Home

Catalog

- Requirements
- Project Architecture
- Meeting

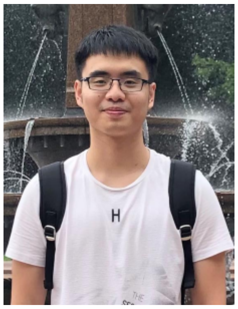
Useful Links

Link	Description
Zoom Meeting	Group meeting
Supervisor Meeting	Weekly meeting with the supervisor
NDIS Code of Conduct	Promote safe and ethical service delivery
Submitting behaviour support plans and reports	For practitioners: how to lodge behaviour support plans
Github	Repository for the project
Trello	Project Trello
Trello_Sprint 2 (Old Version)	Sprint backlog of the sprint 2
Trello_Sprint 3 (Old Version)	Sprint backlog of the sprint 3

Stakeholders

Role	Name	Email
Supervisor	Samodha Pallewatta	**samodha.pallewatta@unimelb.edu.au**
Client	Michael Kirley	**mkirley@unimelb.edu.au**
Client	Hanna Navissi	**hanna.navissi@unimelb.edu.au**

Project Team

Name	Pohto	Email	Index	Role	Responsibility
Donghui Guo		donghuig @student.unimelb.edu.au	Donghui GUO	Back-end Developer	▪ Develop Database ▪ Develop database API ▪ Develop Back-end framework ▪ Test the performance of the system

Aijia Gong		aijiag @student.unimeb.edu.au	Aijia Gong	Product owner	• Develop pdf data extraction • clarify project requirement • test and compare test result
Xuande Li		xuandel @stundent.unimelb.edu.au	Xuande LI	Scrum Master	• Develop Database • Develop database API • Develop Back-end framework
Jiyuan Wang		jiyuanw @student.unimelb.edu.au	Jiyuan Wang	Back-end Developer	• pdf data extraction • data extraction test
Ruize Fu		rufu @student.unimelb.edu.au	Ruize Fu		

Requirements

Catalog

Background

Background

The national disability insurance scheme (NDIS) code of conduct (the "code") is set out in the national disability insurance scheme (code of conduct) Rules 2018, which is the NDIS rule established in accordance with the national disability insurance scheme act 2013 (NDIS act). The association employs service providers and workers, and the service providers develop plans to support the disabled. The workers receive plans from the practitioners and help the disabled according to the plans.
This guideline is intended to work with other elements of quality and assurance arrangements to promote safe and skilled labor within NDIS. Providing high-quality support for the disabled involves not only the right ability, but also the right attitude. Non-disability information service providers and their participants need to be familiar with the basic principles of non-disability information, respect the rights of persons with disabilities, aim at preventing injuries, and respond appropriately when injuries occur.

Project overview

In order to provide support to the disabled, NDIS behavioral support practitioners develop a written plan in the document and upload it to the NDIS Committee portal using the professional behavioral support provider's own template. The clients of the project are the NDIS behavior support practitioners who can download and fill out the behavior support plan to learn feedback. The aim of this project is about providing a training system for NDIS practitioners to develop good support plans for the participants.

By developing this system, the practitioners can efficiently get feedback on their plans and make changes to the implementation instead of waiting a long time for human judgments. Therefore, more and more disabilities are able to get better care and help through NDIS.

The whole system includes a front-end port that can receive the PDF files submitted by practitioners. Besides, a back-end software should be designed for scanning and extracting useful information from PDF files submitted by practitioners. These data should be structured and layered, and then stored in a structured database. The data stored in the database can be further used as input to the NLP language model and generate feedback for written plans submitted by NDIS behavior support practitioners, which is not included in the project.

In this project, the application mainly focuses on front-end development, which is about designing an API for receiving PDF files, and back-end development which is about extracting key information from the PDF planning document and saving the information in the database. The development of the NLP language model is not included in the project.

Goal

Provide a backend software that can scan and extract useful information from PDF files submitted by practitioners, store the data structurally and hierarchically in our own designed database, generate feedback for written plans submitted by NDIS behavior support practitioners, and download planning documents through the canvas.

References

Anon, NDIS code of Conduct. *NDIS Quality and Safeguards Commission*. Available at: https://www.ndiscommission.gov.au/about/ndis-code-conduct [Accessed August 19, 2022].

Anon, Submitting behaviour support plans and reports. *NDIS Quality and Safeguards Commission*. Available at: https://www.ndiscommission.gov.au/providers/understanding-behaviour-support-and-restrictive-practices-providers/submitting-behaviour [Accessed August 19, 2022].

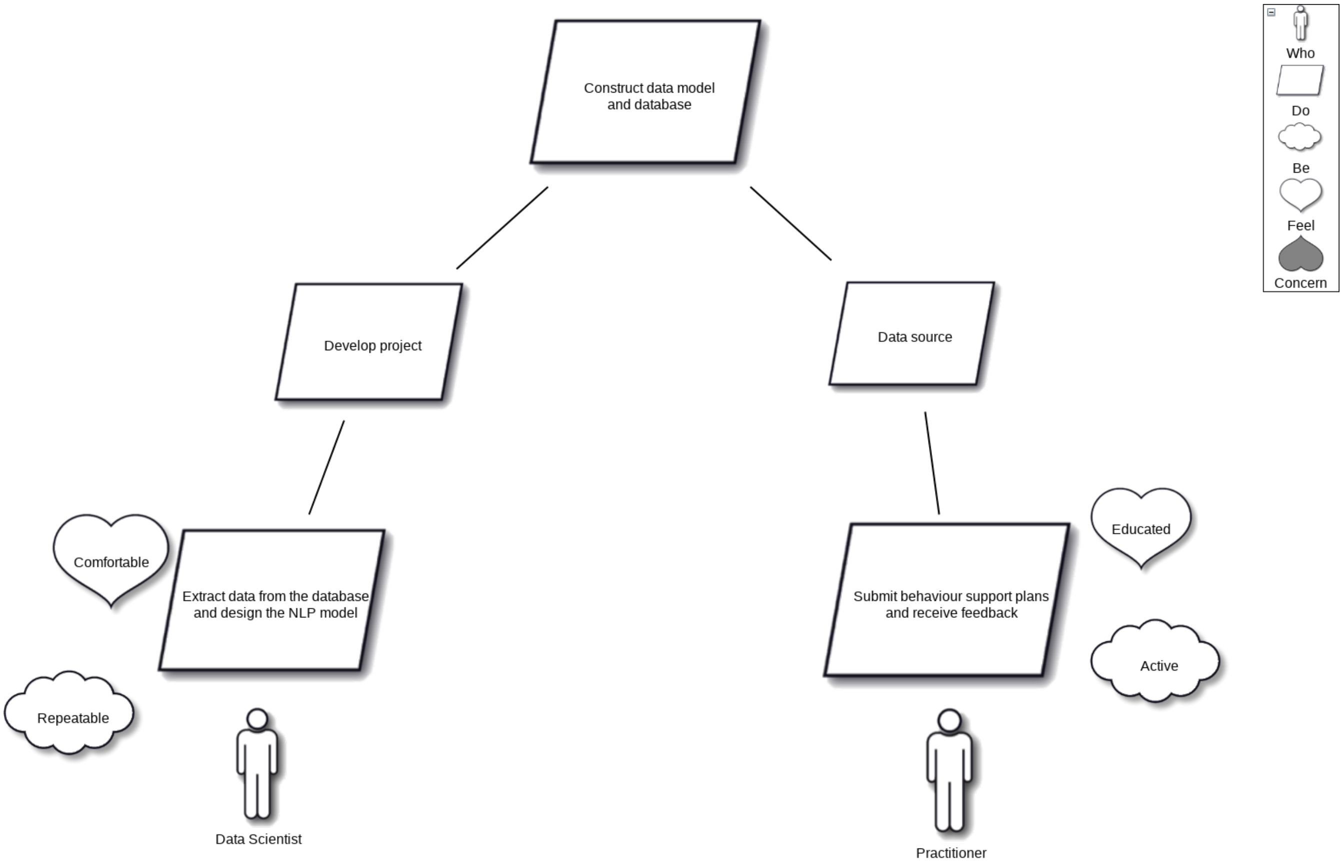
Motivational Model

Version 2.0

Do-Be-Feel-Who List

Who(Roles)	Do(Functional Goal)	Be(Quality Goal)	Feel(Emotion Goal)
Behaviour Support Practitioner	Submit behaviour support plans and receive feedback	Active	Educated
Data Scientist	Extract data from the database and design the NLP model	Repeatable	Comfortable

Goal Model


Construct data model and database
Develop project
Data source
Comfortable
Extract data from the database and design the NLP model
Repeatable
Data Scientist
Educated
Submit behaviour support plans and receive feedback
Active
Practitioner
Who
Do
Be
Feel
Concern

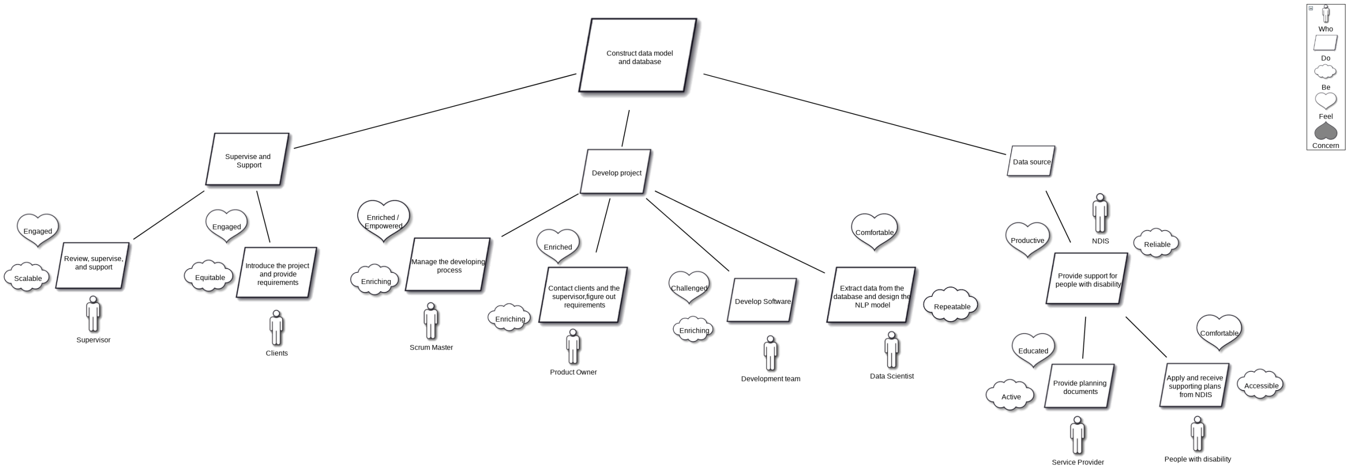
Version 1.0

Do-Be-Feel-Who List

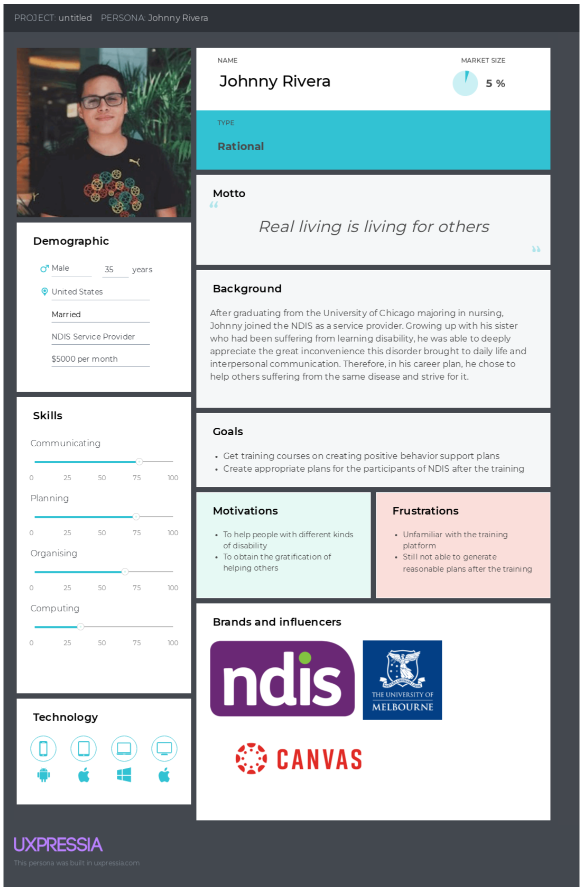
Who(Roles)	Do(Functional Goal)	Be(Quality Goal)	Feel(Emotion Goal)
Supervisor	Review information, supervise, and support the development team	Scalable	Engaged
Clients	Introduce the project and provide requirements	Equitable	Engaged
Scrum Master	Manage the developing process	Enriching	Enriched / Empowered
Product Owner	Contact clients and the supervisor, and figure out requirements	Enriching	Enriched
Development team	Develop Software	Enriching	Challenged
NDIS	Provide support for people with disability	Reliable	Productive
Service Provider	Provide planning documents	Active	Educated

People with disability	Apply and receive supporting plans from NDIS	Accessible	Comfortable
Data Scientist	Extract data from the database and design the NLP model	Repeatable	Comfortable

Goal Model



Personas



PROJECT: untitled PERSONA: Johnny Rivera

NAME

Johnny Rivera

MARKET SIZE

5 %

TYPE

Rational

Motto

“Real living is living for others”

Demographic

Male 35 years

United States

Married

NDIS Service Provider

$5000 per month

Background

After graduating from the University of Chicago majoring in nursing, Johnny joined the NDIS as a service provider. Growing up with his sister who had been suffering from learning disability, he was able to deeply appreciate the great inconvenience this disorder brought to daily life and interpersonal communication. Therefore, in his career plan, he chose to help others suffering from the same disease and strive for it.

Skills

Communicating

0 25 50 75 100

Planning

0 25 50 75 100

Organising

0 25 50 75 100

Computing

0 25 50 75 100

Goals

- Get training courses on creating positive behavior support plans
- Create appropriate plans for the participants of NDIS after the training

Motivations

- To help people with different kinds of disability
- To obtain the gratification of helping others

Frustrations

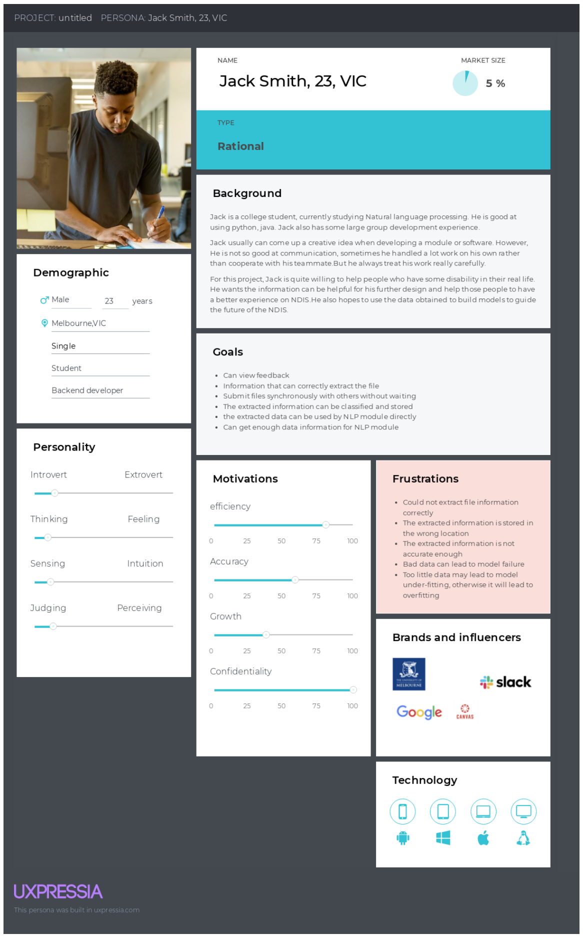
- Unfamiliar with the training platform
- Still not able to generate reasonable plans after the training

Brands and influencers

Technology

UXPRESSIA

This persona was built in uxpressia.com



PROJECT: untitled PERSONA: Jack Smith, 23, VIC

NAME

Jack Smith, 23, VIC

MARKET SIZE

5 %

TYPE

Rational

Background

Jack is a college student, currently studying Natural language processing. He is good at using python, java. Jack also has some large group development experience.

Jack usually can come up a creative idea when developing a module or software. However, He is not so good at communication, sometimes he handled a lot work on his own rather than cooperate with his teammate.But he always treat his work really carefully.

For this project, Jack is quite willing to help people who have some disability in their real life. He wants the information can be helpful for his further design and help those people to have a better experience on NDIS.He also hopes to use the data obtained to build models to guide the future of the NDIS.

Demographic

Male 23 years

Melbourne,VIC

Single

Student

Backend developer

Goals

- Can view feedback
- Information that can correctly extract the file
- Submit files synchronously with others without waiting
- The extracted information can be classified and stored
- the extracted data can be used by NLP module directly
- Can get enough data information for NLP module

Personality

Introvert Extrovert

Thinking Feeling

Sensing Intuition

Judging Perceiving

Motivations

efficiency

0 25 50 75 100

Accuracy

0 25 50 75 100

Growth

0 25 50 75 100

Confidentiality

0 25 50 75 100

Frustrations

- Could not extract file information correctly
- The extracted information is stored in the wrong location
- The extracted information is not accurate enough
- Bad data can lead to model failure
- Too little data may lead to model under-fitting, otherwise it will lead to overfitting

Brands and influencers

Technology

UXPRESSIA

This persona was built in uxpressia.com

Plan

Version 1.0

Module 1: Data Extracting

1. Scan information from PDF file
2. Analysis and process information
    a. Figure out the relationship among the extracted data.
    b. Determine the type of the extracted data.

Module 2: Database Design

1. Definition of Data (Integer, Array, String, Boolean)
2. Database Server Hosting
    a. Need to choose an appropriate host for Database
    b. Determine budget for hosting server bandwidth
    c. Test Hosting ability for RAM etc.
3. Database Coding
    a. Determine Database management language: MySQL, NoSQL, etc.
    b. Determine Database type: MongoDB, Neo4J, etc.
    c. Coding
    d. Debugging
    e. Run on localhost to test availability and efficiency

Module 3: Connection

1. Connect Database with hosting server
    a. Upload Database to Cloud Computer
    b. Import runtime environment
    c. Test on localhost, debugging
    d. Open hosting, test connectivity
    e. Upload Data
2. Store the extracted information into the database

Product Backlog

Version 4.0

Epic	ID	User	User Story	Story Estimate	MoSCoW Priority	Task	Task Estimate	Subtask (Transfer to sprint backlog)	Sprint ID
Prototype	1	Data Scientist	As a data scientist, I want to obtain the information of BSP so that I can process the data and give feedback.	23	Must Have	1.1 Extract data from PDF documentation	25	• 1.1.1 Load the document from the target location • 1.1.2 Extract information based on the text headings of different areas	2,3
						1.2 Design the database	13	• 1.2.1 Design the structure of the database • 1.2.2 Create database • 1.2.3 Create tables and related attributes	2
						1.3 Store the information in the database	5	• 1.3.1 Link the database with the extracting module • 1.3.2 Write information into corresponding tables	2
	3	Service Provider Practitioner	As a service provider practitioner, I want to submit the behaviour support plan through the website easily.	5	Could Have	3.1 Integrate the API with the whole system for document interaction	4	• 3.1.1 Select a web port for PDF files submission (e. g. canvas) • 3.1.2 Use postman to extract uploaded PDF files • 3.1.3 Integration with the data extraction module	3
Optimisation	2	Data Scientist	As a data scientist, I want to get access to the collected information easily so that I can preprocess data as soon as possible.	5	Should Have	2.1 Simplify the process of searching for information in the database	16	• 2.1.1 Optimize the structure design of the database • 2.1.2 Reduce the duplication in table data storage • 2.1.3 Design the database API that can gain information from the database	3

Total Story Point63

Version 3.0

Epic	ID	User	User Story	Story Estimate	MoSCoW Priority	Task	Task Estimate	Subtask (Transfer to sprint backlog)	Sprint ID
Prototype	1	Data Scientist	As a data scientist, I want to obtain the information of BSP so that I can process the data and give feedback.	23	Must Have	1.1 Extract data from PDF documentation	17	• 1.1.1 Load the document from the target location • 1.1.2 Extract information based on the text headings of different areas	2
						1.2 Design the database	13	• 1.2.1 Design the structure of the database • 1.2.2 Create database • 1.2.3 Create tables and related attributes	2
						1.3 Store the information in the database	5	• 1.3.1 Link the database with the extracting module • 1.3.2 Write information into corresponding tables	2
	3	Service Provider Practitioner	As a service provider practitioner, I want to submit the behaviour support plan through the website easily.	5	Could Have	3.1 Integrate the API with the whole system for document interaction	5	• 3.1.1 Select a web port for PDF files submission (e.g. canvas) • 3.1.2 Use postman to extract uploaded PDF files • 3.1.3 Integration with the data extraction module	3

	6	Service Provider Practitioner	As a service provider practitioner, I want to submit the BSP concurrently with other staff so that I can shorten the time for waiting online.	8	Must Have	6.1 Apply the concurrency characteristic to the whole system	3	• 6.1.1 Design a multithreading extracting module • 6.1.2 Design a multithreading processing module • 6.1.3 Design a multithreading database that can read and write data concurrently	3
Optimisation	2	Data Scientist	As a data scientist, I want to get access to the collected information easily so that I can preprocess data as soon as possible.	5	Should Have	2.1 Simplify the process of searching for information in the database	12	• 2.1.1 Optimize the structure design of the database • 2.1.2 Reduce the duplication in table data storage • 2.1.3 Design the database API that can gain information from the database	3
	5	Service Provider Practitioner	As a service provider practitioner, I want the website can process my behavior support plans efficiently so that I can get feedback in a short period and correct my plan early.	12	Should Have	5.1 Optimize the design of the processing module	4	• 5.1.1 Avoid coding loops and coding redundancy • 5.1.2 Reduce the time complexity of the algorithms	3
						5.2 Optimize the design of the linking module	4	• 5.2.1 Optimize the interaction between the two modules to reduce the time • 5.2.2 Lazy loading	3
						5.3 Optimize the database	4	• 5.3.1 Proper indexing • 5.3.2 Avoid unused and temporary tables	3

	7	Service Provider Practitioner	As a service provider practitioner, I want my submitted BSP secured so that the personal information of participants and staff will be protected.	10	Could have	7.1 The system requires certain identification and encryption processing functions	3	- 7.1.1 Staff need to be authenticated with a unique employee ID	3

Total Story Point70

Version 2.0

Epic	ID	User	User Story	Story Estimate	MoSCoW Priority	Task	Task Estimate	Subtask (Transfer to sprint backlog)	Sprint ID
Prototype	1	Data Scientist	As a data scientist, I want to obtain the information of BSP so that I can process the data and give feedback.	23	Must Have	1.1 Extract data from PDF documentation	17	- 1.1.1 Load the document from the target location - 1.1.2 Extract information based on the text headings of different areas	2
						1.2 Design the database	13	- 1.2.1 Design the structure of the database - 1.2.2 Create database - 1.2.3 Create tables and related attributes	2
						1.3 Store the information in the database	5	- 1.3.1 Link the database with the extracting module - 1.3.2 Write information into corresponding tables	2
	3	Service Provider Practitioner	As a service provider practitioner, I want to submit the behaviour support plan through the website easily.	5	Could Have	3.1 Integrate the API with the whole system for document interaction	5	- 3.1.1 Select a web port for PDF files submission (e.g. canvas) - 3.1.2 Use postman to extract uploaded PDF files - 3.1.3 Integration with the data extraction module	3

	6	Service Provider Practitioner	As a service provider practitioner, I want to submit the BSP concurrently with other staff so that I can shorten the time for waiting online.	8	Must Have	6.1 Apply the concurrency characteristic to the whole system	3	• 6.1.1 Design a multithreading extracting module • 6.1.2 Design a multithreading processing module • 6.1.3 Design a multithreading database that can read and write data concurrently	3
Optimisation	2	Data Scientist	As a data scientist, I want to get access to the collected information easily so that I can preprocess data as soon as possible.	5	Should Have	2.1 Simplify the process of searching for information in the database	5	• 2.1.1 Optimize the structure design of the database • 2.1.2 Reduce the duplication in table data storage	3
	5	Service Provider Practitioner	As a service provider practitioner, I want the website can process my behavior support plans efficiently so that I can get feedback in a short period and correct my plan early.	12	Should Have	5.1 Optimize the design of the processing module	4	• 5.1.1 Avoid coding loops and coding redundancy • 5.1.2 Reduce the time complexity of the algorithms	3
						5.2 Optimize the design of the linking module	4	• 5.2.1 Optimize the interaction between the two modules to reduce the time • 5.2.2 Lazy loading	3
						5.3 Optimize the database	4	• 5.3.1 Proper indexing • 5.3.2 Avoid unused and temporary tables	3

	7	Service Provider Practitioner	As a service provider practitioner, I want my submitted BSP secured so that the personal information of participants and staff will be protected.	10	Could have	7.1 The system requires certain identification and encryption processing functions	10	• 7.1.1 People with high permission need voice recognization and additional verification • 7.1.2 The user needs to authenticate if he or she has entered an error more than three times • 7.1.3 Staff need to be authenticated with a unique employee ID	3

Total Story Point70

Version 1.0

Epic	ID	User	User Story	Story Estimate	MoSCoW Priority	Task	Task Estimate	Subtask (Transfer to sprint backlog)	Sprint ID
Prototype	1	Data Scientist	As a data scientist, I want to obtain the information of BSP so that I can process the data and give feedback.	23	Must Have	Extract data from PDF documentation	4	• Load the document from the target location • Extract information based on the text headings of different areas	2
						Design the database	12	• Design the structure of the database • Create database • Create tables and related attributes	2
						Store the information in the database	7	• Link the database with the extracting module • Write information into corresponding tables	2
	3	Data Scientist	As a data scientist, I want to obtain meaningful and well-classified data so that I can improve the quality of the feedback generated by my NLP language models.	5	Could Have	Analyze and structure the extracted data	5	• Preprocess the extracted data • Develop a well-structured database for storing the preprocessed data	2

	7	Service Provider	As a service provider, I want to submit the BSP concurrently with other staff so that I can shorten the time for waiting online.	8	Must Have	Apply the concurrency characteristic to the whole system	8	• Design a multithreading extracting module • Design a multithreading processing module • Design a multithreading database that can read and write data concurrently	2
Optimisation	2	Data Scientist	As a data scientist, I want to get access to the collected information easily so that I can preprocess data as soon as possible.	5	Should Have	Simplify the process of searching for information in the database	5	• Optimize the structure design of the database • Reduce the duplication in table data storage	3
	5	Service Provider	As a service provider, I want to upload my behavior support plan without restriction so that I can practice as many times as I want.	9	Could Have	Check whether a similar document is submitted by the same service provider	5	• Build an NLP model to decide the document similarity • Check the similarity of the document based on the participant that is submitted by the same service provider	3
						Update data that is repeatedly committed in the database	4	• Save the whole data extracted from the document with low similarity • Update the table based on the new-submitted document with high similarity	3
	6	Service Provider	As a service provider, I want the website can process my behavior support plans efficiently so that I can get feedback in a short period and correct my plan early.	12	Should Have	Optimize the design of the processing module	4	• Avoid coding loops and coding redundancy • Reduce the time complexity of the algorithms	3
						Optimize the design of the linking module	4	• Optimize the interaction between the two modules to reduce the time • Lazy loading	3
						Optimize the database	4	• Proper indexing • Avoid unused and temporary tables	3
	8	Service Provider	As a service provider, I want my submitted BSP secured so that the personal information of participants and staff will be protected.	10	Could have	The system requires certain identification and encryption processing functions	10	• People with high permission need voice recognization and additional verification • The user needs to authenticate if he or she has entered an error more than three times • Staff need to be authenticated with a unique employee ID	3

Total Story Point72

Road Map

Sprint 1: Inception (or design sprint)	**Due** Aug 22 at 13:59
Sprint 2: Development	**Due** Sep 19 at 13:00
Sprint 3: Development	**Due** Oct 21 at 13:00
Sprint 4: Product	**Due** Nov 4 at 13:00

User Stories

Version 3.0

ID	User	User Story	MoSCoW Priority
1	Data Scientist	As a data scientist, I want to obtain information of BSP so that I can process the data and give feedback.	Must Have
2	Data Scientist	As a data scientist, I want to get access to the collected information easily so that I can preprocess data as soon as possible.	Should Have
3	Service Provider Practitioner	As a service provider practitioner, I want to submit the behaviour support plan through Canvas easily.	Could Have
4	Service Provider Practitioner	As a service provider practitioner, I want to receive feedback from the system so that I can improve my behavior support plan based on the feedback.	Won't Have This Time
5	Service Provider Practitioner	As a service provider practitioner, I want the website can process my behavior support plans efficiently so that I can get feedback in a short period and correct my plan early.	Should Have
6	Service Provider Practitioner	As a service provider practitioner, I want to submit the BSP concurrently with other staff so that I can shorten the time for waiting online.	Should Have
7	Service Provider Practitioner	As a service provider practitioner, I want my submitted BSP secured so that the personal information of participants and staff will be protected.	Could Have

Version 2.0

ID	User	User Story	MoSCoW Priority
1	Data Scientist	As a data scientist, I want to obtain the information of BSP so that I can process the data and give feedback.	Must Have
2	Data Scientist	As a data scientist, I want to get access to the collected information easily so that I can preprocess data as soon as possible.	Should Have
3	Service Provider Practitioner	As a service provider practitioner, I want to submit the behaviour support plan through the website easily.	Could Have
4	Service Provider Practitioner	As a service provider practitioner, I want to receive feedback from the system so that I can improve my behavior support plan based on the feedback.	Won't Have This Time
5	Service Provider Practitioner	As a service provider practitioner, I want the website can process my behavior support plans efficiently so that I can get feedback in a short period and correct my plan early.	Should Have
6	Service Provider Practitioner	As a service provider practitioner, I want to submit the BSP concurrently with other staff so that I can shorten the time for waiting online.	Must Have
7	Service Provider Practitioner	As a service provider practitioner, I want my submitted BSP secured so that the personal information of participants and staff will be protected.	Could Have

Version 1.0

ID	User	User Story	MoSCoW Priority
1	Data Scientist	As a data scientist, I want to obtain the information of BSP so that I can process the data and give feedback.	Must Have
2	Data Scientist	As a data scientist, I want to get access to the collected information easily so that I can preprocess data as soon as possible.	Should Have
3	Data Scientist	As a data scientist, I want to obtain meaningful and well-classified data so that I can improve the quality of the feedback generated by my NLP language models.	Could Have
4	Service Provider	As a service provider, I want to receive feedback from the system so that I can improve my behavior support plan based on the feedback.	Won't Have This Time
5	Service Provider	As a service provider, I want to upload my behavior support plan without restriction so that I can practice as many times as I want.	Could Have
6	Service Provider	As a service provider, I want the website can process my behavior support plans efficiently so that I can get feedback in a short period and correct my plan early.	Should Have
7	Service Provider	As a service provider, I want to submit the BSP concurrently with other staff so that I can shorten the time for waiting online.	Must Have

8	Service Provider	As a service provider, I want my submitted BSP secured so that the personal information of participants and staff will be protected.	Could Have

Acceptance Criteria

Version 2.0

ID	User	User Story	Given	When	Then
1	Data Scientist	As a data scientist, I want to obtain the information of BSP so that I can process the data and give feedback.	I have a database that contains all the information of positive behaviour support plans	I search for data through the corresponded location where data is stored	I can obtain all needed data needed to give suitable feedback
2	Data Scientist	As a data scientist, I want to get access to the collected information easily so that I can preprocess data as soon as possible.	I have a database that contains all the information of positive support behaviour plans	I try to access a target instance	I can easily get permission for loading the required data in a short time
3	Service Provider Practitioner	As a service provider practitioner, I want to submit the behaviour support plan through the website easily.	I have generated several positive behaviour support plans	I upload the document through the given port	I successfully submit the document without issue
4	Service Provider Practitioner	As a service provider practitioner, I want the website can process my behavior support plans efficiently so that I can get feedback in a short period and correct my plan early.	I have generated several positive behaviour support plans	I submit the BSP through the user interface	I receive the feedback in a short time

Version 1.0

ID	User	User Story	Given	When	Then
1	Data Scientist	As a data scientist, I want to obtain the information of BSP so that I can process the data and give feedback.	I have a database that contains all the information of positive behaviour support plans	I search for data through the corresponded location where data is stored	I can obtain all needed data needed to give suitable feedback
2	Data Scientist	As a data scientist, I want to get access to the collected information easily so that I can preprocess data as soon as possible.	I have a database that contains all the information of positive support behaviour plans	I try to access a target instance	I can easily get permission for loading the required data in a short time
3	Service Provider Practitioner	As a service provider practitioner, I want to submit the behaviour support plan through the website easily.	I have generated several positive behaviour support plans	I upload the document through the given port	I successfully submit the document without issue
5	Service Provider Practitioner	As a service provider practitioner, I want the website can process my behavior support plans efficiently so that I can get feedback in a short period and correct my plan early.	I have generated several positive behaviour support plans	I submit the BSP through the user interface	I receive the feedback in a short time
6	Service Provider Practitioner	As a service provider practitioner, I want to submit the BSP concurrently with other staff so that I can shorten the time for waiting online.	I have generated several positive behaviour support plans	I submit the BSP through the user interface	I successfully submit the document immediately
7	Service Provider Practitioner	As a service provider practitioner, I want my submitted BSP secured so that the personal information of participants and staff will be protected.	I have generated several positive behaviour support plans	I upload the BSP document	The personal information in the files is protected and cannot be accessed by other people without permission

Sprint Artifacts

Sprint planning, review, and retrospective for Sprint 2 and Sprint 3

Catalog

Sprint Artifacts - Sprint 2

Sprint Planning:

At the beginning of this sprint, we made a plan to complete task1.1, task1.2 and task1.3.

task1.1 is information extraction. We plan to complete the main function of the information extraction part in sprint2, and we hope that the information we want can be successfully extracted by the end of sprint2. This part is in charge of jiyuan wang and aijia gong. They need to decide which way to extract the table information. One of them is responsible for the brute force method, and the other tries a third-party library, decides whether to use the third-party library or brute force according to the result, and then further develop the basic functions according to the selected scheme. Then they also need to build a good structure to classify and encapsulate the data. Finally, in order to successfully transfer the extracted data to the database, they also need to output the extracted data into a file format acceptable to the database.

Task1.2 is in charge of xuande li and donghui guo. The content of task1.2 is mainly to design and develop the database. After an initial discussion they decided to use a relational database. They need to design the ER model and physical model according to the relationship between various questions and answers in the table, and then build a real database system based on the designed results. During the design process, they also need to communicate with the person in charge of the information extraction task to ensure that the format of the extracted information file can be directly imported into the database.

task1.3 is the connection work to be done when the information extraction of task1.1 and the database of task1.2 are completed. They need to use API or library to extract the file from task1.1, and import the content of the file into the database. This task is mainly to connect the first two tasks successfully so as to ensure the successful operation of the entire system.

Sprint Review:

At the end of sprint2, we have completed all the contents of task1.2 perfectly, and the database has been completed. task1.1 can also correctly extract most of the information in the table, but we also found more bugs and some vague details that need to be further discussed with the client. We found issues with two test files where the order was reversed, and when the user was asked to make a choice, the user gave up. In addition, there are some details in the table that may require us to modify the code.The work of task1.3 went very smoothly, we successfully connected the database and the information extraction system, and now the system is ready to run.

We submitted the code for all our tasks and also wrote readme files for each task explaining how to run our files.

Sprint Artifacts - Sprint 3

Sprint Planning:

In this sprint, based on the sprint backlog, we need to finish tasks 2.1, 3.1, and 1.1.2.

Task 1.1.2 belongs to the section of the data extraction module and should be carried on by Aijia and Jiyuan. Currently, they have finished coding the basic structure of the information extraction, but there are still some bugs that need to be fixed in the following period. For example, they found that the order of the problems in the two different test files was reversed and parts of the function module can be repeated if the user needs it. To solve the problems, they should have further discussions with themselves and try different algorithms in programming. And if the problems cannot be solved due to programming skills, they might need to ask the supervisor for help and ask the client to check the requirements in detail if the problem is caused by lacking sufficient document materials.

For task 2.1 and task 3.1, Xuande Li and Donghui Guo should cooperate to finish the assigned tasks. Firstly, they need to optimize the structure of the database by deleting duplicate tables and combining some tables together to make the architecture of the database clearer. Secondly, with the final optimized database, they need to design the database API that can gain information from the database. Basically, the API should include 7 functions, one connecting function and six functions that can gain information according to the documentation pages. Besides, they should use the front-end API and framework to deal with issues of sending pdf documents through web services and integrating the data extraction module and data saving module.

Additionally, our group members have some confusion about some requirements and the given pdf document. Therefore, we will hold a client meeting with Michael on 28/09 to deal with these issues.

What's more, we need to check the performance of the project by applying test cases to each module and the whole system.

Sprint Review:

At the end of Sprint 3, task 2.1 and task 3.1 are perfectly completed, and task 1.1.2 is almost finished with some small problems that do not affect the routine of the program.

All sections of our project are finally presented with several python files, and we have uploaded the programming files that can basically implement all functions on GitHub.

In addition, we have also submitted the readme document and image files that contain testing steps for running the whole program and results of the test cases based on each task.

And these documents can be seen as the reference of our work completion and guidelines for the client for testing the project.

Project Architecture

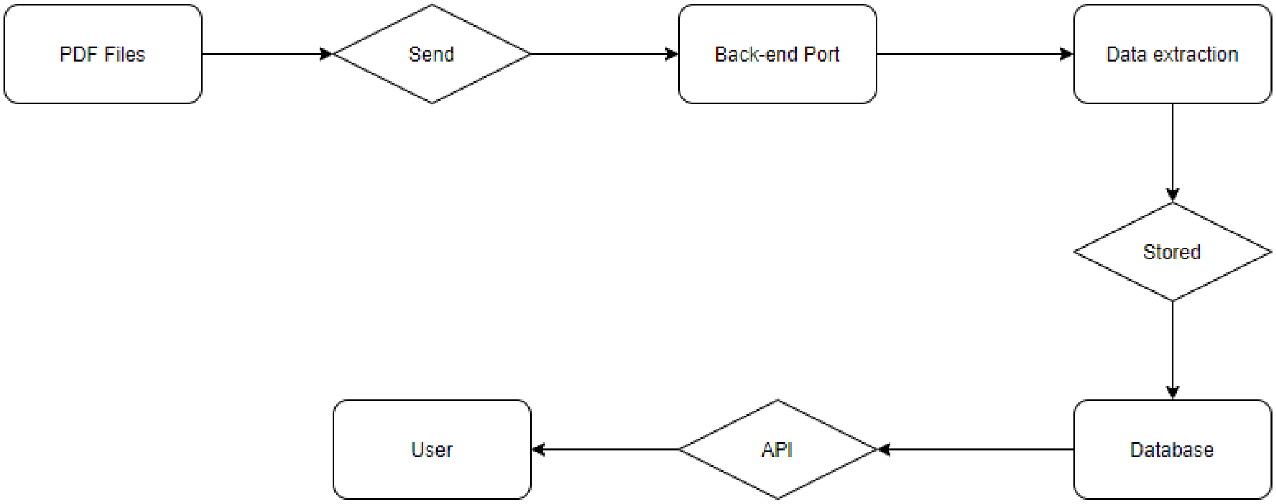
This section mainly concentrates on the developing progress of the whole project with the following pages.

Catalog

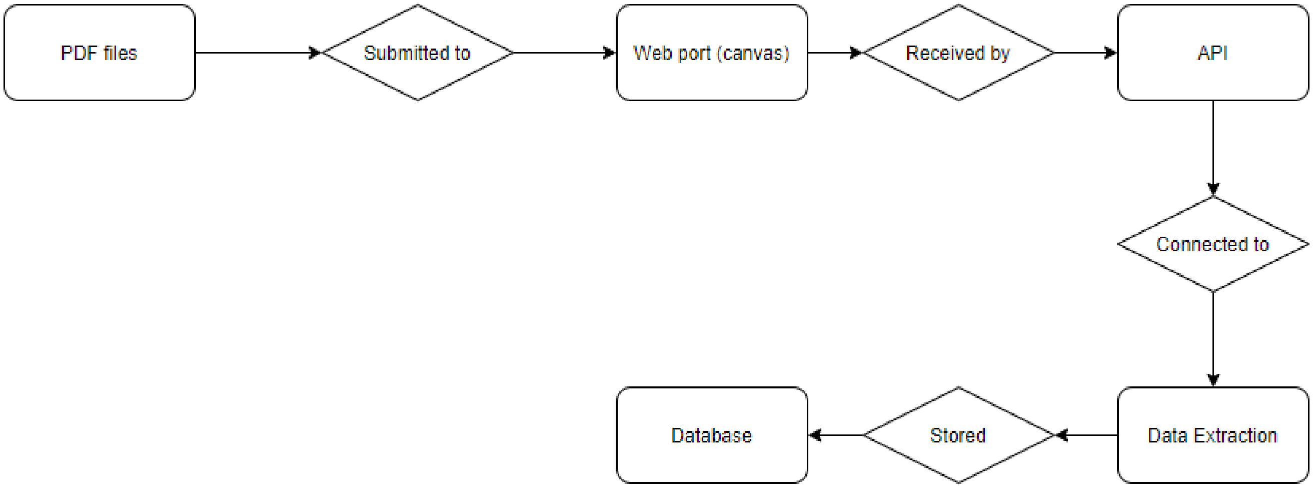
- Design Diagram
- Database
- Data Extraction
- Data Saving and Database API
- Framework
- Test Cases

Design Diagram

Version 2.0



Version 1.0

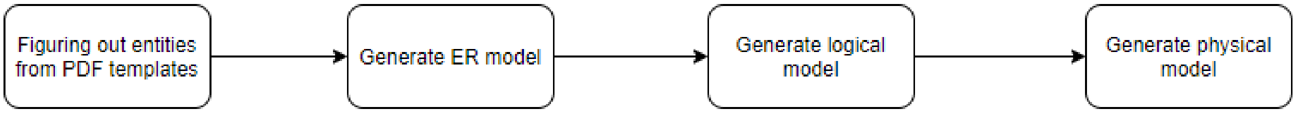


- The tasks for sprint2 include the design of the database, data extraction, and the integration between these two modules.
- The connection with the web port and API and some optimization measures of the whole system will be performed in sprint 3.

Database

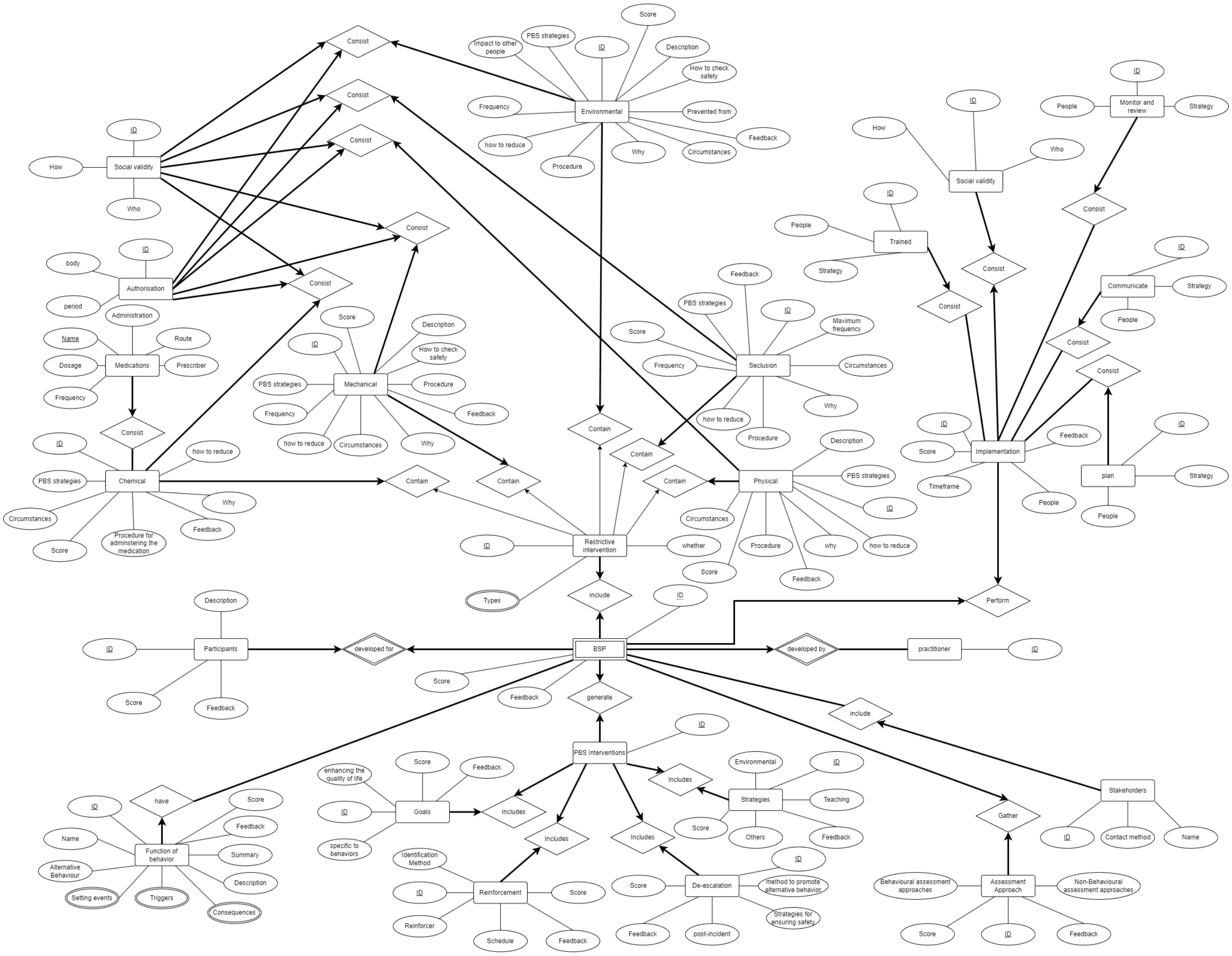
Catalog

Database Design Diagram

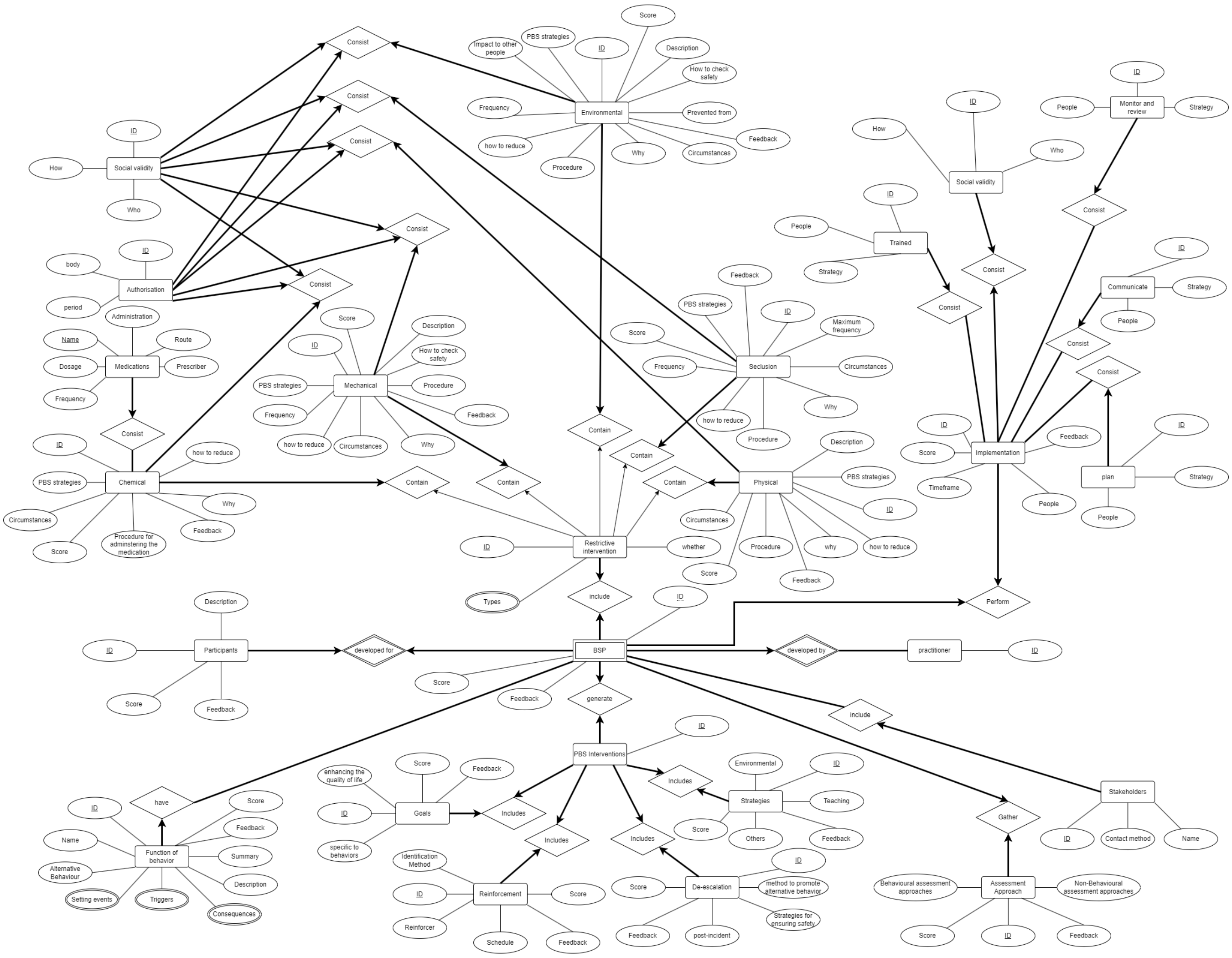


ER Model

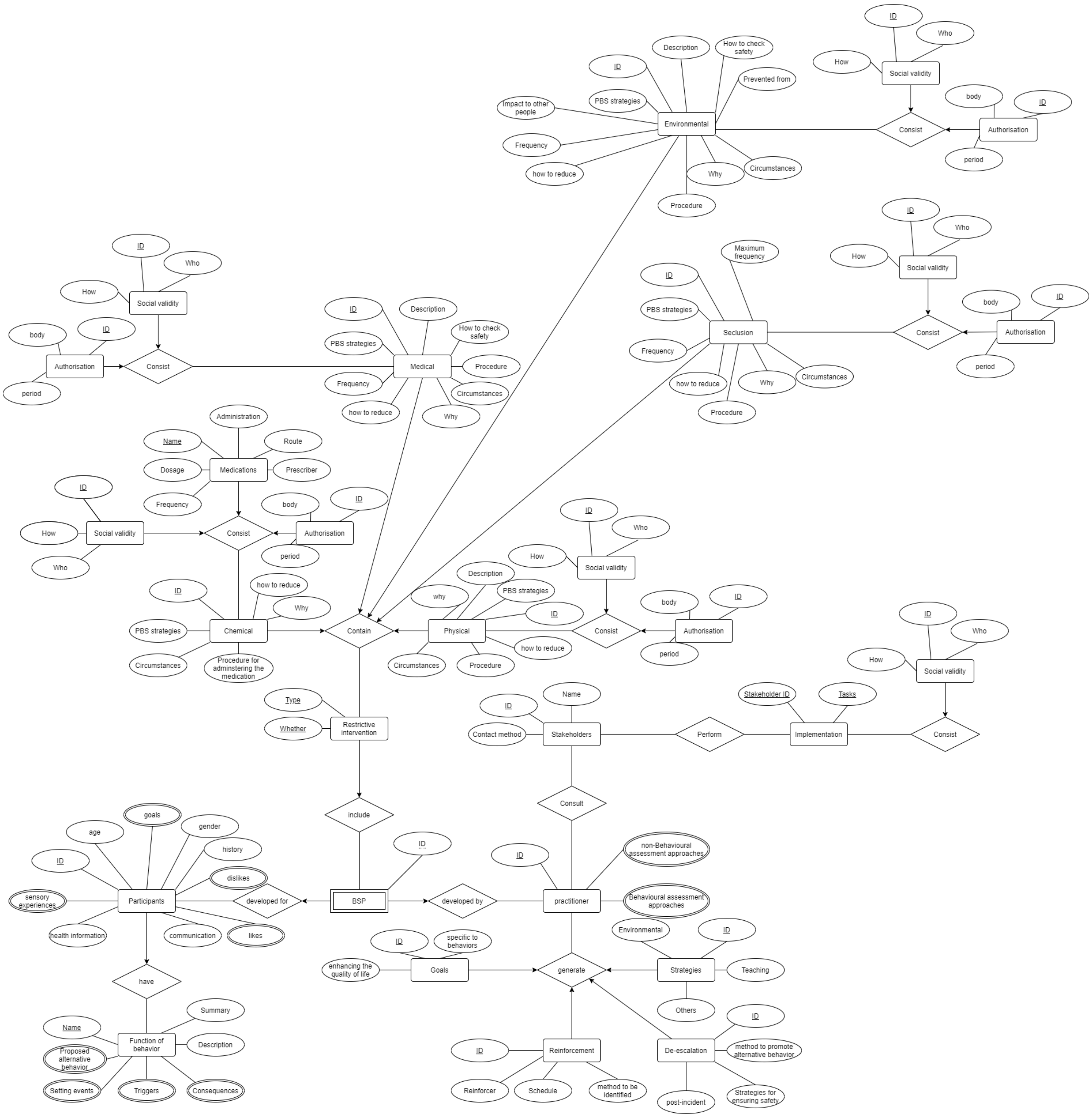
Version 3.0



Version 2.0


Consist
Consist
Consist
Consist
Consist
Score
Impact to other people
PBS strategies
ID
Description
How to check safety
Frequency
Environmental
Prevented from
Feedback
how to reduce
Why
Circumstances
Procedure
ID
How
Social validity
Who
ID
body
Authorisation
period
Administration
Name
Route
Dosage
Medications
Prescriber
Frequency
Consist
ID
how to reduce
PBS strategies
Chemical
Why
Circumstances
Feedback
Score
Procedure for adminstering the medication
Score
Description
ID
How to check safety
PBS strategies
Mechanical
Procedure
Frequency
Feedback
how to reduce
Circumstances
Why
Contain
Contain
Contain
Contain
Contain
Feedback
PBS strategies
ID
Maximum frequency
Score
Frequency
Seclusion
Circumstances
Why
how to reduce
Procedure
Description
Physical
PBS strategies
ID
Circumstances
Procedure
why
how to reduce
Score
Feedback
ID
Restrictive intervention
whether
Types
include
ID
Description
ID
Participants
developed for
BSP
developed by
practitioner
ID
Score
Feedback
Score
Feedback
generate
Perform
ID
How
Social validity
Who
ID
People
Monitor and review
Strategy
ID
People
Trained
Strategy
Consist
Consist
Consist
ID
Communicate
Strategy
People
Consist
Consist
ID
Score
Implementation
Feedback
Timeframe
People
plan
ID
Strategy
People
include
ID
PBS Interventions
Score
Feedback
enhancing the quality of life
ID
Goals
Includes
specific to behaviors
Includes
Environmental
ID
Strategies
Teaching
Score
Others
Feedback
Includes
Identification Method
ID
Reinforcement
Score
Reinforcer
Schedule
Feedback
Includes
ID
Score
De-escalation
method to promote alternative behavior
Feedback
post-incident
Strategies for ensuring safety
have
Score
ID
Feedback
Name
Function of behavior
Summary
Alternative Behaviour
Description
Setting events
Triggers
Consequences
Stakeholders
ID
Contact method
Name
Gather
Behavioural assessment approaches
Assessment Approach
Non-Behavioural assessment approaches
Score
ID
Feedback

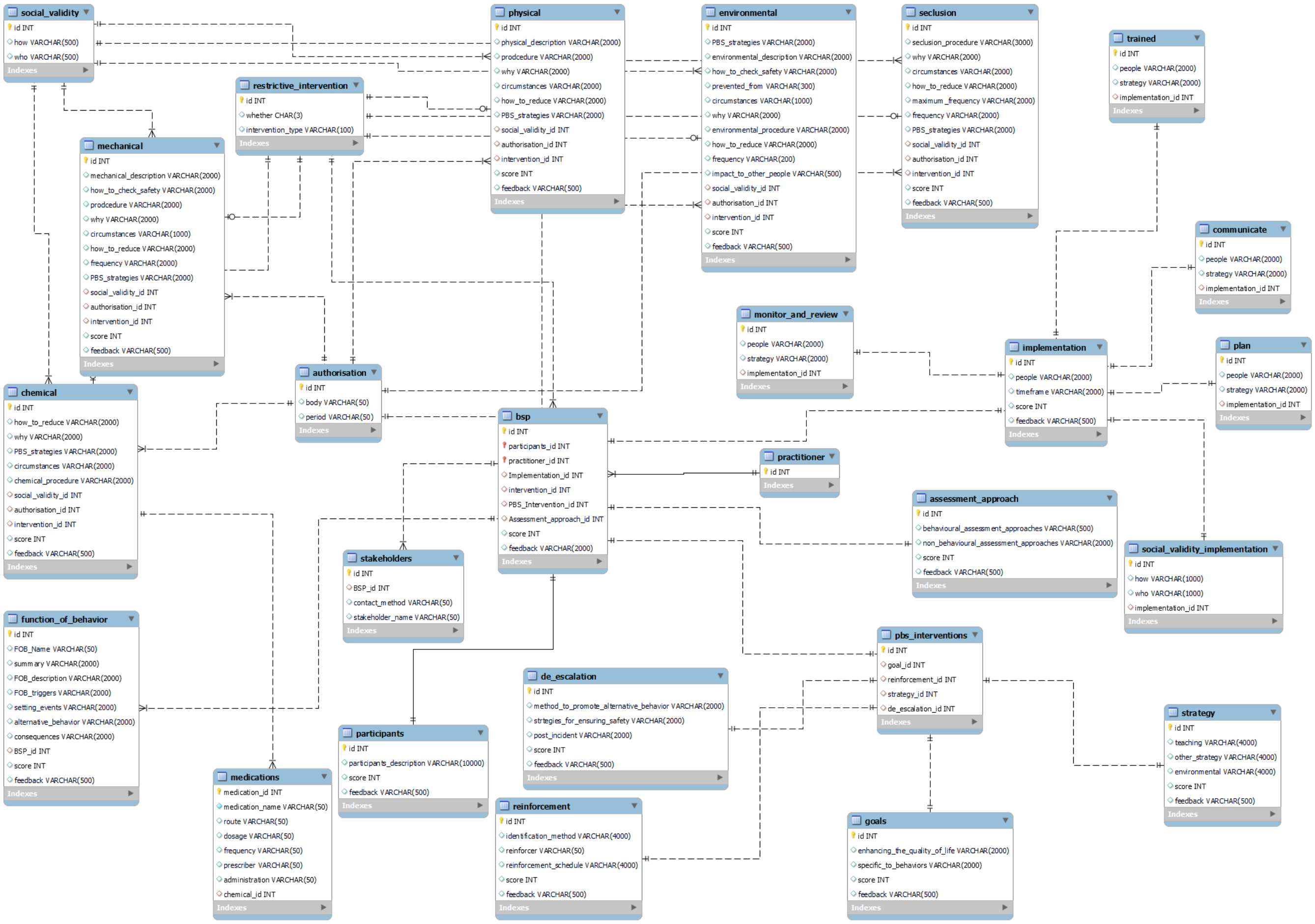
Version 1.0


ID
Who
How
Social validity
Consist
body
ID
Authorisation
period
Description
How to check safety
ID
Prevented from
PBS strategies
Impact to other people
Environmental
Frequency
how to reduce
Why
Circumstances
Procedure
ID
Who
How
Social validity
Maximum frequency
ID
PBS strategies
Seclusion
Consist
body
ID
Authorisation
Frequency
how to reduce
Why
Circumstances
period
Procedure
ID
Who
How
Social validity
body
ID
Authorisation
Consist
period
ID
Description
How to check safety
PBS strategies
Medical
Procedure
Frequency
Circumstances
how to reduce
Why
Administration
Name
Route
Dosage
Medications
Prescriber
Frequency
ID
body
ID
How
Social validity
Consist
Authorisation
Who
period
ID
Who
How
Social validity
how to reduce
ID
Why
PBS strategies
Chemical
Contain
Circumstances
Procedure for adminstering the medication
Description
why
PBS strategies
ID
Physical
Consist
body
ID
Authorisation
how to reduce
period
Circumstances
Procedure
ID
Who
How
Social validity
Type
Whether
Restrictive intervention
ID
Name
Stakeholder ID
Tasks
Contact method
Stakeholders
Perform
Implementation
Consist
Consult
include
goals
age
gender
ID
history
ID
dislikes
non-Behavioural assessment approaches
sensory experiences
Participants
developed for
BSP
developed by
practitioner
Behavioural assessment approaches
health information
communication
likes
Environmental
ID
ID
specific to behaviors
have
enhancing the quality of life
Goals
generate
Strategies
Teaching
Others
Summary
Name
Function of behavior
Description
Proposed alternative behavior
ID
ID
Reinforcement
De-escalation
method to promote alternative behavior
Setting events
Triggers
Consequences
Reinforcer
Schedule
method to be identified
post-incident
Strategies for ensuring safety

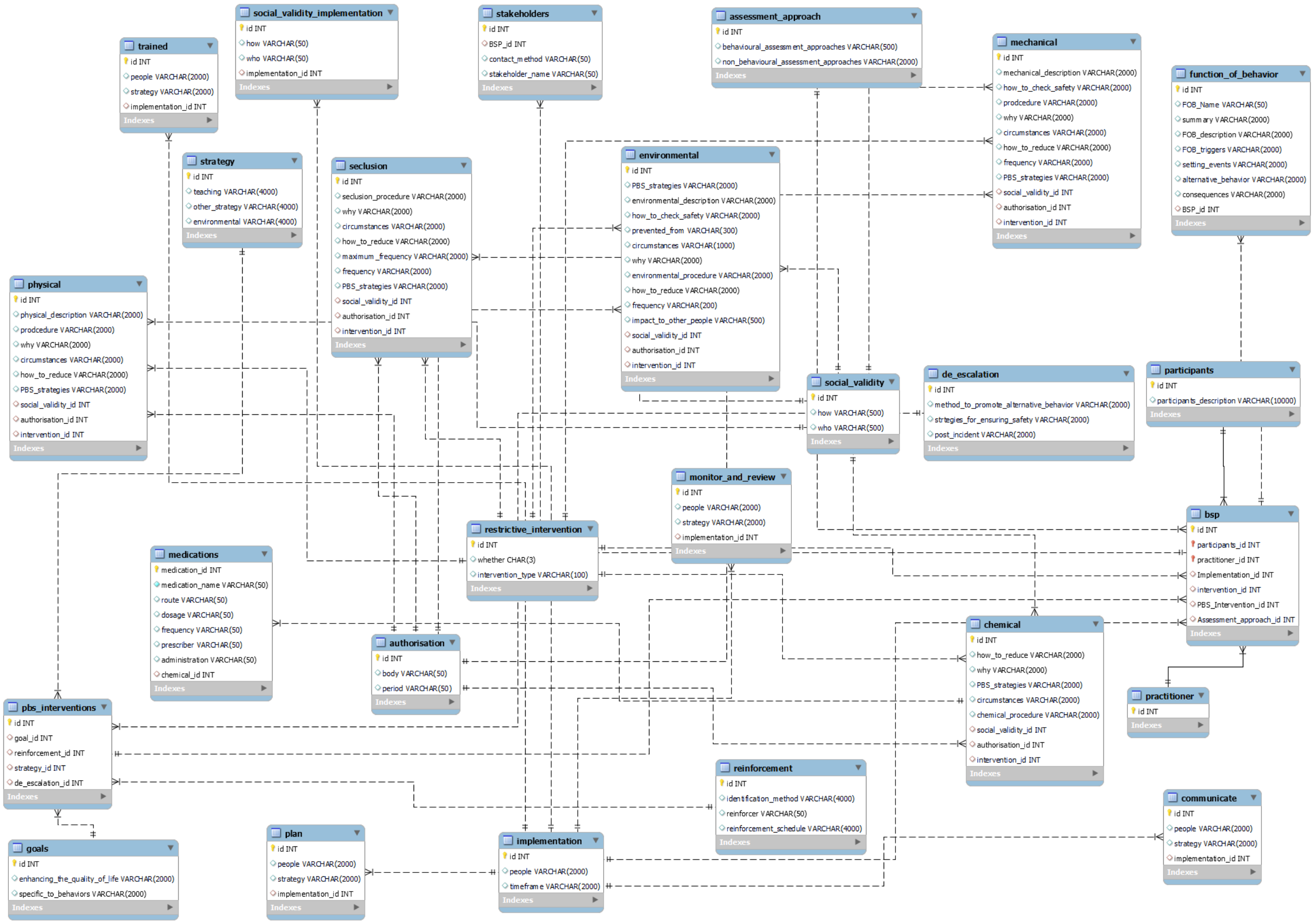
Physical Model

Note: The difference between a logical model and a physical model is whether the data type is determined. For convenience, here only shows the physical model.

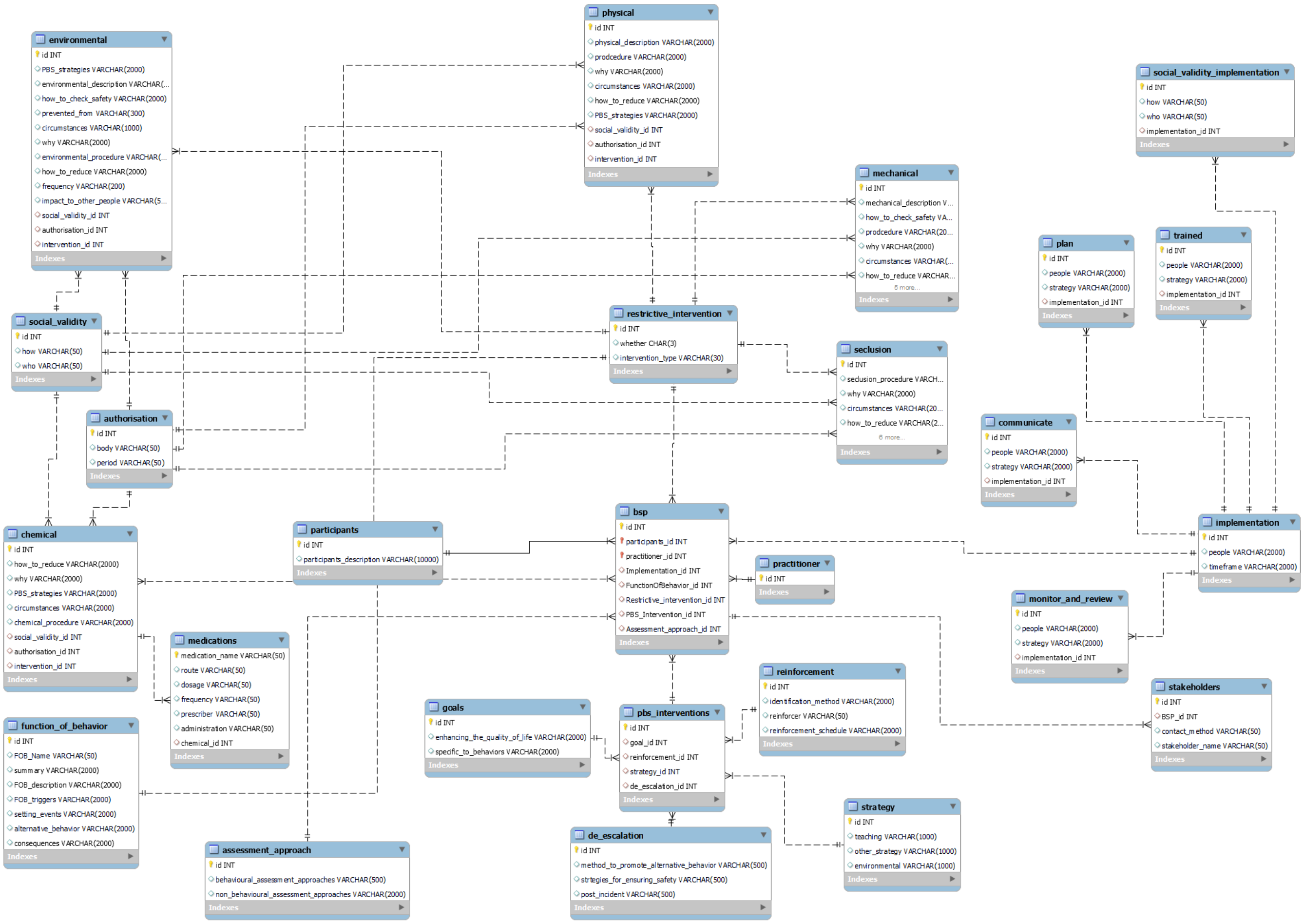
Version 3.0



Version 2.0


trained
id INT
people VARCHAR(2000)
strategy VARCHAR(2000)
implementation_id INT
Indexes
social_validity_implementation
id INT
how VARCHAR(50)
who VARCHAR(50)
implementation_id INT
Indexes
stakeholders
id INT
BSP_id INT
contact_method VARCHAR(50)
stakeholder_name VARCHAR(50)
Indexes
assessment_approach
id INT
behavioural_assessment_approaches VARCHAR(500)
non_behavioural_assessment_approaches VARCHAR(2000)
Indexes
mechanical
id INT
mechanical_description VARCHAR(2000)
how_to_check_safety VARCHAR(2000)
prodcedure VARCHAR(2000)
why VARCHAR(2000)
circumstances VARCHAR(2000)
how_to_reduce VARCHAR(2000)
frequency VARCHAR(2000)
PBS_strategies VARCHAR(2000)
social_validity_id INT
authorisation_id INT
intervention_id INT
Indexes
function_of_behavior
id INT
FOB_Name VARCHAR(50)
summary VARCHAR(2000)
FOB_description VARCHAR(2000)
FOB_triggers VARCHAR(2000)
setting_events VARCHAR(2000)
alternative_behavior VARCHAR(2000)
consequences VARCHAR(2000)
BSP_id INT
Indexes
strategy
id INT
teaching VARCHAR(4000)
other_strategy VARCHAR(4000)
environmental VARCHAR(4000)
Indexes
seclusion
id INT
seclusion_procedure VARCHAR(2000)
why VARCHAR(2000)
circumstances VARCHAR(2000)
how_to_reduce VARCHAR(2000)
maximum_frequency VARCHAR(2000)
frequency VARCHAR(2000)
PBS_strategies VARCHAR(2000)
social_validity_id INT
authorisation_id INT
intervention_id INT
Indexes
environmental
id INT
PBS_strategies VARCHAR(2000)
environmental_description VARCHAR(2000)
how_to_check_safety VARCHAR(2000)
prevented_from VARCHAR(300)
circumstances VARCHAR(1000)
why VARCHAR(2000)
environmental_procedure VARCHAR(2000)
how_to_reduce VARCHAR(2000)
frequency VARCHAR(200)
impact_to_other_people VARCHAR(500)
social_validity_id INT
authorisation_id INT
intervention_id INT
Indexes
physical
id INT
physical_description VARCHAR(2000)
prodcedure VARCHAR(2000)
why VARCHAR(2000)
circumstances VARCHAR(2000)
how_to_reduce VARCHAR(2000)
PBS_strategies VARCHAR(2000)
social_validity_id INT
authorisation_id INT
intervention_id INT
Indexes
social_validity
id INT
how VARCHAR(500)
who VARCHAR(500)
Indexes
de_escalation
id INT
method_to_promote_alternative_behavior VARCHAR(2000)
strtegies_for_ensuring_safety VARCHAR(2000)
post_incident VARCHAR(2000)
Indexes
participants
id INT
participants_description VARCHAR(10000)
Indexes
monitor_and_review
id INT
people VARCHAR(2000)
strategy VARCHAR(2000)
implementation_id INT
Indexes
restrictive_intervention
id INT
whether CHAR(3)
intervention_type VARCHAR(100)
Indexes
bsp
id INT
participants_id INT
practitioner_id INT
Implementation_id INT
intervention_id INT
PBS_Intervention_id INT
Assessment_approach_id INT
Indexes
medications
medication_id INT
medication_name VARCHAR(50)
route VARCHAR(50)
dosage VARCHAR(50)
frequency VARCHAR(50)
prescriber VARCHAR(50)
administration VARCHAR(50)
chemical_id INT
Indexes
authorisation
id INT
body VARCHAR(50)
period VARCHAR(50)
Indexes
chemical
id INT
how_to_reduce VARCHAR(2000)
why VARCHAR(2000)
PBS_strategies VARCHAR(2000)
circumstances VARCHAR(2000)
chemical_procedure VARCHAR(2000)
social_validity_id INT
authorisation_id INT
intervention_id INT
Indexes
practitioner
id INT
Indexes
pbs_interventions
id INT
goal_id INT
reinforcement_id INT
strategy_id INT
de_escalation_id INT
Indexes
reinforcement
id INT
identification_method VARCHAR(4000)
reinforcer VARCHAR(50)
reinforcement_schedule VARCHAR(4000)
Indexes
communicate
id INT
people VARCHAR(2000)
strategy VARCHAR(2000)
implementation_id INT
Indexes
goals
id INT
enhancing_the_quality_of_life VARCHAR(2000)
specific_to_behaviors VARCHAR(2000)
Indexes
plan
id INT
people VARCHAR(2000)
strategy VARCHAR(2000)
implementation_id INT
Indexes
implementation
id INT
people VARCHAR(2000)
timeframe VARCHAR(2000)
Indexes

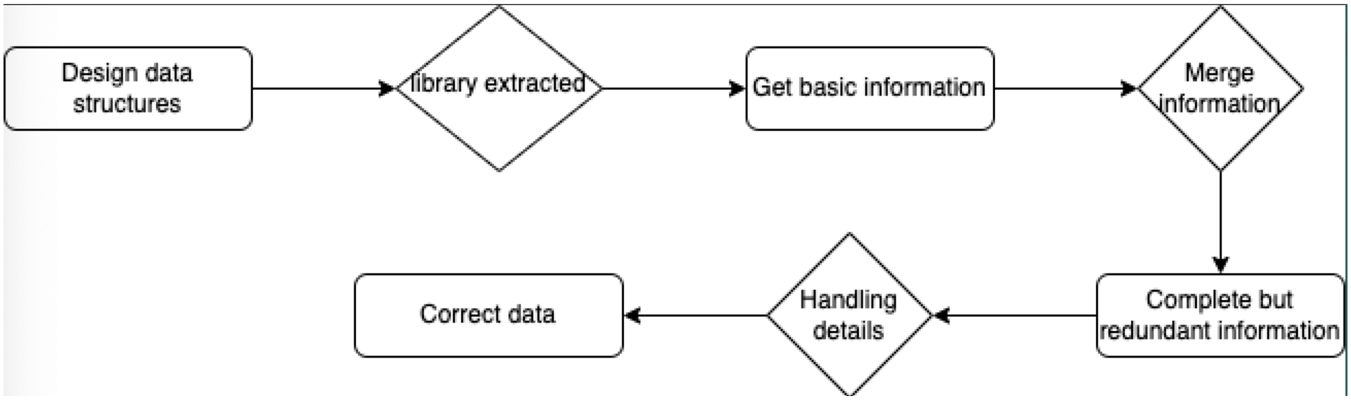
Version 1.0


environmental
id INT
PBS_strategies VARCHAR(2000)
environmental_description VARCHAR(...
how_to_check_safety VARCHAR(2000)
prevented_from VARCHAR(300)
circumstances VARCHAR(1000)
why VARCHAR(2000)
environmental_procedure VARCHAR(...
how_to_reduce VARCHAR(2000)
frequency VARCHAR(200)
impact_to_other_people VARCHAR(5...
social_validity_id INT
authorisation_id INT
intervention_id INT
Indexes
physical
id INT
physical_description VARCHAR(2000)
prodcedure VARCHAR(2000)
why VARCHAR(2000)
circumstances VARCHAR(2000)
how_to_reduce VARCHAR(2000)
PBS_strategies VARCHAR(2000)
social_validity_id INT
authorisation_id INT
intervention_id INT
Indexes
social_validity_implementation
id INT
how VARCHAR(50)
who VARCHAR(50)
implementation_id INT
Indexes
mechanical
id INT
mechanical_description V...
how_to_check_safety VA...
prodcedure VARCHAR(20...
why VARCHAR(2000)
circumstances VARCHAR(...
how_to_reduce VARCHAR...
5 more...
Indexes
plan
id INT
people VARCHAR(2000)
strategy VARCHAR(2000)
implementation_id INT
Indexes
trained
id INT
people VARCHAR(2000)
strategy VARCHAR(2000)
implementation_id INT
Indexes
social_validity
id INT
how VARCHAR(50)
who VARCHAR(50)
Indexes
restrictive_intervention
id INT
whether CHAR(3)
intervention_type VARCHAR(30)
Indexes
seclusion
id INT
seclusion_procedure VARCH...
why VARCHAR(2000)
circumstances VARCHAR(20...
how_to_reduce VARCHAR(2...
6 more...
Indexes
authorisation
id INT
body VARCHAR(50)
period VARCHAR(50)
Indexes
communicate
id INT
people VARCHAR(2000)
strategy VARCHAR(2000)
implementation_id INT
Indexes
bsp
id INT
participants_id INT
practitioner_id INT
Implementation_id INT
FunctionOfBehavior_id INT
Restrictive_intervention_id INT
PBS_Intervention_id INT
Assessment_approach_id INT
Indexes
participants
id INT
participants_description VARCHAR(10000)
Indexes
implementation
id INT
people VARCHAR(2000)
timeframe VARCHAR(2000)
Indexes
chemical
id INT
how_to_reduce VARCHAR(2000)
why VARCHAR(2000)
PBS_strategies VARCHAR(2000)
circumstances VARCHAR(2000)
chemical_procedure VARCHAR(2000)
social_validity_id INT
authorisation_id INT
intervention_id INT
Indexes
practitioner
id INT
Indexes
monitor_and_review
id INT
people VARCHAR(2000)
strategy VARCHAR(2000)
implementation_id INT
Indexes
medications
medication_name VARCHAR(50)
route VARCHAR(50)
dosage VARCHAR(50)
frequency VARCHAR(50)
prescriber VARCHAR(50)
administration VARCHAR(50)
chemical_id INT
Indexes
reinforcement
id INT
identification_method VARCHAR(2000)
reinforcer VARCHAR(50)
reinforcement_schedule VARCHAR(2000)
Indexes
stakeholders
id INT
BSP_id INT
contact_method VARCHAR(50)
stakeholder_name VARCHAR(50)
Indexes
goals
id INT
enhancing_the_quality_of_life VARCHAR(2000)
specific_to_behaviors VARCHAR(2000)
Indexes
pbs_interventions
id INT
goal_id INT
reinforcement_id INT
strategy_id INT
de_escalation_id INT
Indexes
function_of_behavior
id INT
FOB_Name VARCHAR(50)
summary VARCHAR(2000)
FOB_description VARCHAR(2000)
FOB_triggers VARCHAR(2000)
setting_events VARCHAR(2000)
alternative_behavior VARCHAR(2000)
consequences VARCHAR(2000)
Indexes
strategy
id INT
teaching VARCHAR(1000)
other_strategy VARCHAR(1000)
environmental VARCHAR(1000)
Indexes
assessment_approach
id INT
behavioural_assessment_approaches VARCHAR(500)
non_behavioural_assessment_approaches VARCHAR(2000)
Indexes
de_escalation
id INT
method_to_promote_alternative_behavior VARCHAR(500)
strtegies_for_ensuring_safety VARCHAR(500)
post_incident VARCHAR(500)
Indexes

Data Extraction

Catalog

Diagram


Design data structures
library extracted
Get basic information
Merge information
Complete but redundant information
Handling details
Correct data

Third-party library

We found three third-party libraries at the beginning of the project. We have conducted preliminary understanding and testing of these three libraries respectively.

library	pro	con
pdfplumber	The features of pdfplumber are very helpful for us. It can extract the table content by page or the table content of all pages. It can also store the extracted content into files of various formats. We originally used pdfplumber as our main 3rd party library.	Pdfplumber performed poorly when extracting some non-standard complex tables. Garbled characters may occur. As well as the extracted text is not clean enough, we need to spend a lot of time on the extracted text to separate lines and delete some unnecessary symbols.
camelot	camelot didn't do well in the beginning. The results it extracts seem to be inferior to pdfplumber. But after a careful study of the library, we found that this is because we don't know enough about the library's performance. When we read more about the operation instructions of camelot, we found that as long as the line_scale parameter in camelot is adjusted, camelot can accurately identify the content of 99% of the table, so we decided to use camelot as the main third party library.	Camelot has a very tricky trouble. When the table is only a single cell, Camelot will ignore the table and think that the table is only text.
pyPDF2	We found that pyPDF2 can disassemble and merge pdf files, and the library mainly operates on the whole file. It can also get some important properties of pdf.	For extracting table content, it can extract all the text in the pdf file, but not only the text in the table in the pdf file.

Method

As we have two members working for dataExtraction, we firstly separately using different libraries to extract data to get some ideas and have initial results. Then we find that nearly all the important information is stored in tables, so the main task is how to extract information from tables and classify them.

Then, we tried different approaches we can came out to extract data , such as ,adding many if-else condition to extract table one by one(code is complex and not regular,not good for further use abandoned),try to use one function to process multiple tables(the tables does have something in common. however, it’s too difficult to classify every detail in different tables using one function ,abandoned).

As the main problem of processing tables is the pages may divide the table into two or three part which breaks the original table.So one of us came out of a good idea that, before extracting information of the tables of different pages,we can firstly bring the separate tables together to form the original one. Then we use different functions to process them.This part is will need camelot library.

As for detecting bold text, we use pdfplumber to filter the text that is not bold,then we can know which option the user had choose.As for some tables that can not be detected by the library.We used text detection and add missing information to dictionary.Although,there are still limits for these method,but common situations are considered.

The information will all be stored in dictionary waiting for data transmission test.

Challenge and solution

Sprint1

	challenge	solution
1	To extract the text in the table, choose to convert all the text into text format and then use brute force to obtain the content, or choose to use a third-party library	After trying several third-party libraries and text extraction methods, we decided to use a third-party library as the direction of information extraction. Although the content extracted by the third-party library has different degrees of bugs, it seems to be more lightweight and flexible than the text extraction method.
2	Among the alternative third-party libraries, which one is more suitable for us	The two people responsible for the information extraction task chose a library to try, and finally we found that the information extracted by camelot was more accurate and comprehensive due to the adjustable parameters.
3	As the final input data to the database, which format should we adopt that is convenient for the database and easy for developers to use at the same time?	At the beginning we considered JSON, EXCEL, CSV and python's dict. We ruled out excel and csv after further discussions with database members. Although third-party libraries can automatically generate files in JSON format, we need to do a lot of processing on the contents of the files, which means that we cannot directly use the transfer function of third-party libraries, so we decided to store the extracted information by manual processing and manual storage.

Sprint2

	challenge	solution
1	We have made initial progress in extracting tabular information, but since the scale_line parameter has a very large range of values, we need to choose a value with the best generalization ability to apply to most documents.	We take every 10 values as an option, from 10 to 150 (the maximum value of this parameter is 150). Then we delineated the value range of 30-50. Finally, after repeated attempts, we chose 40 as the parameter to determine the value. But 40 is not a perfect value, in fact there is no perfect value, but 40 can be widely used in most of the needs, we still need to make further adjustments to the program
2	When a complete form is divided into two pages by pdf, the third-party library will store the complete form as two forms.	We added the ability to merge tables. When the first row of a table is not a table problem and the last or penultimate row of the previous table of this table is a problem, we need to merge the two tables into one.
3	Third-party libraries cannot read single-line tables, and third-party libraries default to single-line tables as text.	We have additionally added the ability to read text, once there is no table on the page, read the text content and add it to the last table on the previous page.

Sprint3

	challenge	solution
1	We found that the order of the problems in the two different test files was reversed, we need more details to understand if it is a file error or need to handle the situation	The change of files is very fast, we need to be flexible to deal with many possibilities, but since we only have two files for testing, we have added an if statement to deal with the order of these two files. We also need to discuss more flexible ways to deal with more situations that may arise.
2	There is a position in the file that the user needs to select where the user has not selected any of the options. We would like to know more details in order to cope.	We learned that this situation is possible, so we adjusted the code part of this option. If the user chooses nothing, the value part corresponding to the key value is empty.
3	We found that parts of the function module can be repeated if the user needs it. There are also some tables that can be arbitrarily added by the user. But we read the table by a fixed number of questions.	We tried to write code to deal with recurring function modules, but gave up for the time being because there were no testable files. but the possibility for users to add multiple lines has been improved.

Data Saving and Database API

- Data saving module is used for saving the extracted data from the data extraction module into the database.
- Database API is used for retrieving data from the database. The following developers can use the API to extract data stored in the database and perform Natural Language Processing.

- Library used

Pymysql is a powerful pure-python MySQL client library. It can be used for creating, reading, updating and deleting the content in the database. In this project, we mainly focus on the reading and updating function.

- An example of forming connection with a database, updating and reading the information in the database

```
import pymysql.cursors

# Connect to the database
connection = pymysql.connect(host='localhost',
                             user='user',
                             password='passwd',
                             database='db',
                             cursorclass=pymysql.cursors.DictCursor)

with connection:
    with connection.cursor() as cursor:
        # Updating the database
        sql = "INSERT INTO `users` (`email`, `password`) VALUES (%s, %s)"
        cursor.execute(sql, ('webmaster@python.org', 'very-secret'))

    connection.commit()

    with connection.cursor() as cursor:
        # Read a record
        sql = "SELECT `id`, `password` FROM `users` WHERE `email`=%s"
        cursor.execute(sql, ('webmaster@python.org',))
        result = cursor.fetchone()
        print(result)
```

Framework

- Flask

Flask is a popular web framework written in Python. As it does not require any special tools or libraries, it is classified as a microframework. No database abstraction layer, form validation, or any other components are included in this framework. However, it allows extensions and can add application features as if they were implemented in Flask itself, which is quite flexible.

- Components
    - Werkzeug
    - Jinja
    - MarkupSafe
    - ItsDangerous

- Example

```
from flask import Flask
app = Flask(__name__)

@app.route("/")
def hello() -> str:
    return "Hello World"


if __name__ == "__main__":
    app.run(debug=False)
```

Test Cases

Acceptance Criteria #	Test case #	Test case	Pre Requirements	Test steps	Test data	Expected result
1, 2	1.1	Data extraction	Download the required third-party libraries	- Running the data extraction program - Check the information stored in the dictionary	PBSP Summary Document (Draft V3 MV 170822) - QLD Model Plan - No Comments. pdf	All the data is stored in the dictionary correctly. Test result link: Task 1.1
1, 2	1.2	Database Creation	MySQL Database Installation	- Configure the MySQL database. - Running the programming file named Database_Create.sql to see the results.	N/A	1. A scheme called 'new_schema' should be created from the database. 2. 28 tables with their attributes should be created from the new schema. Test result link; Task 1.2
1, 2	1.3	Storing data into the database	Test case 1.1	- Running the program of data extraction module and data storing module - Check the data stored in the database tables	PBSP Summary Document (Draft V3 MV 170822) - QLD Model Plan - No Comments. pdf	All the data is stored correctly in the database. Test result link: Task 1.3
3, 4	2.1.3	Database API	Test case 1.3	- Keep connection with database. - Run the program named 'DB_API_test.py'. - Check the output files. - Check the txt results with the PBSP pdf document to make sure whether the contents are the same.	Database with stored information from test case 1.3	1. The output files should be six text files named with their page numbers. 2. The text contents should be the same as the PBSP pdf document except for some small errors. Test result link: Task 2.1.3
3, 4	3.1	Integration of the whole system	Test case 1.1, 1.2, 1.3, 2.1.3	- Keep connection with the database - Start the back-end server - Send a PDF file to the back-end port by using Postman - Use database API to retrieve the information stored in the database and store the output data into six text files - Check the content of the new generated text files.	PBSP Summary Document (Draft V3 MV 170822) - QLD Model Plan - No Comments. pdf	1. The output files should be six text files named with their page numbers. 2. The text contents should be the same as the PBSP pdf document except for some small errors. Test result link: Task 3.1