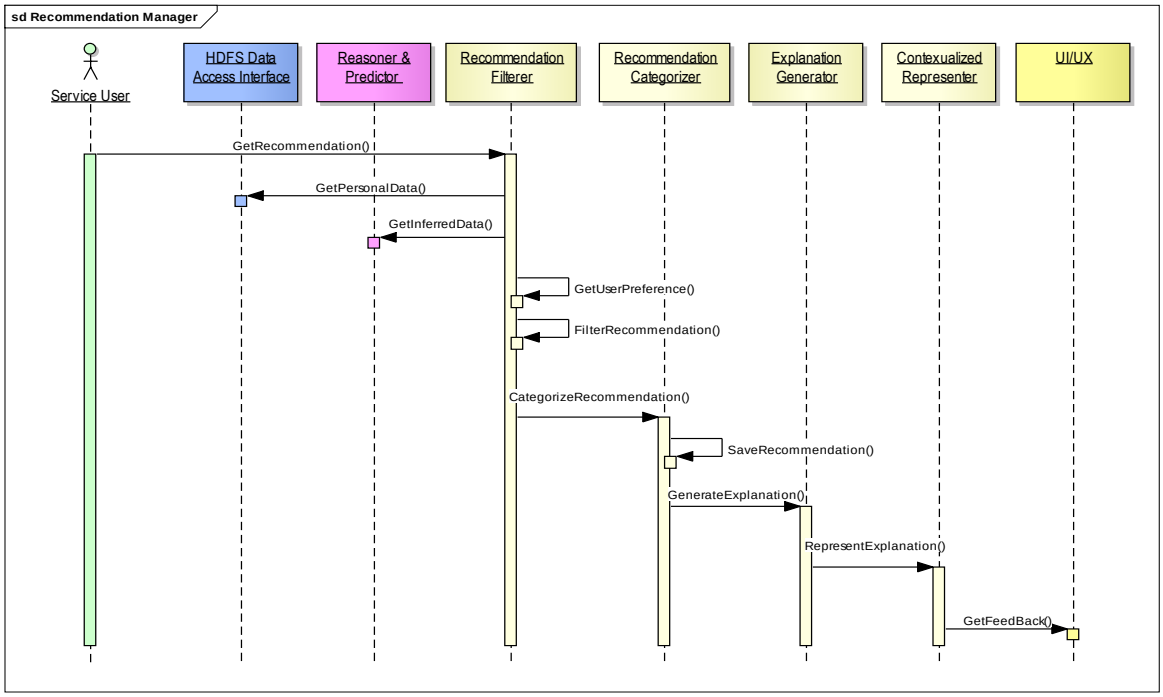


Figure 4.18: Recommendation Manager Sequence Diagram

- Component Diagram

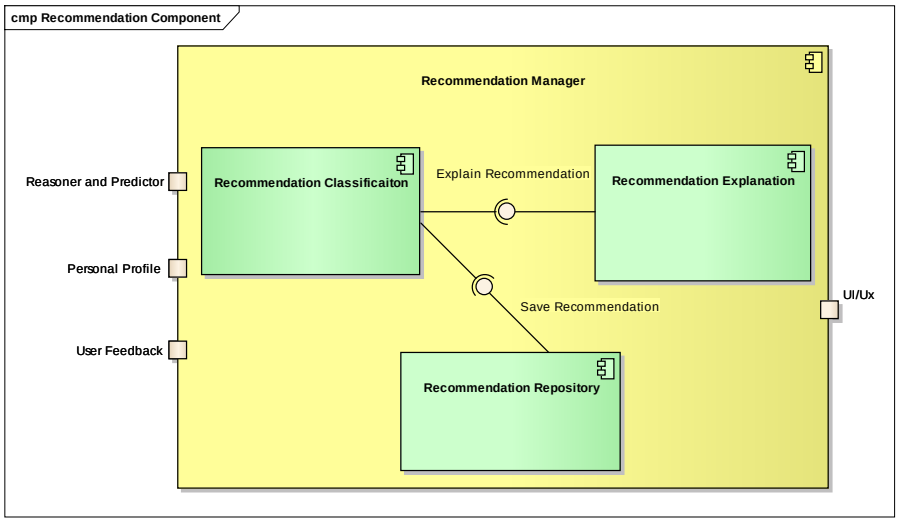


Figure 4.19: Recommendation Manager Component Diagram

4.8.5. As Is – To Be Systems

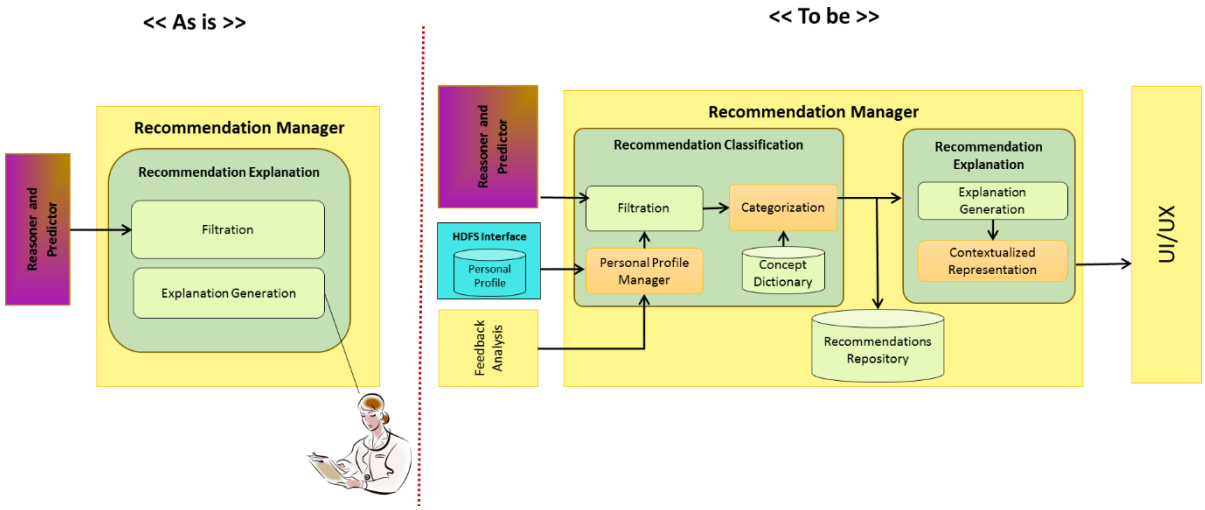


Figure 4.20: Recommendation Manager As Is – To Be

4.8.6. Uniqueness

- Cross domain recommendation with knowledgeable explanations
- Recommendation filtration based on user preferences and context
- Answers to not only "How" question but also the "Why" question
- Contextualized Representation based on the nature of the recommendation

4.8.7. Summary

The systems that show recommendations without explanation depend on the knowledge in the form of rules in knowledge base. While the systems that show recommendations with explanation either exploits the internal logic of the rules participated in the decision. Such kinds of explanation

sometimes become difficult for the users to understand due to logic complexities. Unless users provided with proper representation of the explanation along with the contextual information, it is hard to believe the decision made by the system. We propose a recommendation manager that produces cross-domain recommendation with knowledgeable explanation.

4.9. UX/ UI Authoring Tool

4.9.1. Introduction

The main reason of less adaptation of expert systems is because of the user interfaces. The two main areas that are recently under thorough observation is related to user interface (UI) and user experience (UX). Most of the researchers considers UI and UX as same entities but there exists huge amount of difference between them, although they are related to one another. The main objectives that we are dealing with in this module is the adaptive and personalized approach towards building and managing the user interfaces.

4.9.2. Related Work

- Prior adaptive user interfaces (AUI), including Letizia [Liberman1995] and Let' Browse [Vivacqua1997], they tracked the user behavior patterns and its interest via www browsing. Nowadays, AUI mostly use the computational approaches such as model-based, Bayesian network, and mixed-initiative. He, Jiang, and I-Ling Yen. [Jiang2007] Proposed an AUI generation framework for Web Services that use service oriented approach along with programming language approach and the rule-based approach.
- Peissner, Matthias, and Rob Edlin-White. [Patent2010] developed a MyUI system that made numerous experiments on user in order to examine the effectiveness and acceptability of various UI/UX adaptation patterns throughout the interaction. The UI/UX adaptation patterns were executed before adaptation that decreased the costs of an adaptation by requesting an explicit user confirmation.
- A lot of Performance-Oriented Interface Design Models ware purposed. Ibrahim et al. [Ibrahim2011] and Rozilawati Razali et al. [Matthias2013] considered many factors of UI design that impacts on the performance of web applications based on systematic Literature Review (SLR).they proposed a model which provide guidelines for refining existing web applications in order to used resources in effective manner via its UI.
- Patent [Wu2012] on AUI proposed a method of detecting user left and right handed grip of electronic devices and activated the UI according it.
- In existing AUI/AUX systems, the AUI are not based on context: in which context an application is used and how information is input.in our system along with context, we considered user experience based on many measurement factors such as trust, interaction, reaction, functionality, predictability, and individuality.

4.9.3. Components Description

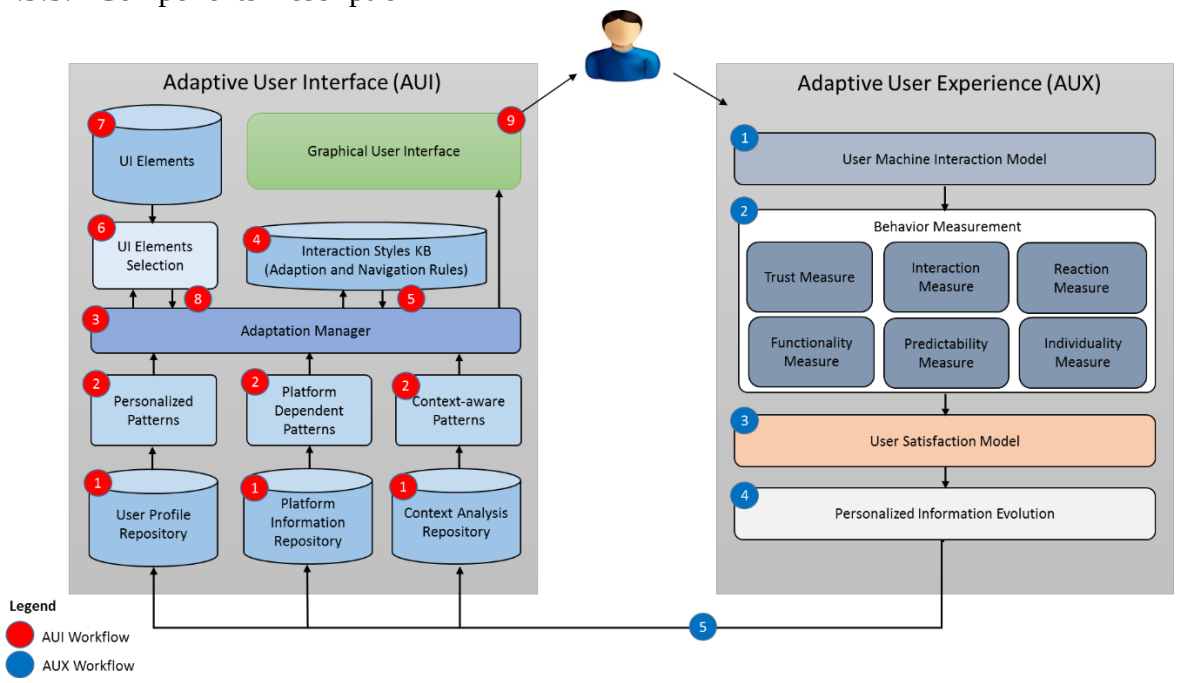


Figure 4.21: UI/ UX Architecture

- Adaptive User Interface (AUI)

Personalized Repositories

- Input: User personalized data
- Output: User preferences, platforms used, and contextual information
- Description: This layer consists of three repositories: User Profile Repository, Platform Information Repository, and Context Analysis Repository. User Profile Repository is used for storing the user preferences in his daily life activities. Platform Information Repository stores the platform related information of the user. Context Analysis Repository stores the current context about the user.

Personalized Patterns

- Input: User preferences, platforms used, and contextual information
- Output: Personalized preferences, platform, and contextual patterns
- Description: This layer consists of three modules: Personalized Patterns, Platform Dependent Patterns, and Context Aware Patterns. Personalized Patterns are based on the preferences information stored in the repository. Platform Dependent Patterns are the patterns derived from the specific platforms and its specifications used by the users. Context Aware Patterns are the patterns based on the current contextual information of the user.

Interactive Styles KB

- Input: Personalized preferences, platform, and contextual patterns
- Output: Individualized Pattern

- Description: This Knowledge Base stores the adaption and navigation rules that are personalized to specific users. These are derived from the personalized patterns. These are used for adaption of the user interface and also navigation between different graphical user interfaces.

UI Elements Selection

- Input: Individualized Pattern
- Output: UI Elements
- Description: This component retrieves the UI elements from the UI Elements repository. This is used by the Adaption Manager for building the graphical user interface.

Adaptation Manager

- Description: This component behaves as a coordinator between different internal components. It takes initial input from the Personalized Patterns layer and forwards it to the Interactive KB. It then takes input from the Interactive KB and provides it to UI Elements Selection. It finally takes output of the UI Elements Selection and builds Graphical User Interface accordingly.

➢ Adaptive User Experience (AUX)

User Machine Interaction Model

- Input: User response
- Output: Feedback
- Description: This component interacts with the user through questionnaire, dialogs, and games to find out the user feedback that will be used for the adaption of the user interfaces.

Behavior Measurement

- Input: User response
- Output: Feedback
- Description: The User Machine Interaction Model is dependent on the Indicator Measures as well. These measures the indicators that can be used for the evaluation of user response after initially seeing the personalized adaptive user interfaces.

User Satisfaction Model

- Input: Feedback
- Output: Personalized Information Categories
- Description: The feedback obtained from the different measurement indicators are then processed for finding out the user satisfaction. A threshold value is set for comparison and degree of adaptation to be carried out in the user interface.

Personalized Information Evolution

- Input: Personalized Information Categories
- Output: Evolution of Personalized Repositories
- Description: This modules takes as input the categories form the user satisfaction model and forwards the information to the Adaptive User Interface for evolution of the personalized repositories.

4.9.4. UML Diagrams

- Sequence Diagram

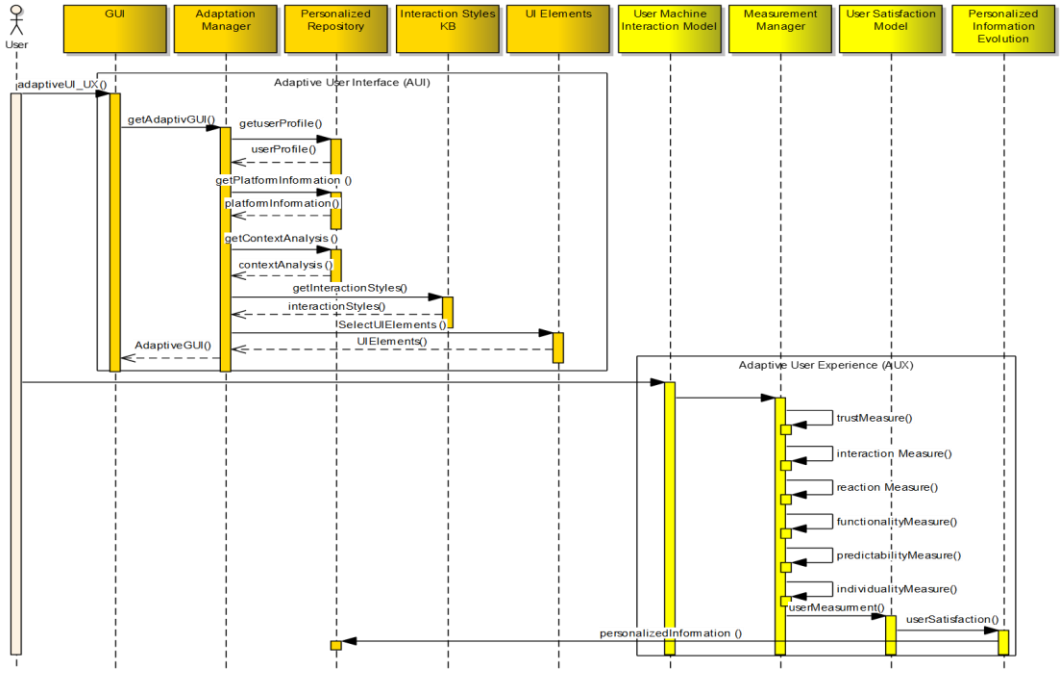


Figure 4.22: AUI/AUX Sequence Diagram

- Component Diagram

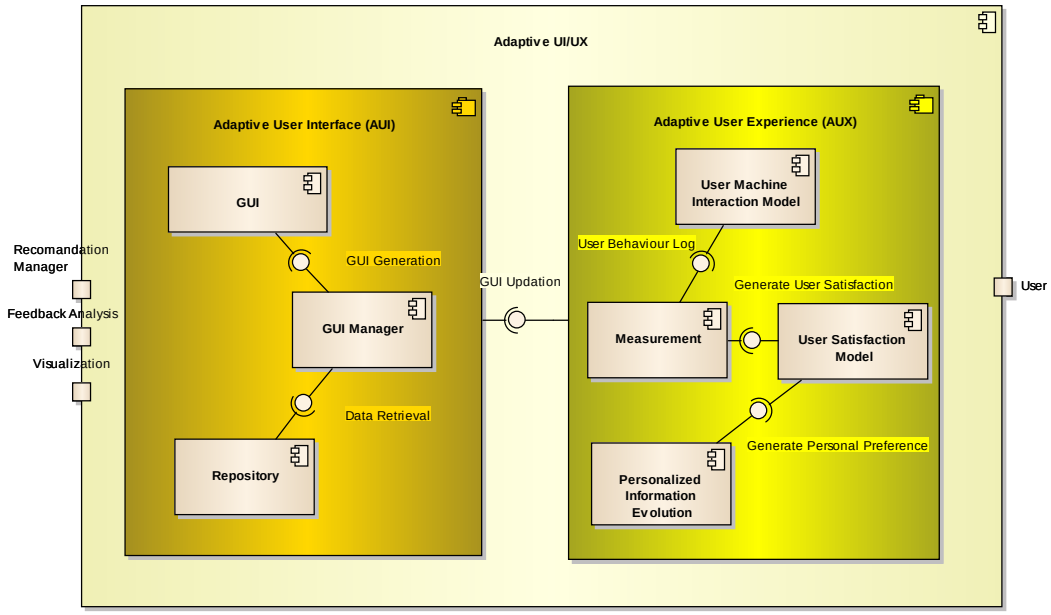


Figure 2.23: AUI/ AUX Component Diagram

4.9.5. UI/ UX Authoring Tool Service Scenario

AUI/ AUX presents personalized user interfaces to the users based on their preferences, platform they use, and contextual information. We describe a scenario showing the personalized user interfaces and its adaption with user experience input to a changed layout and content. A diabetes patient scenario is explained for AUI/ AUX in this section to demonstrate the personalized user interfaces presentation to the user using UI/ UX Authoring Tool.

AUI Workflow

The AUI Workflow is shown in Figure 2.24 and its steps are described as follows:

Step 1: Initially, personalized information is obtained from the user and stored in the repositories. User preferences information is stored in the User Profile Repository; user platform use information is stored in the Platform Information Repository; and user context information is stored in Context Analysis Repository. In the scenario, the user preferences are watching movies, listening to music, playing games, and he is diabetes patient. The user also uses Samsung S3 smartphone which is his platform related information, and currently he is having high glucose level as contextual information. This contextual information is obtained from the Recommendation Manager component of our Mining Minds framework.

Step 2: Personalized patterns are created from the User Profile Repository information. The user likes entertainment because of being fond of movies, games, and music. Also, his health status is Diabetes in combination to the entertainment.

Step 3: The user's platform related information is obtained from the Platform Information Repository. In this case, we suppose that the user is using Samsung S3 smartphone.

Step 4: The current contextual information about the user is obtained from the Context Analysis Repository that is fed by the Recommendation Manager component. The user in the scenario is having high glucose level.

Step 5: The Adaption Manager behaves as a coordinator and concatenates the three kinds of patterns from the previous step modules and forwards it to the Interactive KB, for finding the specific theme or layout for this context based on user profile. The theme in this case should be based on the information of High Glucose Level and displaying the related information on Samsung S3 smartphone.

Step 6: The specific UI Elements are selected from the UI Repository based on the theme selected with the information high glucose level and smartphone in the previous step.

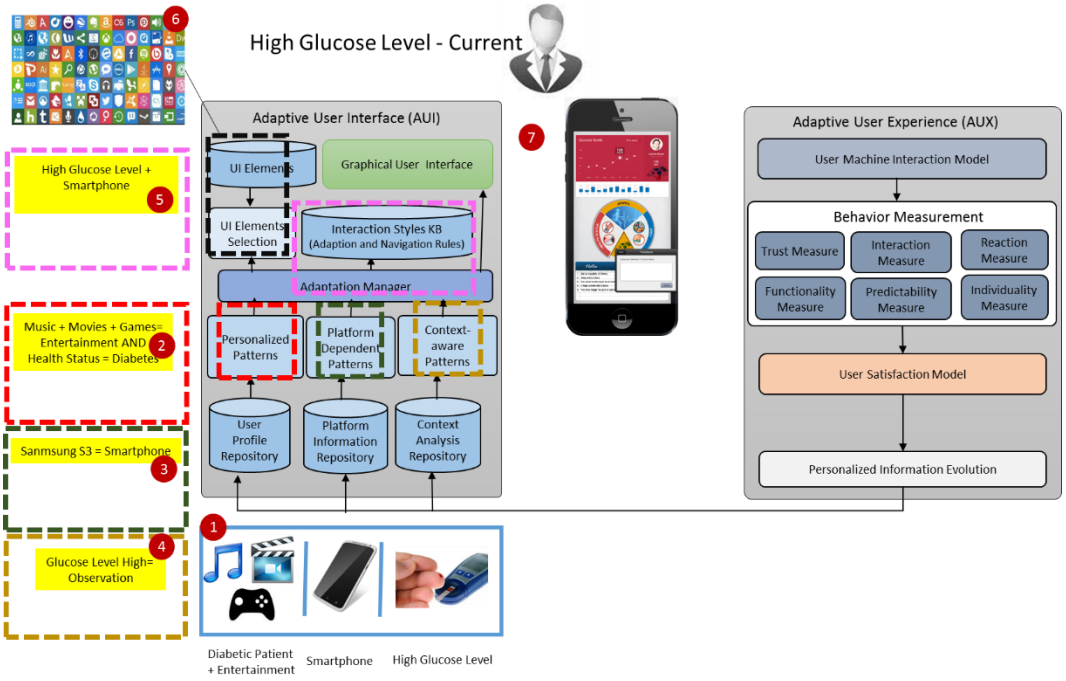


Figure 2.24: AUI Workflow

Step 7: The AUI Workflow output is shown in Figure 2.25. In the figure it can be seen that the information is categorized content and layout. Initially, due to not very much interaction, the user experience is lacking and the layout is specific to the theme. The content are obtained from Visualization and Recommendation Manager Component. Also, the Feedback component is presented to the user, to obtain the use response to be used for user experience. From the visualization part the user glucose information is displayed, showing the user glucose level in the past week and also showing that currently the user is having high glucose value. Based on this visualization, the UI/UX Authoring tool is also fed with recommendations information provided by the Recommendation Manager module. This information is displayed in the Notes section of the AUI. Based on the recommendations some related application are also shown on the AUI, that are related to activities such as Sports, Health Status, and Diet Management. The important thing is to have user experience and therefore feedback module is presented to the user in the user interface.

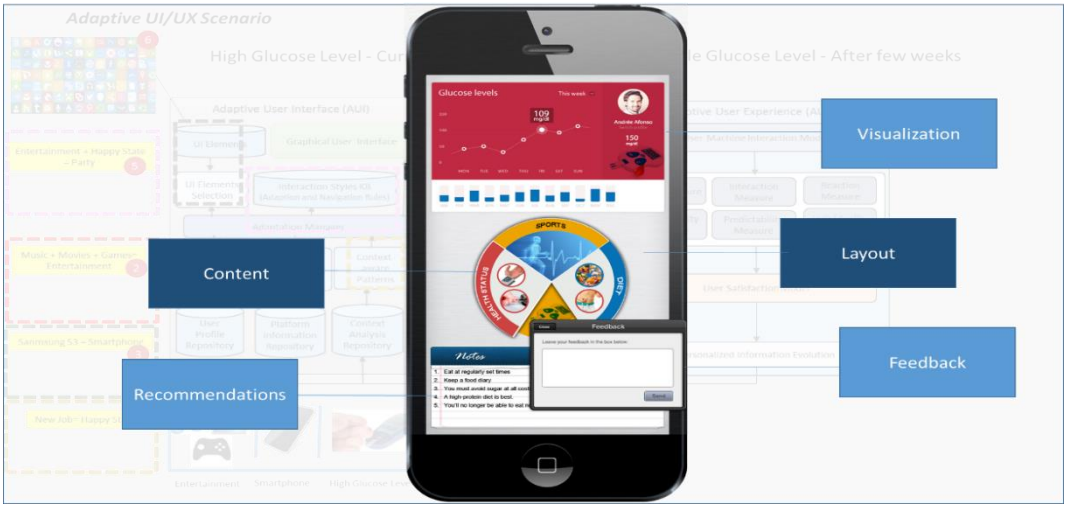


Figure 2.25: AUI Workflow Output

AUX Workflow

Step 8: The user after some time or few weeks finds his/her glucose level to be stable. To get user experience AUX part of the UI/UX Authoring tool is responsible. It interacts with the user with User Machine Interaction Model. The interaction with user is performed using Human Computer Interaction techniques, few of them are questionnaire, dialogs, and interactive games as shown in Figure 2.26.

Step 9: The input of the user using User Machine Interaction model is measured by Behavior Measurement module. There are different metrics associated with the user that includes trust measurement (less application usage), interaction measure (less no clicks), reaction measure (complexity), functionality measure (less features), predictability measure (easy navigation), and individuality measure (color scheme). All this information is obtained and provided to the user satisfaction model.

Step 10: Based on the Behavior Measurement output, the User Satisfaction Model calculates the user satisfaction of the use of our system. If the threshold value is achieved then it means that user is satisfied with the current AUI and based on the information in the repositories carry on with the personalized interface. Otherwise, as is in the scenario, if threshold is not achieved then user is not satisfied and information needs to be categorized and repositories should be evolved. The current information extracted by the behavior measurement module are entertainment is the preference of the user but entertainment stuff was missing in the previous AUI. Also, user has started using Tablet (I-Pad), and color schemes should be changed to purple and blue with more interactive GUI.

Step 11: The entertainment, tablet, and current status of glucose stable information is categorized by the Personalized Information Model. This information is forwarded to the repositories for evolution of the new knowledge.

Step 12: The personalized, platform and context analysis patterns are also updated with the new information of entertainment, I-Pad, and stable glucose level respectively.

Step 13: The color schemes, I-Pad setting, and entertainment information is matched for a particular theme to be selected from the Interactive KB. The final theme is selected and then information is forwarded to the UI Elements Selection module. The UI Elements selection retrieves the UI elements from the UI repository and new adaptive user interface is displayed to the user. The new adaptation is based on the personalized information.

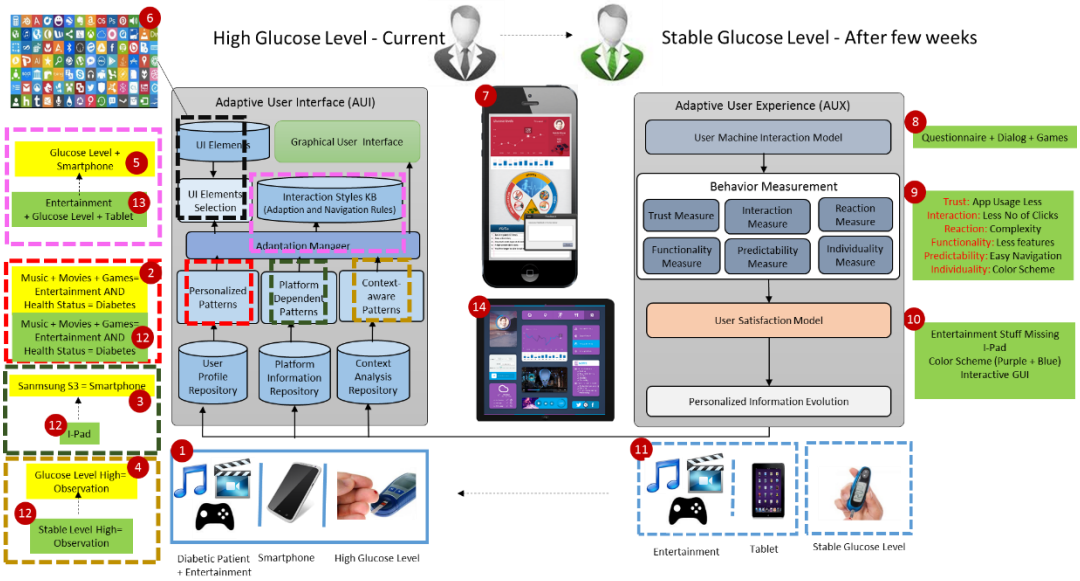


Figure 2.26: AUX Workflow

Step 14: The new AUI based on the AUX output is shown in Figure 2.27. The new navigation style with use preferred color schemes and also adding the entertainment and more features stuff to the user is presented in the AUI.

Figure 2.27: AUX and AUI Workflow Output

4.9.6. As Is – To Be Systems

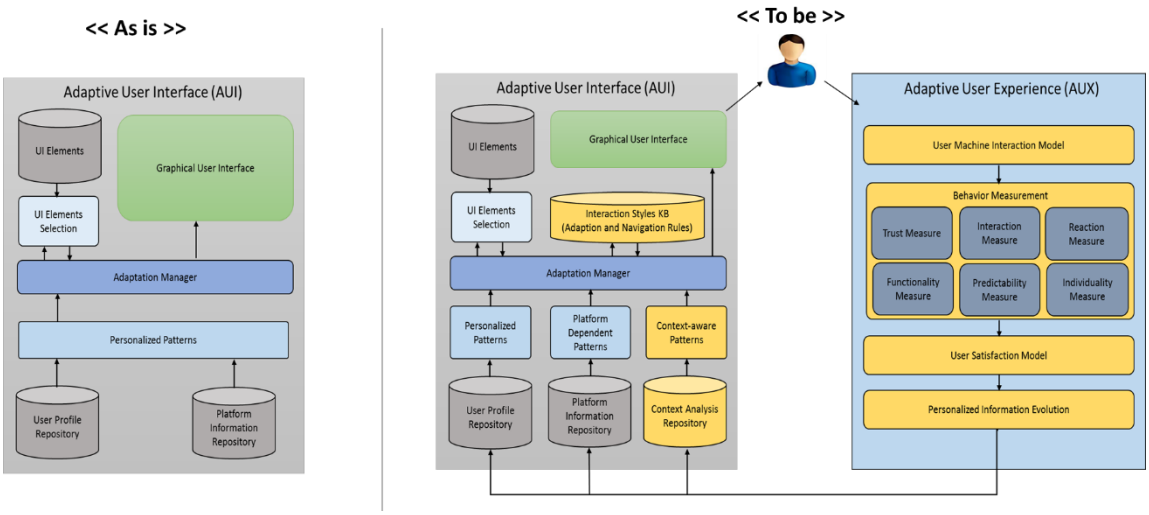


Figure 4.28: UI/UX Authoring Tool As Is – To Be

4.9.7. Uniqueness

- User Experience based on User Machine interaction
- Personalized, Platform and contextual information are combined for AUI
- Long term flexibility by involving user in design
- User satisfaction as a feedback metric
- Continuous evolution of AUI with contextual information change

4.9.8. Summary

AUI/ AUX module takes the personal preferences and contextual information of the user into account for adapting the user interface. User experience controls the evolutionary process of the adaptation of the user interfaces and also maintaining the personalization aspect.

4.10. Service Curation Layer Overall Uniqueness

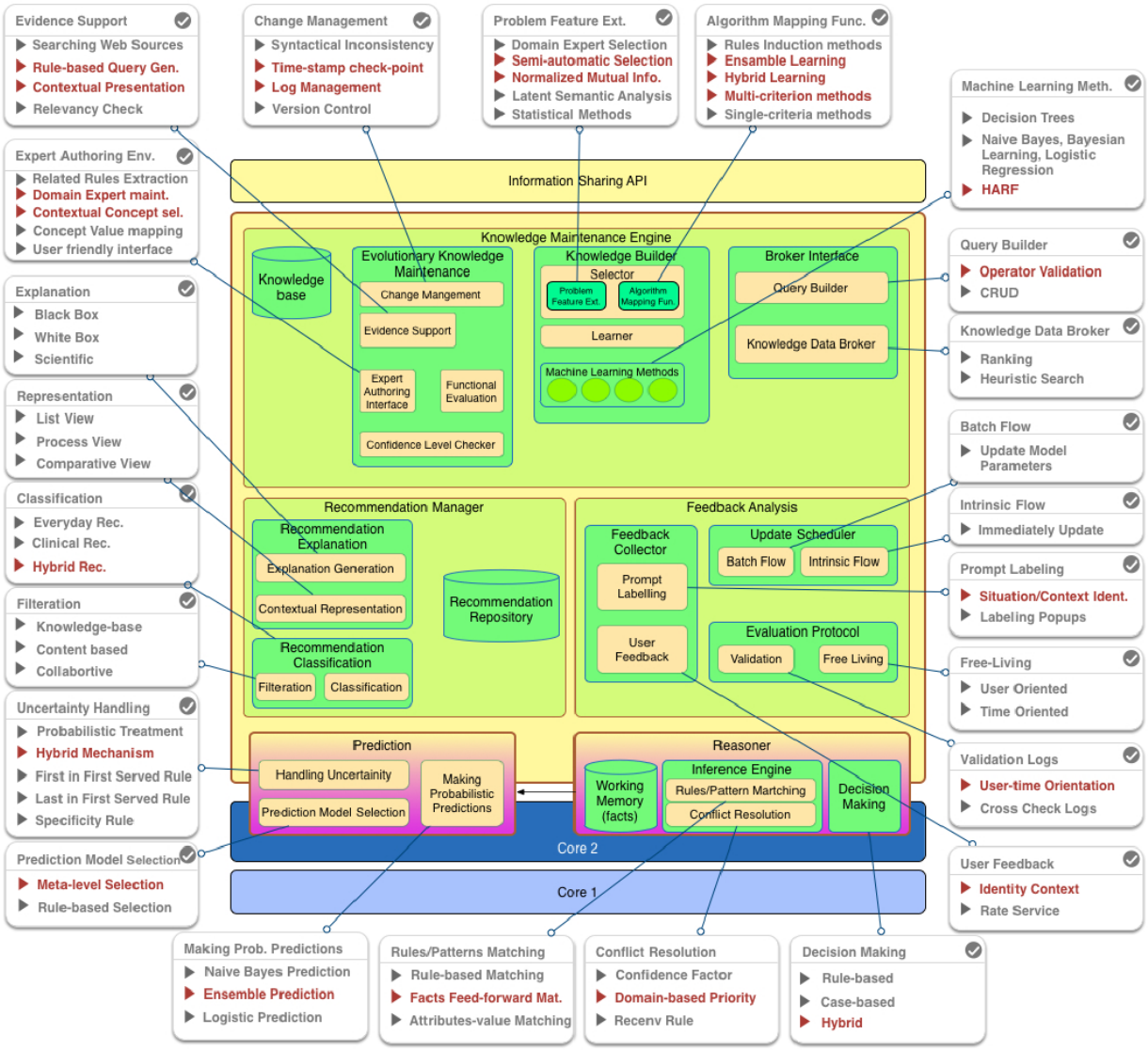


Figure 4.29: Service Curation Layer Uniqueness

Chapter 5
Services Scenarios

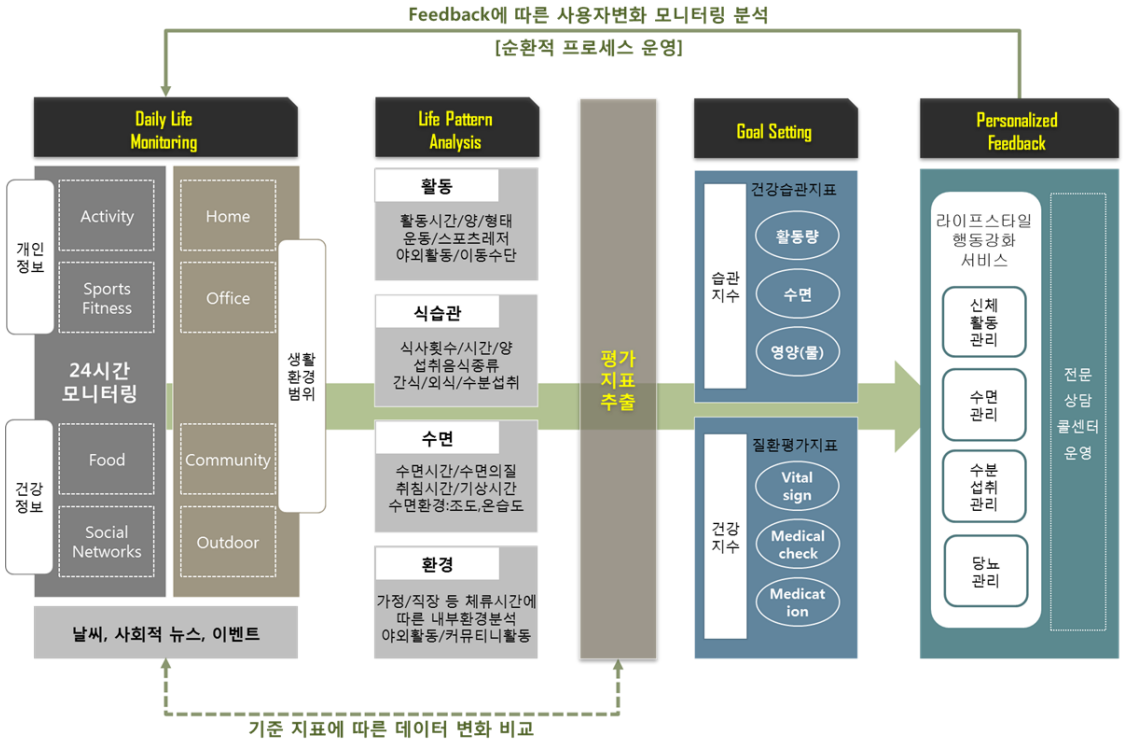
5. 서비스 시나리오

본 과제에서 개발될 연구결과물인 마이닝마인즈 플랫폼을 활용하여 다양한 종류의 개인화된 서비스를 사용자에게 제공할 수 있을 것으로 기대된다. 본 장에서는 과제에서 핵심적으로 추구하는 뉴 라이프 매니지먼트에 기반한 서비스에 대하여 설명하고, 각 서비스 시나리오별 현재 상황 및 추진 방향을 제시하는 As is – To be 전략에 대하여 기술한다.

5.1. 전체 서비스 시나리오

- 본 과제에서 핵심적으로 추구하는 마이닝마인즈 플랫폼을 활용한 서비스는 “4050 을 위한 생활밀착형 습관형성 서비스”는 시니어 진입 전 세대의 라이프 매니지먼트를 통해 만성질환 이환율을 감소시키고 건강한 시니어 라이프를 영위하기 위한 예방적 차원의 서비스 모델로, 서비스 영역별로 기본 목표를 제시하고 빅 데이터를 활용하여 개인별 맞춤 서비스를 제공한다.
    - 다양한 웨어러블 센서로부터 사용자의 행위/행동/환경 정보/생체 데이터, 비디오/오디오 등의 멀티미디어 데이터, SNS/웹 등의 소셜 미디어 데이터 등의 사용자 정보를 수집하여 사용자의 개인정보 및 건강정보를 파악함.
    - 대상자의 특성, 식습관, 활동, 수면, 생활환경, 생활이벤트에 따른 일상 생활 패턴을 분석하고, 그에 따른 평가 지표를 추출하여 기준 지표에 따른 데이터 변화와 비교함
    - 습관지수, 건강지수를 기반으로 개별 사용자에 맞는 건강습관지표 및 질환평가지표 작성과 그에 대한 목표치 설정
    - 활동량 기반 신체활동 관리 서비스, 수면 모니터링을 통한 수면의 질 개선 서비스, 스마트 컵을 활용한 물 섭취 관리 서비스, 당뇨 관리 서비스 등 라이프스타일의 행동을 강화할 수 있는 서비스 제공
    - 사용자의 피드백을 분석, 시스템에 반영하여 개개인을 위한 진화하는 맞춤형 서비스 제공이 가능

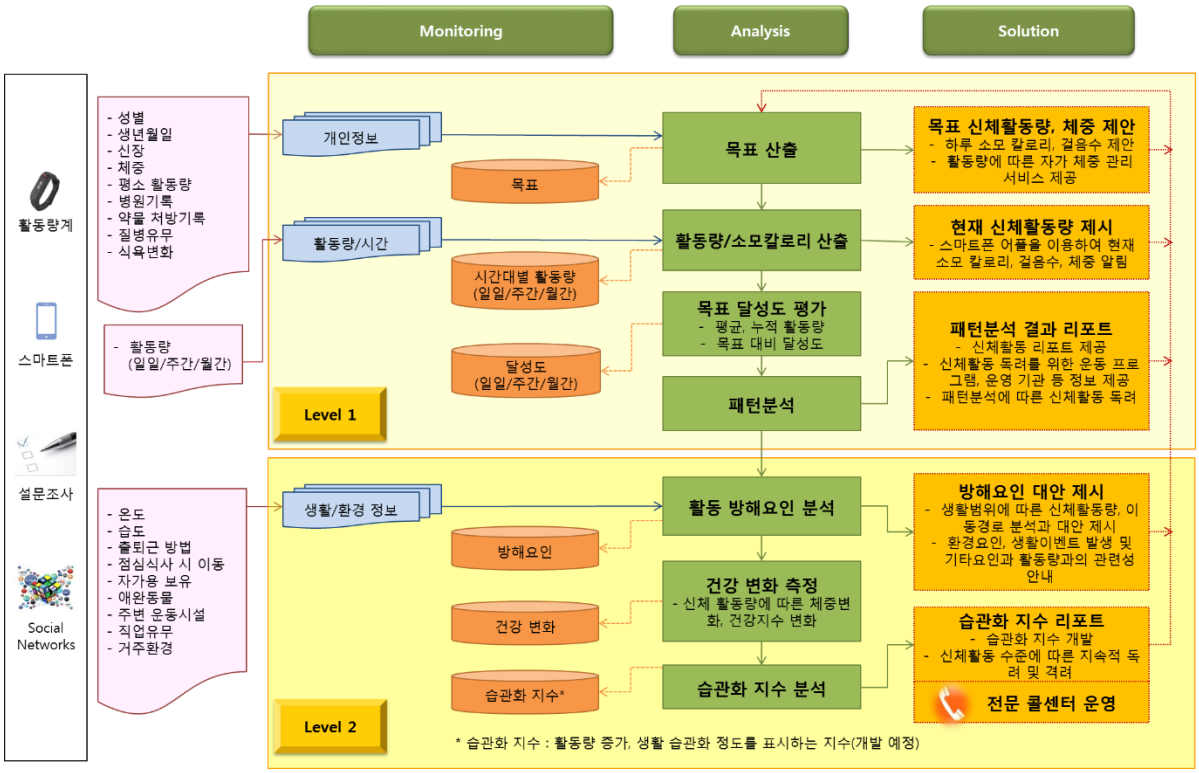
[4050을 위한 생활밀착형 습관형성 서비스]



5.1.1. 신체활동 관리를 위한 Exercise curation 서비스

- 신체활동패턴 분석에 따른 신체활동 증진 목표 및 실천 가이드라인 제시를 통해 자가관리 서비스를 구현하며, 다음은 생활환경 기반 개인 맞춤형 서비스의 리스트이다.
    - 목표 신체활동량 제시, 신체활동량 및 소모 칼로리 모니터링
    - 건강위험군, 만성질환자의 경우 위험도에 따른 운동처방 및 관리 솔루션서비스
    - 빅 데이터를 통해 신체활동에 방해 요소를 찾아내어 습관을 개선해 나갈 수 있도록 맞춤형 서비스 제공
    - 생활 환경(교통수단 및 활동 패턴 분석 등)에 따른 활동량 증대 방안 서비스
    - 운동 종목, 운동 강도, 운동 방법 등을 모니터 해서 정확한 동작, 효과적이고 올바른 방법 안내 서비스
    - 개별 성향(취미, 활동정도 등)에 따른 맞춤 서비스: 스포츠 및 여가활동 프로그램을 운영하는 기관, 문화센터, 피트니스 센터 등의 추천 및 운영 정보 제공

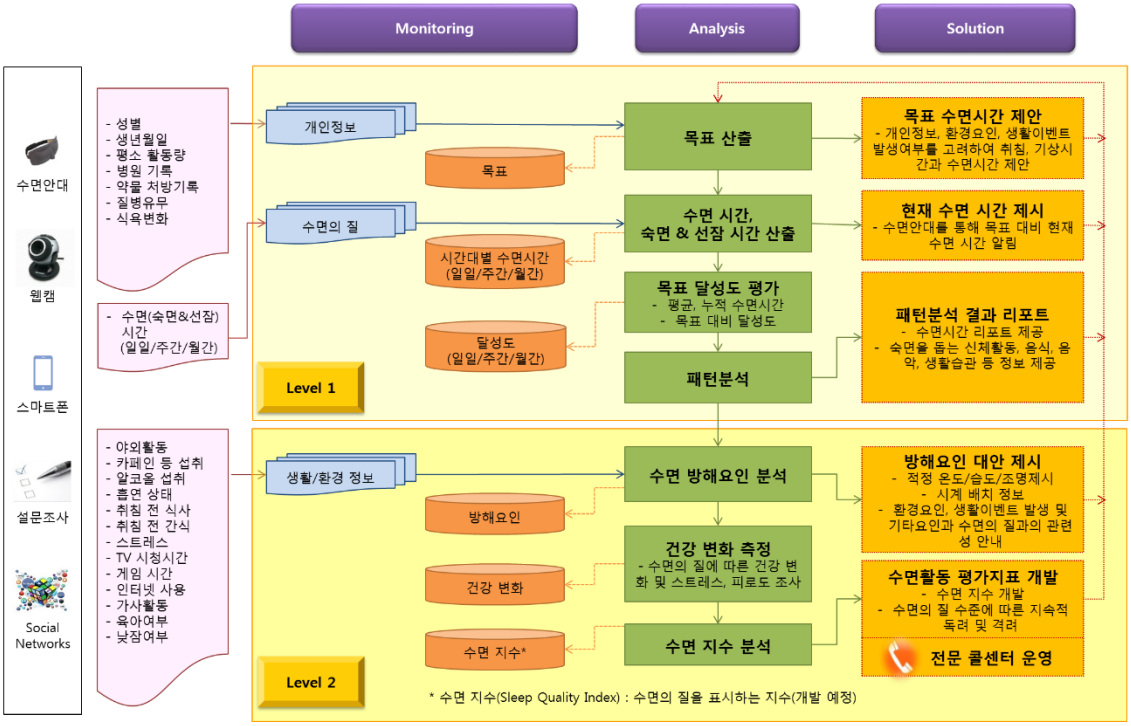
4050 Healthy Life Service Model – 신체 활동 관리를 위한 『Exercise Curation』



5.1.2. 수면관리를 위한 Quality of Sleep Curation 서비스

- 수면패턴 분석에 따른 최적화된 수면 계획 목표 및 실천 가이드라인 제시를 통해 수면의 질을 개선하기 위한 서비스이며, 다음과 같은 생활환경 기반 개인 맞춤형 서비스를 제공함
    - 수면 상태(시간, 수면 단계, 코골이, 무호흡 등), 수면 환경(조도, 온도, 습도, 소음 등), 생활 환경 등을 모니터링
    - 빅 데이터 분석과 개인 데이터를 기반한 수면 방해 요소 분석
    - 개인 맞춤형 수면 개선 프로그램 제시

4050 Healthy Life Service Model – 수면 섭취 관리를 위한『Quality of Sleep Curation 』

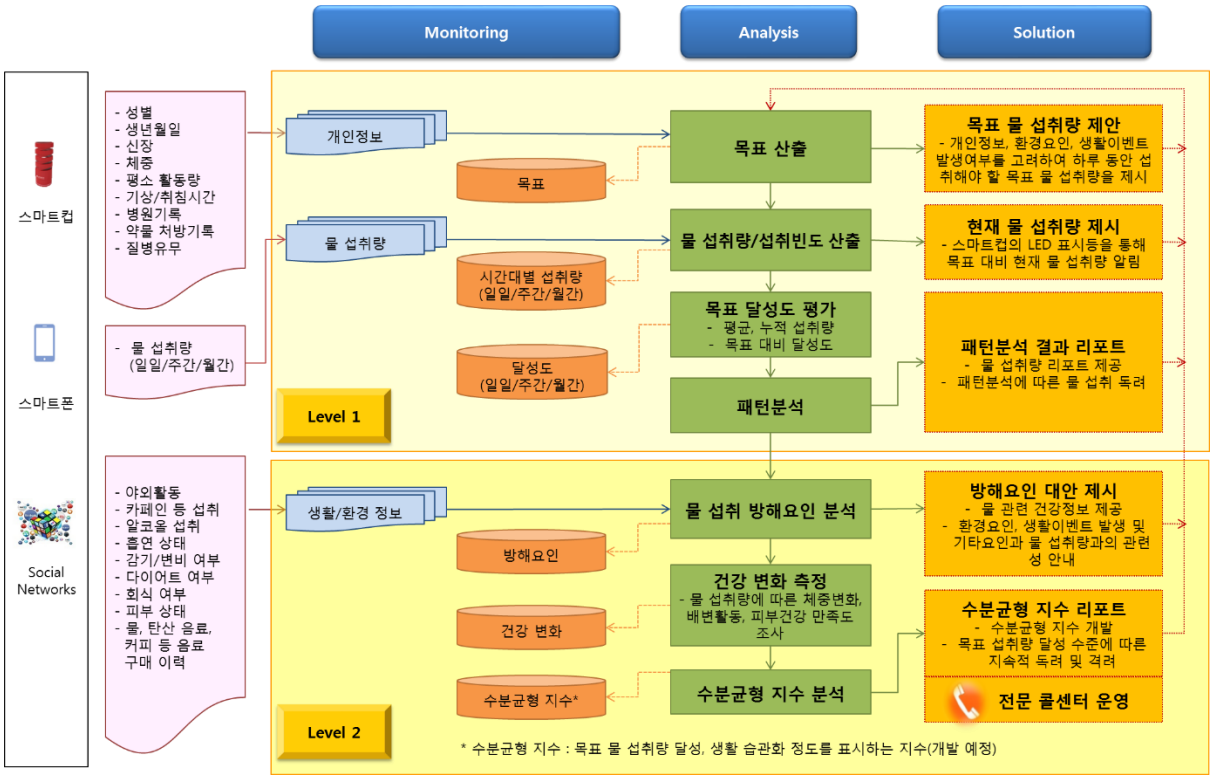


5.1.3. 물 섭취를 위한 Drinking Water Curation 서비스

• 물 섭취에 관한 관리 목표 및 실천 가이드라인 제시를 통해 자가관리서비스를 구현하며, 다음과 같은 생활환경 기반 개인 맞춤형 서비스를 제공함

- 물 섭취량 모니터링 및 섭취 패턴 평가
- 환경 요인 모니터링
- 생활 환경 및 이벤트 발생 모니터링
- 개인별 신체, 질병, 생활 패턴 분석
- 개인 맞춤형 물 섭취 관리 서비스 제공

4050 Healthy Life Service Model – 물 섭취 관리를 위한『Drinking Water Curation 』



5.2. 서비스 시나리오별 As is – To be 전략

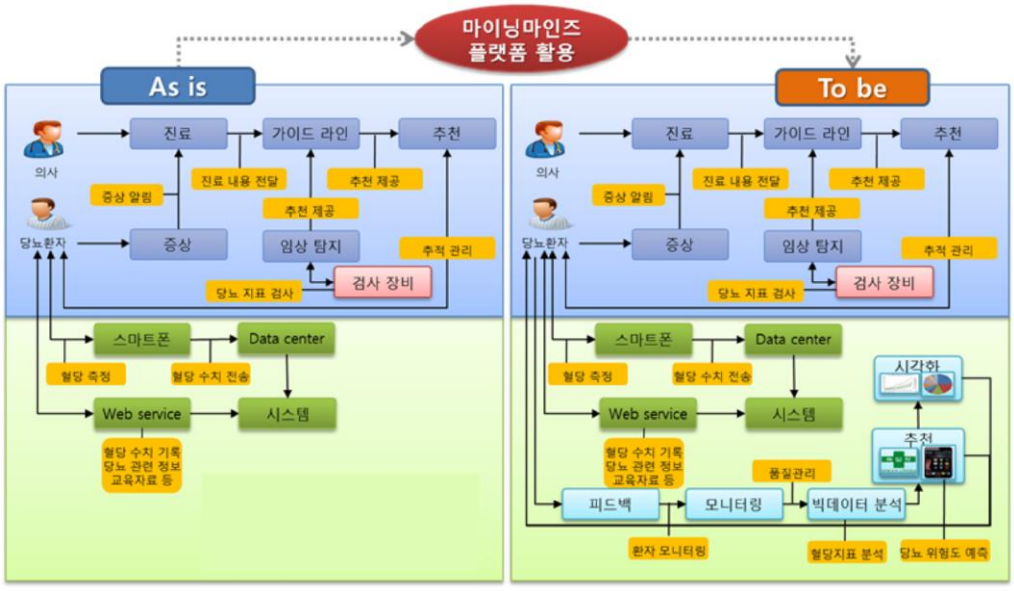
5.2.1. 활동량 기반 4050 Healthy life management 서비스

- 사용자의 일상 생활을 기반으로 퍼스널 라이프 이벤트 데이터를 수집하고, 마이닝마인즈 엔진을 통해 사용자의 생활 패턴을 추출하여 그에 맞는 맞춤형 서비스를 제공한다.

- As Is
  - 정보 전달 수준이 낮아 사용자가 직접 선택을 해야 함
  - 개별화 서비스 제공으로 독립된 서비스를 추구함
  - 단방향 서비스 제공으로 정보 공유 체계가 미흡함
- To Be
  - 예측 서비스 제공으로 데이터 상관관계 분석을 통해 건강 위험모델 수립이 가능함
  - 사용자 맞춤형 서비스 제공으로 개인 맞춤형 건강관리 콘텐츠를 제공함
  - 활동량계, 스마트 컵, 수면 모니터링 등 통합된 서비스 제공이 가능함

5.2.2. 당뇨관리 서비스

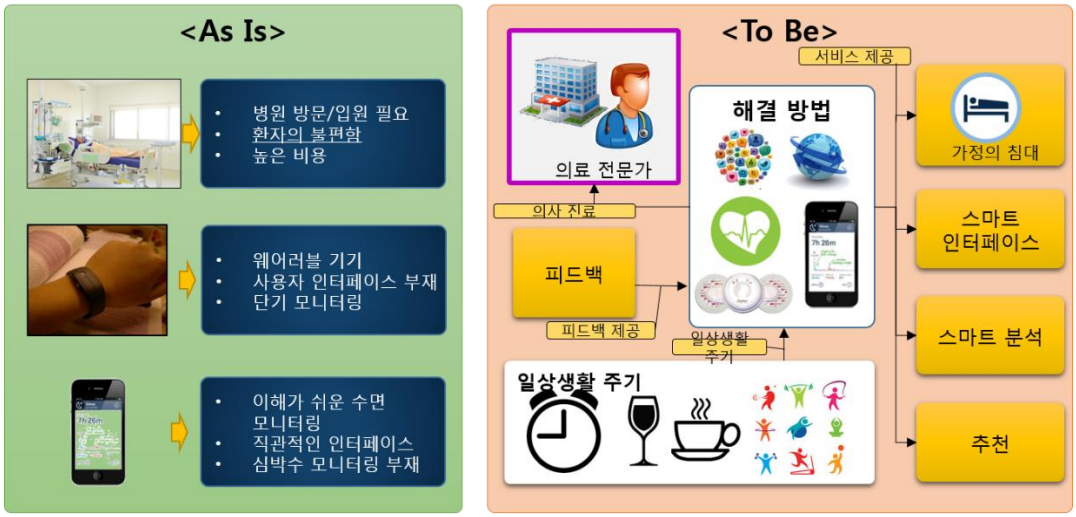
- 기존의 당뇨환자 관리 서비스는 수동적이며 일관적인 서비스가 제공되었으나, 마이닝마인즈 플랫폼을 활용하여 능동적인 참여시스템으로 변화시킬 수 있다.



- As Is
  - Data 관리 중심의 정보제공서비스
  - 수동적인 모니터링
  - 대체 서비스 추천 기능 부재
  - 개인화된 맞춤 추천 부재
- To Be
  - 당뇨 위험도 예측에 따른 생활습관관리(운동, 영양) 서비스
  - 서비스의 질 향상 및 모니터링, 관리 및 피드백 기능
  - 사용자 맞춤형 UI/UX 제공 및 이해도 높이는 시각화 기능
  - 개인화된 서비스 제공 및 대체 서비스 추천 가능

5.2.3. 수면 모니터링 서비스

- 기존의 수면모니터링 서비스는 병원 방문의 불편함, 단순 수면 상태만을 관찰하였으나, 마이닝마인즈 플랫폼을 활용하여 일상 생활에서 모니터링이 가능하고, 일상 생활 정보 반영으로 정확도를 향상시킬 수 있다.



- As Is
  - 수면 전과 후의 일상생활 주기를 알 수 없음.
  - 스마트 인터페이스의 부재
  - 수준 낮은 서비스
- To Be
  - 수면 전과 후의 상황정보 획득 가능
  - 스마트 인터페이스 기능
  - 피드백을 통한 서비스의 질 향상 가능
  - 수면 패턴을 가시화하기 위한 스마트 분석 기능

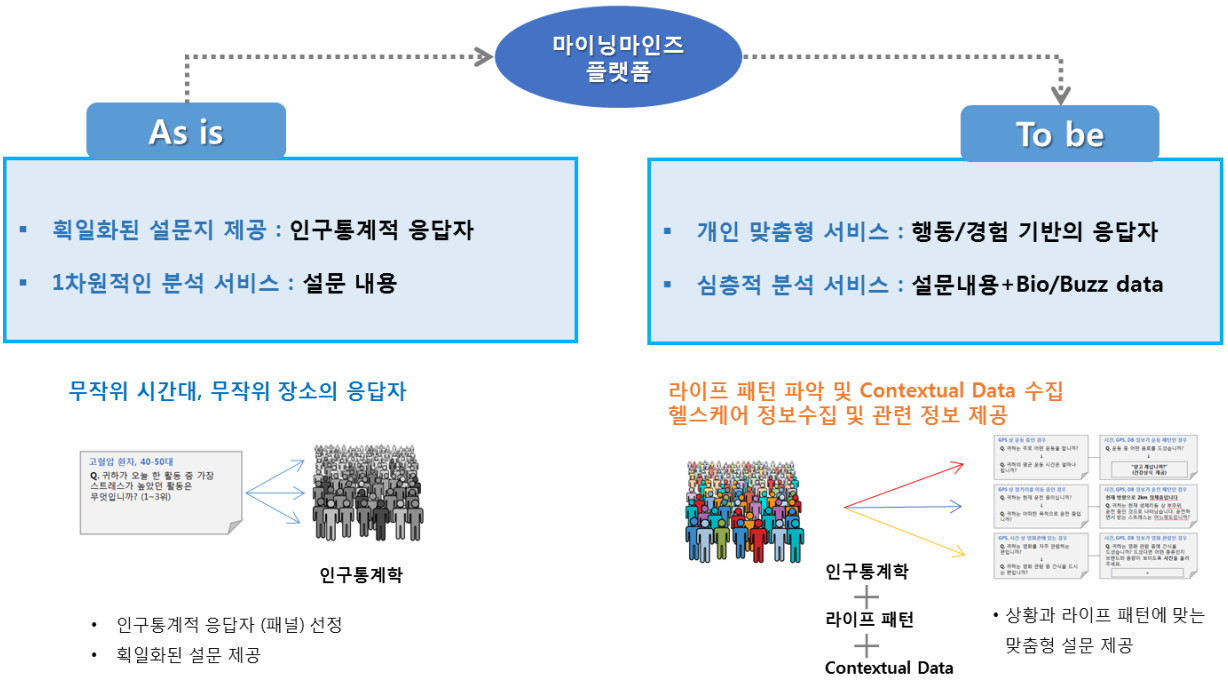
5.2.4. 물 섭취 관리 서비스

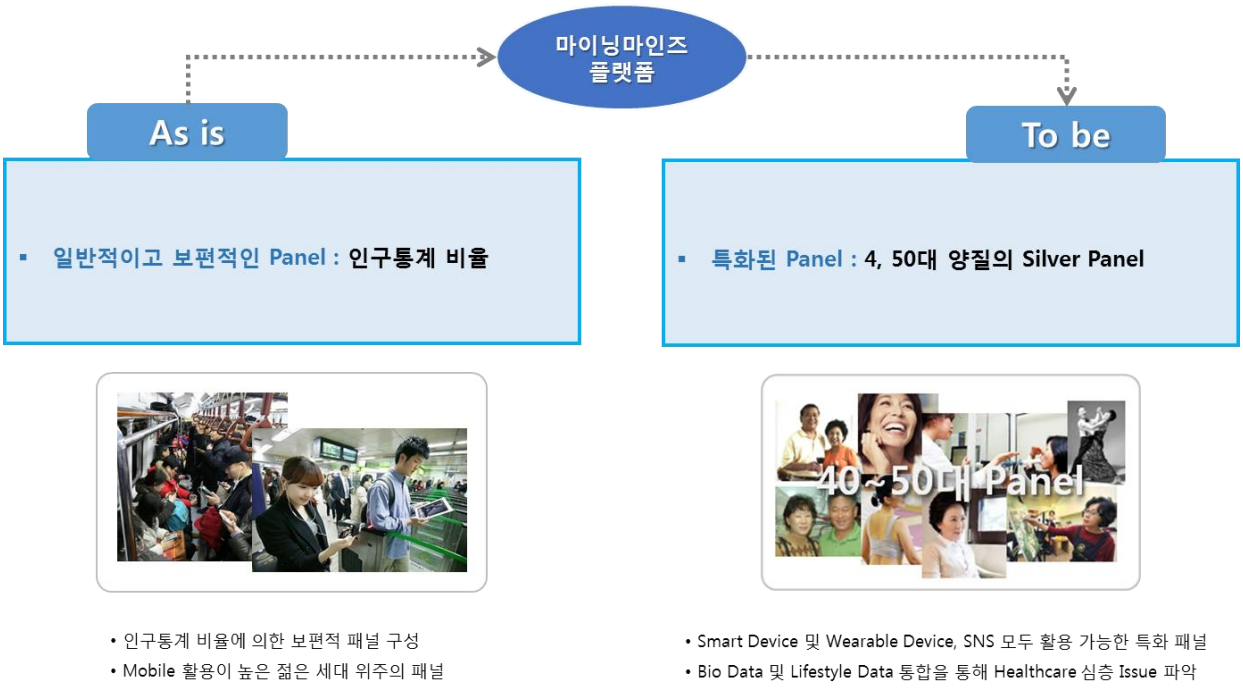
- 기존의 물 섭취 관리 서비스는 사용자가 수동적으로 실시할 수 밖에 없었으나, 마이닝 마인즈 플랫폼을 활용하여 사용자에게 자동으로 맞춤형 서비스의 제공이 가능하다.

- As Is
  - 스마트폰 앱을 통한 단순한 물 섭취량 입력 수준
  - 목표 물 섭취량에 대한 수동적인 모니터링 방법
- To Be
  - 생활 환경범위에 따른 목표 섭취량 제공
  - 생활이벤트 발생에 따른 목표 섭취량 제공
  - 전문가 모니터링 서비스 제공

5.2.5. Survey platform

- 기존의 기술과 마이닝마인즈 플랫폼을 활용하여 개발될 mobile survey open platform 을 기반으로 survey authoring tool 을 고도화하여 라이프매니지먼트 영역 외 모바일 광고, Political Campaign 등의 분야로 사업 확장이 가능하다.

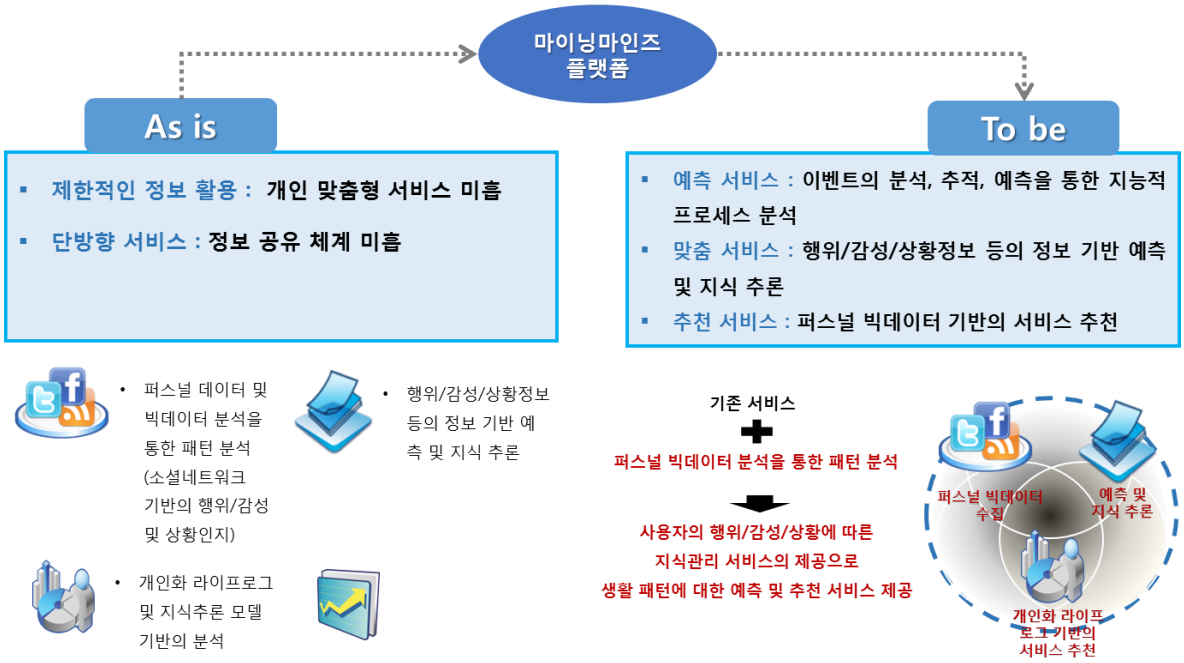




- As Is
    - 획일화된 설문지를 제공하고, 이를 기반으로 인구통계적 응답자를 선정함
    - 설문 내용에 따라 1 차원적인 분석 서비스가 이루어짐
    - 인구통계 비율에 따른 일반적이고 보편적인 Panel 로 구성됨
- To Be
    - 행동/경험 기반의 응답자에 따른 개인 맞춤형 서비스를 제공함
    - 설문 내용과 Bio/Buzz Data 결합을 통한 심층적인 분석 서비스를 제공함
    - 40~50 대를 위한 양질의 특화된 Silver Panel 을 제공함

5.2.6. SNS 기반 빅 데이터 분석

- 지식관리 엔진을 활용한 서비스 타겟 커스터마이징 및 보유중인 SNS 분석 엔진과 마이닝마인즈 엔진을 통합하여 사용자 개개인 맞춤형 예측 및 추천이 가능한 서비스 제공이 가능하다.



- As Is
  - 제한적인 정보 활용으로 인한 개인 맞춤형 서비스 미흡
  - 단반향 서비스로 정보 공유 체계 미흡
- To Be
  - 이벤트의 분석, 추적, 예측 기반 지능적 프로세스를 통한 예측 서비스 제공
  - 행위/감성/상황정보 등의 정보 기반 예측 및 지식 추론으로 맞춤 서비스 제공
  - 퍼스널 빅 데이터 기반의 서비스 추천 제공

Chapter 6
Conclusion

Conclusion

Mining Minds platform provides personalized services to users and it is based on high quality of content and high quality of services. It analyses the user's behaviors to provide customized services to the users. The quality of contents and presentation are two important aspects to use the system for a long duration. If the quality of content is rich but poorly presented to the users', then there is very low probability of system adoption in daily routines and vice versa. This objective can be achieved by providing virtual personal assistant to support person in better life-care management.

The proposed platform maintains the quality of content and provide quality of services as a virtual personal assistant or coach. Therefore, we introduce the layered architecture to support real time data streaming and processing. We separate the data, information, and service curation layers. Data curation layer deals with the real time data streaming and processing and store into the Hadoop based file system. Furthermore, intermediate data is generated to provide fast access and structured data when it is required by information and service curation layer. Information curation layer extracts the low-level contexts and high-level contexts for human's life modeling. Service curation layer consists of Reasoning and prediction to provide the generalized recommendation. These recommendations are further refined in recommendation manager to compliance with the user's profile. We are collecting the feedback by the user and maintaining through knowledge maintenance engine.

To meet high quality of content requires appropriate reasoning, feedback analysis, and knowledge maintenance engine. The updated personal and contextual knowledge keep on updating in the knowledge bases for processing and providing new personalized services to the users. In the same way, for high quality of services UI/UX Authoring tool is used to present visualization, recommendation and feedback to the user in a more personalized user interface environment. Feedback and web monitoring is used to keep the knowledge in the knowledge bases updated and take user experience as input for adaptive user interfaces.

High quality of contents and personalized services are generated through our proposed platform. The scope of the services includes personalized life-quality improving services, silver business services, life-care services, and avail the proactive approach to control the chronic disease as well as active lifestyle.

References

<Data Streaming and Communication>

[Zahria2010] Zaharia, Matei, et al. "Spark: cluster computing with working sets." Proceedings of the 2nd USENIX conference on Hot topics in cloud computing. 2010.

[Shahrivari2014] Shahrivari, Saeed. "Beyond Batch Processing: Towards Real-Time and Streaming Big Data." arXiv preprint arXiv:1403.3375 (2014).

[Lam2012] Lam, Wang, et al. "Muppet: MapReduce-style processing of fast data." Proceedings of the VLDB Endowment 5.12 (2012): 1814-1825.

[Pallickara2009] Pallickara, Shrideep, Jaliya Ekanayake, and Geoffrey Fox. "Granules: A lightweight, streaming runtime for cloud computing with support, for map-reduce." Cluster Computing and Workshops, 2009. CLUSTER'09. IEEE International Conference on. IEEE, 2009.

[Zinn2010] Zinn, Daniel, et al. "Parallelizing XML data-streaming workflows via MapReduce." Journal of Computer and System Sciences 76.6 (2010): 447-463.

[Ding2011] Ding, Mengwei, et al. "More convenient more overhead: the performance evaluation of Hadoop streaming." Proceedings of the 2011 ACM Symposium on Research in Applied Computation. ACM, 2011.

<Data Representation and Mapping>

[Chen2012] L. Chen, C. Nugent, H. Wang, "A Knowledge-Driven Approach to Activity Recognition in Smart Homes," Knowledge and Data Engineering, IEEE Transactions on , vol.PP, no.99, pp.1, 0, 2012.

[Khattak2011] A. M. Khattak, P. T. H. Truc, L. X. Hung, L. T. Vinh, V. H. Dang, D. Guan, Z. Pervez, M. H. Han, S. Y. Lee and Y. K. Lee, "Towards Smart Homes Using Low Level Sensory Data", Journal of Sensors (SCIE, IF 1.77), ISSN: 1424-8220, Vol. 11, No. 12, pp.11581-11604, 2011

[Hussain2013] M. Hussain, A. M. Khattak, W. A. Khan, I. Fatima, M. B. Amin, Z. Pervez, R. Batool, M. A. Saleem, M. Afzal, M. Faheem, M. H. Saddiqi, S. Y. Lee and K. Latif, "Cloud-based Smart CDSS for chronic diseases", Health and Technology (non-SCI), 2013

[Petersen2005] A. K. Petersen and M. Mikalsen. "Context: Representation and reasoning." Revue d'Intelligence Artificielle on Applying Context-Management (2005).

[Ertel2011] W. Ertel, "First-order Predicate Logic." Introduction to Artificial Intelligence. 31-55, Springer London, 2011

<Data Curation>

[Wang06] Wang. Y, Sunderraman, R. “PDB Data Curation,” Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE

[Mark07] Mark Hedges, Adil Hasan, Tobais Blanke, “Curation and Preservation of Research Data in an iRODS Data Grid,” Third IEEE International Conference on e-Science and Grid Computing 2007

[Sophia11] Sophia Zaimidou, Sjozefvan Baal, Timothy D. Smith, Konstantinos Mitropoulos, Mila Ljujic, Dragica Radojkovic, Richard G. Cotton, and George P. Patrinos “Development of a universal, flexible and freely available database management system for gene-centered data collection, curation and display of DNA variation”, 2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).

[Andrew09] Andrew F. Hart, Chris A. Mattmann, John J. Tran, Daniel J. Crichton, J. Steven Hughes, Heather Kincaid, Sean Kelly, "Enabling Effective Curation of Cancer Biomarker Research Data", 22nd IEEE International Symposium of Computer-Based Medical Systems, 2009. CBMS 2009.

[Cecile11] Cecile Germain-Renaud, Frederic Furst, Michel Jouvin, Gilles Kassel, Julien Nauroy, Guillaume Philippon, "The Green Computing Observatory: a data curation approach for green IT", 2011 Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing.

<Visualization>

[Glatz2014] Glatz, Eduard, et al. "Visualizing big network traffic data using frequent pattern mining and hypergraphs." Computing 96.1 (2014): 27-38.

[Gehlenborg2010] Gehlenborg, Nils, et al. "Visualization of omics data for systems biology." Nature methods 7 (2010): S56-S68.

[Bennett2009] Bennett, Jeremy S., "Massive model visualization: An investigation into spatial partitioning" (2009). Graduate Theses and Dissertations. Paper 10546.

[Dancy2008] Elizabeth Dancy, Uliyana Markova "Autonomic computing in Canadian academia"

[Dietrich2007] Dietrich, Andreas, Enrico Gobbetti, and Sung-Eui Yoon. "Massive-model rendering techniques: a tutorial." IEEE Computer Graphics and Applications27.6 (2007): 20-34.

[Shi2004] Shi, Rui, et al. "An adaptive image content representation and segmentation approach to automatic image annotation." Image and Video Retrieval. Springer Berlin Heidelberg, 2004. 545-554.

[McAdams2000] McAdams, Daniel A., and Kristin L. Wood. "Tuning parameter tolerance design: foundations, methods, and measures." Research in Engineering Design 12.3 (2000): 152-162.

<Security and Privacy>

[Zeeshan2013] Zeeshan Pervez, Ammar Ahmad Awan, Asad Masood Khattak, Sungyoung Lee, and Eui-Nam Huh. Privacy-aware searching with oblivious term matching for cloud storage. The Journal of Supercomputing, 63(2):538–560, 2013.

[Zeeshan2012] Zeeshan Pervez, Asad Masood Khattak, Sungyoung Lee, Young-Koo Lee, and Eui- Nam Huh. Oblivious access control policies for cloud based data sharing systems.Computing, 94(12):915–938, 2012.

[Mahmood2013] Mahmood Ahmad, Zeeshan Pervez, YongIK Yoon, Byeong Ho Kang, and Sungy-oung Lee. Reflection: A lightweight protocol for private matching. In Signal-ImageTechnology & Internet-Based Systems (SITIS), 2013 International Conference on,pages 673–678. IEEE, 2013.

[Michel2005] Michel Abdalla, Mihir Bellare, Dario Catalano, Eike Kiltz, Tadayoshi Kohno, Tanja Lange, John Malone-Lee, Gregory Neven, Pascal Paillier, and Haixia Shi. Searchable encryption revisited: Consistency properties, relation to anonymous ibe, and extensions. In Advances in Cryptology–CRYPTO 2005, pages 205–222. Springer,2005.

[Mihir2007] Mihir Bellare, Alexandra Boldyreva, and Adam ONeill. Deterministic and efficiently searchable encryption. In Advances in Cryptology-CRYPTO 2007, pages 535–552. Springer, 2007.

[Louis2006] Louis Kruger, Somesh Jha, Eu-Jin Goh, and Dan Boneh. Secure function evaluation with ordered binary decision diagrams. In Proceedings of the 13th ACM conference on Computer and communications security, pages 410–420. ACM, 2006.

[Roberto2005] Roberto J Bayardo and Rakesh Agrawal. Data privacy through optimal k-anonymization. In Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on, pages 217–228. IEEE, 2005.

[Ashwin2007] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1):3, 2007.

<Low-level Context Awareness>

[EuropeanCommission2010] "Social networks Overview: Current Trends and Research Challenges," European Commission, Future Media Networks Cluster, October 2010

[Prakash2009] Prakash, B.A.; Seshadri, M.; Sridharan, A.; Machiraju, S.; Faloutsos, C.; , "EigenSpokes: Surprising Patterns and Scalable Community Chipping in Large Graphs," Data Mining Workshops, 2009. ICDMW '09. IEEE International Conference on, vol., no., pp.290-295, 6-6 Dec. 2009

[White2010] White, T.; Chu, W.; Salehi-Abari, A., "Media Monitoring Using Social Networks," Social Computing (SocialCom), 2010 IEEE Second International Conference on , vol., no., pp.661,668, 20-22 Aug. 2010

[Abel2011-3] Abel F, Gao Q, Houben G, Houben K (2011) Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In Proceedings of the ACM WebSci '11, 3rd International Conference on Web Science, Koblenz, Germany, pp. 1-8

[Juyoung2010] Juyoung K, Hwan-Seung Y (2010) Mining Spatio-Temporal Patterns in Trajectory Data. Journal of Information Processing Systems , Vol.6, No.4, pp. 521-536

[Christopher2003] Christopher SC, Maglio PP, Cozzi A, Dom B (2003) Expertise identification using email communications. In Proceedings of the 12th international conference on Information and knowledge management (CIKM '03). ACM, pp.528-531.

[Chen2010] Chen J, Nairn R, Nelson L, Bernstein M, Chi E (2010) Short and tweet: experiments on recommending content from information streams. In Proceedings of the 28th international conference on Human factors in computing systems. ACM, pp. 1185–1194.

[Celik2011] Celik I, Abel F, Houben G (2011) Learning semantic relationships between entities in twitter. In Proceedings of the Web Engineering, pp. 167–181

[Yang2009] Yang Z, Qingquan L, Qingzhou M (2009) Mining time-dependent attractive areas and movement patterns from taxi trajectory data. In Proceedings of the 17th International Conference on Geoinformatics, pp.1-6

[Zhu2011] Zhu F (2011) Mining ship spatial trajectory patterns from AIS database for maritime surveillance. In Proceedings of the 2nd IEEE International Conference on Emergency Management and Management Sciences (ICEMMS), Beijing,China, pp. 772 - 775

[Braga2011] Braga RB. Martin H (2011) CAPTAIN: A Context-Aware system based on Personal TrAckINg. In Proceedings of the 17th International Conference on Distributed Multimedia Systems, Florence, Italy, pp. 130-133

[Lahiri2010] Lahiri M, Tanya YB (2010) Periodic subgraph mining in dynamic networks. Journal Knowledge and Information Systems, Volume 24, Issue 3, pp. 467-497

[Amit2006] Amit AN, Siva G, Gautam D, Dipanjan C, Koustuv D, Sougata M, Anupam J (2006) On the structural properties of massive telecom call graphs: findings and implications. In Proceedings of

the 15th ACM international conference on Information and knowledge management, USA, pp. 435–444

[Hayley2011] Hayley H, Yan H, Friedland G, Gatica-Perez D (2011) Estimating Dominance in Multi-Party Meetings Using Speaker Diarization. IEEE Transactions on Audio, Speech & Language Processing 19(4), pp. 847-860

[Yingjie2010] Yingjie Z, Kenneth RF, William AW (2010) Automatic Text Analysis of Values in the Enron Email Dataset: Clustering a Social Network Using the Value Patterns of Actors. In Proceedings of the 2010 43rd Hawaii International Conference on System Sciences (HICSS '10). IEEE Computer Society, Washington, DC, USA, pp. 1-10.

[Pawel2012] Pawel L, Mikolaj M (2012) Measuring the Importance of Users in a Social Network Based on Email Communication Patterns. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM, pp. 86-90

[Tapia2007] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, R. Friedman, Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor. 11th IEEE International Symposium on Wearable Computers, 2007, pp. 37-40.

[Ayua2012] M. A. Ayua, S. A. Ismail, A. F. Abdul Matina, T. Mantorob, A comparison study of classifier algorithms for mobile-phone's accelerometer based activity recognition. In proceeding of Elsevier Journal of Procedia Engineering, vol. 41, 2012, pp. 224-229.

[Garcia2012] S. Garcia, J. Derrac, J. R. Cano, F. Herrera, Prototype Selection for Nearest Neighbor Classification: Taxonomy and Empirical Study. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(3), 2012.

[Xindong2007] W. Xindong, Kumar, Vipin, R. Quinlan, J. Ghosh, Joydeep, Yang, Qiang, Motoda, Hiroshi, McLachlan, J. Geoffrey, Angus, Liu, Bing, Yu, S. Philip, Zhou, Zhi-Hua, Steinbach, Michael, Hand, J. David, S. Dan, Top 10 algorithms in data mining. Journal of Knowledge and Information Systems, vol. 14(1), 2007, pp. 1-37.

[Bao2004] Bao, Ling, and Stephen S. Intille. "Activity recognition from user-annotated acceleration data.", Pervasive Computing. Springer Berlin Heidelberg, 2004. 1-17.

[Lee2008] Kye-Hwan Lee, Joon-Hyuk Chang and Hyoung-Gon Kim, "Acoustic Environment Classification Algorithm for Context-aware Mobile Phone", Telecommunications Review, vol. 18, no. 1, 139-149p, 2008.02.

[Wang2010] Shuangquan Wang, Canfeng Chen, Jian Ma, "Accelerometer based transportation mode recognition on mobile phones", APWCS '10 Proceedings of the 2010 Asia-Pacific Conference on Wearable Computing Systems, 44-46p

[Han2012] Manhyung Han, La The Vinh, Young-Koo Lee and Sungyoung Lee, "Comprehensive Context Recognizer Based on Multimodal Sensors in a Smartphone", Journal of Sensors, vol. 12, no. 9, pp. 12588-12605, 2012

[Pourjam2013] Pourjam, E., Ide, I., Deguchi, D., & Murase, H. Segmentation of Human Instances Using Grab-cut and Active Shape Model Feedback. In proceddings of MVA2013 IAPR International Conference on Machine Vision Applications, pp. 77–80, May 20–23, 2013.

[Atefian2013] Atefian, M., & Mahdavi-Nasab, H. (2013). A Robust Mean-Shift Tracking Using GMM Background Subtraction, J. Basic. Appl. Sci. Res., vol. 3, no. 4, 596–607, 2013.

[Takur2013] Takur, R., Mehan, N., Namitakakkar. Recognition of human actions using motion history information extracted from the compressed. Int. J. Comput. Vis. Image Process. (IJCVIP), vol. 3, no. 7, pp. 973–977, 2013.

[Everts2014] Everts, I.; Gemert, J.C.V.; Gevers, T. Evaluation of color spatio-temporal interest points for human action recognition. IEEE Trans. Image Process, vol. 23, no. 4, pp. 1861–1869, 2014.

[Wang2011] Wang, H.; Klaser, A.; Schmid, C.; Liu, C.L. Action recognition by dense trajectories. In Proceedings of the 2011 IEEE Conference on. Computer Vision and Pattern Recognition (CVPR), pp. 3169–3176, 20–25 June 2011.

[Ayadi2011] Ayadi, M.E., Kamel, M.S., Karray, F.: Survey on speech emotion recognition: Features, classificat ion schemes, and databases. Pattern Recognition 44 (3), 572 - 587 (2011).

[Xiao2007] Z. Xiao, E. Dellandrea, W. Dou, and L. Chen. Automatic hierarchical classi_cation of emotional sp eech. In Multimedia Workshops, 2007. ISMW '07. Ninth IEEE International Symposium on, pages 291-296, 2007.

[Lee2011] C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan. Emotion recognition using a hierarchic al binary decision tree approach. Speech Commun., 53(9-10):1162-1171, Nov. 2011.

[Mao2010] Q.-R. Mao and Y.-Z. Zhan. A novel hierarchical speech emotion recognition method based on imp roved ddagsvm. Comput. Sci. Inf. Syst., 7(1):211-222, 2010.

[Azami2013] Azami, H., Malekzadeh, M., and Sanei, S. A New Neural Network Approach for Face Recognition based on Conjugate Gradient Algorithms and Principal Component Analysis. Journal of Mathema tics and Computer Science, vol. 6, pp. 166–175, 2013.

[Ghimire2013] Ghimire, D., & Lee, J. A Robust Face Detection Method Based on Skin Color and Edges. Journal of Information Processing Systems, vol. 9, no. 1, pp. 141–156, 2013.

[Beham2013] Beham, M. P., & Roomi, S. M. M. Face recognition using appearance based approach: A literature survey. International Journal of Computer Applications, vol. 12, pp. 16–21, 2012.

[Mistry2013] Mistry, V. J., & Goyani, M. M. A literature survey on facial expression recognition using global fe atures. Int. J. Eng. Adv. Technol, vol. 2, pp. 653–657, 2013.

[Kalita2013] Kalita, J., & Das, K. (2013). Recognition of Facial Expression Using Eigenvector Based Distribute d Features and Euclidean Distance Based Decision Making Technique. International Journal of Ad vanced Computer Science & Applications, vol. 4, no. 2, pp. 196–202, 2013.

[Ahsan2013] Ahsan, T., Jabid, T., & Chong, U. P. Facial Expression Recognition Using Local Transitional Patte rn on Gabor Filtered Facial Images. IETE Technical Review, vol. 30, no. 1, pp. 47–52, 2013.

[Kim2004] K. Kim, S. Bang, and S. Kim, "Emotion recognition system using short-term monitoring of physio logical signals," Medical and Biological Engineering and Computing, vol.42, no. 3, pp. 419-427, 2 004.

[Nasoz2003] F. Nasoz, C. L. Lisetti, K. Alvarez, and N. Finkelstein, "Emotion recognition from physiological si gnals for user modeling of affect," in M'2003, 9th International Conference on User Model, PA, Ju ne 2003.

[Haag2004] A. Haag, S. Goronzy, J. Williams, and P. Schaich, "Emotion recognition using bio-sensors: first st eps towards an automatic system," in Tutorial and Research Workshop on Affective Dialogues, vo

l. 1, June 2004, pp. 36 - 48.

[Setz2009] C. Setz, J. Schumm, C. Lorenz, B. Arnrich, and G. Tröster, "Using ensemble classifier systems for handling missing data in emotion recognition from physiology: one step towards a practical syste m" in International Conference on Affective Computing & Intelligent Interaction. Amsterdam, Net herlands, September 2009.

[Leon2007] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda.,"A user-independent real-time emotion recog nition system for software agents in domestic environments," Engineering Applications of Artifici al Intelligence, vol. 20, no. 3, pp. 337-345, April 2007.

<High-level Context Awareness>

[Abowd1997] Abowd, Gregory D., Christopher G. Atkeson, Jason Hong, Sue Long, Rob Kooper, and Mike Pinkerton. "Cyberguide: A mobile context-aware tour guide." Wireless networks, vol. 3, no. 5, pp. 421–433, 1997.

[Want1992] Want, Roy, Andy Hopper, Veronica Falcao, and Jonathan Gibbons. "The active badge location system." ACM Transactions on Information Systems (TOIS), vol. 10, no. 1, pp. 91–102, 1992.

[Ranganathan2003] Ranganathan, Anand, and Roy H. Campbell. "A middleware for context-aware agents in ubiquitous computing environments." In Middleware 2003, pp. 143–161. Springer Berlin Heidelberg, 2003.

[Ejigu2007] Ejigu, Dejene, Marian Scuturici, and Lionel Brunie. "Coca: A collaborative context-aware service platform for pervasive computing." In Information Technology, 2007. ITNG'07. Fourth International Conference on, pp. 297–302. IEEE, 2007.

[Yanwei2011] Yanwei, Song, Zeng Guangzhou, and Pu Haitao. "Research on the context model of intelligent interaction system in the internet of things." In IT in Medicine and Education (ITME), 2011 International Symposium on, vol. 2, pp. 379–382. IEEE, 2011.

[Guillemin2009] Guillemin, Patrick, and Peter Friess. "Internet of things strategic research roadmap." The Cluster of European Research Projects, Tech. Rep (2009).

[Dey2000] Dey, Anind K., Gregory D. Abowd, and Daniel Salber. "A context-based infrastructure for smart environments." In Managing Interactions in Smart Environments, pp. 114–128. Springer London, 2000.

[Carnot2011] Carnot Institutes, "Smart networked objects and internet of things." Carnot Institutes' Information Communication Technologies and Micro Nano Technologies alliance, White Paper, January 2011.

[Nalepa2014] Nalepa, Grzegorz J., and Szymon Bobek. "Rule-based solution for context-aware reasoning on mobile devices." Computer Science and Information Systems, vol. 11, no. 1, pp. 171–193, 2014.

[Saeedi2014] Saeedi, Sara, Adel Moussa, and Naser El-Sheimy. "Context-Aware Personal Navigation Using Embedded Sensor Fusion in Smartphones." Sensors, vol. 14, no. 4, pp. 5742–5767, 2014.

[Charith2013] Perera, Charith, Arkady Zaslavsky, Peter Christen, and Dimitrios Georgakopoulos. "Context aware computing for the internet of things: A survey." pp. 1–41, 2013.

[Banaee2013] Banaee, Hadi, Mobyen Uddin Ahmed, and Amy Loutfi. "Data Mining for Wearable Sensors in Health Monitoring Systems: A Review of Recent Trends and Challenges." Sensors, vol. 13, no. 12, pp. 17472–17500, 2013.

[Tenorth2013] Tenorth, Moritz, Fernando De la Torre, and Michael Beetz. “Learning probability distributions over partially-ordered human everyday activities.” In Robotics and Automation (ICRA), 2013 IEEE International Conference on, pp. 4539–4544. IEEE, 2013.

[Kayes2013] Kayes, A. S. M., Jun Han, and Alan Colman. “An ontology-based approach to context-aware access control for software services." In Web Information Systems Engineering–WISE 2013, pp. 410–420. Springer Berlin Heidelberg, 2013.

[Attard2013] Attard, Judie, Simon Scerri, Ismael Rivera, and Siegfried Handschuh. “Ontology-based situation recognition for context-aware systems.” In Proceedings of the 9th International Conference on Semantic Systems, pp. 113-120. ACM, 2013.

[Brian2009] Ziebart, Brian D., Andrew L. Maas, J. Andrew Bagnell, and Anind K. Dey. “Human Behavior Modeling with Maximum Entropy Inverse Optimal Control.” InAAAI Spring Symposium: Human Behavior Modeling, pp. 92–97, 2009.

[Institute2010] https://www.cso.nato.int/pubs/rdp.asp?RDP=RTO-TR-SAS-053

[American2010] http://www.auburn.edu/~newlamc/apa_div25/

[Michigan] https://www.wmich.edu/psychology/academics/graduate/ba/

[Helaoui2013] Helaoui, Rim, Daniele Riboni, and Heiner Stuckenschmidt. "A probabilistic ontological framework for the recognition of multilevel human activities." InProceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, pp. 345-354. ACM, 2013.

[Rodríguez2014] Rodríguez, Natalia Díaz, M. P. Cuéllar, Johan Lilius, and Miguel Delgado Calvo-Flores. "A survey on ontologies for human behavior recognition." ACM Computing Surveys (CSUR) 46, no. 4, pp. 2–33, 2014.

[Gonzalez2013] Gonzalez, Cleotilde. "The boundaries of Instance-Based Learning Theory for explaining decisions from experience." Decision making: Neural and behavioural approaches, vol. 202, pp. 73–98, 2013.

[Dandekar2013] Dandekar, Devendra B., and Vinay S. Kapse. “A Survey on Analysis of Evolving User Behavior Profile for Predicting Masquerades.” In International Journal of Engineering Research and Technology, vol. 2, no. 1, pp. 1–5, 2013.

[Rowe2013] Rowe, Matthew, Miriam Fernandez, Sofia Angeletou, and Harith Alani. "Community analysis through semantic rules and role composition derivation." Web Semantics: Science, Services and Agents on the World Wide Web 18, no. 1, pp. 31–47, 2013.

[Zhu2011] Zhu, Tian, Bai Wang, Bin Wu, and Chuanxi Zhu. “Role defining using behavior-based clustering in telecommunication network.” Expert Systems with Applications, vol. 38, no. 4, pp. 3902–3908, 2011.

[Knyazev2013] Knyazev, Boris. “Human nonverbal behavior multi-sourced ontological annotation.” In Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications, p. 2. ACM, 2013.

[Á lvarez2012] Álvarez, Nahum, and Federico Peinado. “Exploring body language as narrative interface.” In Interactive Storytelling, pp. 196-201. Springer Berlin Heidelberg, 2012.

<Reasoner and Prediction>

[Zazzi2010] M. Zazzi, R. Kaiser, A. Sonnerborg, et al., Prediction of response to antiretroviral therapy by human experts and by the EuResist data-driven expert system (the EVE study). HIV Medicine, 2010.

[Husain2010] Husain, W. and P. Lim Thean, The development of personalized wellness therapy recommender system using hybrid case-based reasoning. Computer Technology and Development (ICCTD), 2010 2nd International Conference on.

[Drăgulescu2007] Drăgulescu, Doina, and Adriana Albu. "Medical Predictions System." Acta Polytechnica Hungarica 4, no. 3 (2007): 89-101.

[Pandey2012] Dhiraj Pandey, Santosh kumar, "prediction system to support medical information system using data mining approach", International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 3, May-Jun 2012, pp.1988-1996.

[Yuan2014] Yuan, Bingchuan, and John Herbert. "Context-aware hybrid reasoning framework for pervasive healthcare. Personal and Ubiquitous Computing: April 2014, Volume 18, Issue 4, pp 865-881.

<Knowledge Maintenance Engine>

[Lars2013] Kotthoff, Lars. "Ranking Algorithms by Performance." arXiv preprint arXiv: 1311.4319 (2013).

[Quan2013] Sun, Quan, and Bernhard Pfahringer. "Pairwise meta-rules for better meta-learning-based algorithm ranking." Machine learning 93.1 (2013): 141-161.

[Marco2013] Maratea, Marco, Luca Pulina, and Francesco Ricca. "Automated selection of grounding algorithm in Answer Set Programming." AI* IA 2013: Advances in Artificial Intelligence. Springer International Publishing, 2013. 73-84.

[Hutter2013] Hutter, Frank, Holger H. Hoos, and Kevin Leyton-Brown. "Identifying key algorithm parameters and instance features using forward selection." Learning and Intelligent Optimization. Springer Berlin Heidelberg, 2013. 364-381.

[Debenham2003] Debenham, John. "A rigorous approach to knowledge base maintenance. "Developments in Applied Artificial Intelligence. Springer Berlin Heidelberg, 2003. 219-228.

[Debenham2005] Debenham, John. "An analysis of knowledge base maintenance." Computer Science and Information Systems 2.1 (2005): 1-29.

[McAdams2000] McAdams, Daniel A., and Kristin L. Wood. "Tuning parameter tolerance design: foundations, methods, and measures." Research in Engineering Design 12.3 (2000): 152-162.

[Regier2009] Regier, R., Gurjar, R., & Rocha, R. A. (2009). A clinical rule editor in an electronic medical record setting: development, design, and implementation. InAMIA Annual Symposium Proceedings (Vol. 2009, p. 537). American Medical Informatics Association.

[Kaljurand2008] Kaljurand, K. (2008, October). ACE View---an Ontology and Rule Editor based on Attempto Controlled English. In OWLED.

[Hulse2005] N. Hulse, R. Rocha, G. Del Fiol, R. Bradshaw, T. Hanna, and L. Roemer. Kat: a flexible xml-based knowledge authoring environment. Journal of the American Medical Informatics Association, 12(4):418{430, 2005.

[Dancy2008] Elizabeth Dancy, Uliyana Markova "Autonomic computing in Canadian academia"

[Dietrich2007] Dietrich, Andreas, Enrico Gobbetti, and Sung-Eui Yoon. "Massive-model rendering techniques: a tutorial." IEEE Computer Graphics and Applications27.6 (2007): 20-34.

[Shi2004] Shi, Rui, et al. "An adaptive image content representation and segmentation approach to automatic image annotation." Image and Video Retrieval. Springer Berlin Heidelberg, 2004. 545-554.

[Geissbuhler1999] Geissbuhler, Antoine, and Randolph A. Miller. "Distributing knowledge maintenance for clinical decision-support systems: the" knowledge library" model." Proceedings of the AMIA symposium. American Medical Informatics Association, 1999.

[Friedrich2011] Friedrich, Gerhard, and Markus Zanker. "A taxonomy for generating explanations in recommender systems." AI Magazine 32.3 (2011): 90-98.

[Duan2011] Duan, Lian, W. Nick Street, and E. Xu. "Healthcare information systems: data mining methods in the creation of a clinical recommender system." Enterprise Information Systems 5.2 (2011): 169-181.

[MEDLINE2014] Statistical Reports on MEDLINE/PubMed Baseline Data http://www.nlm.nih.gov/bsd/licensee/baselinestats.html [Accessed on 25th Jan, 2014]

<Feedback Analysis>

[xbox2014] http://www.digitaltrends.com/gaming/xbox-one-system-update-set-community-feedback-addressed-2014/#ixzz2y642sHSh

[Vashisth2011] Vashisth Pooja, and Punam Bedi. "Interest-Based personalized Recommender System." IEEE World Congress on Information and Communication Technologies (WICT), 2011.

[Zhang2011] Zhang, Hongxia, and Yuan Yang. "An E-Commerce Personalized Recommendation System Based on Customer Feedback." IEEE International Conference on Management and Service Science (MASS), 2011

Recommendation Manager

[Jannach2010] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, Recommender Systems - An Introduction. Cambridge University Press (CUP), 2010.

[Tintarev2007] N. Tintarev, “Explanations of recommendations,” in Proceedings of the 2007 ACM Conference on Recommender Systems (RecSys’07). Minneapolis, MN, USA: ACM, 2007, pp. 203–206.

[Masthoff2007] N. Tintarev and J. Masthoff, “Effective explanations of recommendations: user-centered design,” in Proceedings of the 2007 ACM Conference on Recommender Systems (RecSys’07). Minneapolis, MN, USA: ACM, 2007, pp. 153–156.

<UI/UX Authoring Tool>

[Liberman1995] H. Lieberman, “Letizia: An Agent That Assists Media Lab, 1995, http://web.media.mit.edu/~lieber/Lieberary/Letizia/Letizia-AAAI/Letizia.html.

[Vivacqua1997] Vivacqua, H. Lieberman, and N.V. Dyke, “Let’s Browse: A Collaborative Web Browsing Agent,” MIT Media Lab, 1997 http://web.media.mit.edu/~lieber/Lieberary/Lets-Browse/Lets-Browse-Intro.html

[Jiang2007] He, Jiang, and I-Ling Yen. "Adaptive user interface generation for web services." e-Business Engineering, 2007. ICEBE 2007. IEEE International Conference on. IEEE, 2007.

[Patent2010] "ADAPTIVE USER INTERFACE." U.S. Patent 20,100,097,331, issued April 22, 2010.

[Matthias2013] Peissner, Matthias, and Rob Edlin-White. "User Control in Adaptive User Interfaces for Accessibility." Human-Computer Interaction–INTERACT 2013. Springer Berlin Heidelberg, 2013. 623-640.

[Ibrahim2011] Ibrahim, Razale, and Rozilawati Razali. "A performance-oriented interface design model of web applications." Electrical Engineering and Informatics (ICEEI), 2011 International Conference on. IEEE, 2011.

[Wu2012] Wu, Wen-Yi. "Adaptive user interface." U.S. Patent Application 12/334,720.