# Appendix I. Experts for incorporating domain knowledge

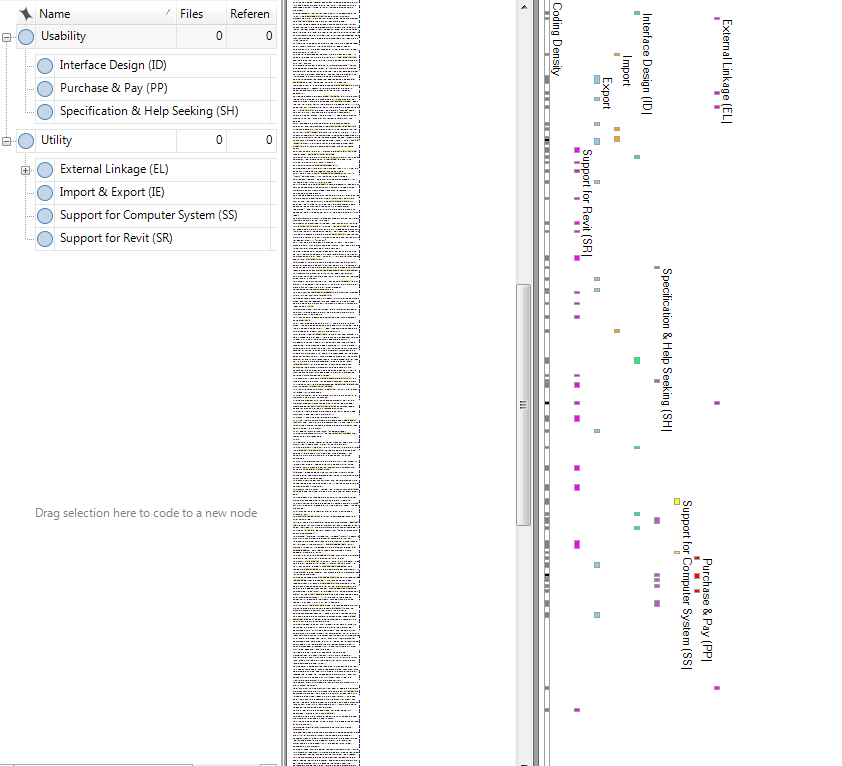
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Experts-related works*** | ***Code*** | ***Institutions*** | ***Position*** | ***Domain*** | ***Experience*** |
| Generate ideas regarding the potential label-related tasks  **(Corresponding to 3.2.1)** | 1 | The University of Hong Kong | M.Phil., PhD student | BIM | >5 years |
| 2 | The University of Hong Kong | M.Phil., PhD student | BIM | >5 years |
| 3 | JieHong Engineering Consulting Co., Ltd. | M.Sc., Senior engineer | BIM | >5 years |
| 4 | The University of Hong Kong | PhD, Assistant Professor | BIM & Computer science (CS) | >15 years |
| 5 | The University of Hong Kong | PhD, Postdoctoral researcher | CS & Machine Learning | >20 years |
| Do labelling work respectively & Reach a consensus  **(Corresponding to 3.2.2)** | 1 | The University of Hong Kong | M.Phil., PhD student | BIM | >5 years |
| 2 | The University of Hong Kong | M.Phil., PhD student | BIM | >5 years |
| Workshop for explaining the significant relationship  **(Corresponding to 3.4.2)** | 1 | The University of Hong Kong | M.Phil., PhD student | BIM | >5 years |
| 2 | The University of Hong Kong | M.Phil., PhD student | BIM | >5 years |
| 3 | JieHong Engineering Consulting Co., Ltd. | M.Sc., Senior engineer | BIM | >5 years |
| 4 | The University of Hong Kong | PhD, Assistant Professor | BIM & Computer science (CS) | >15 years |
| 5 | The University of Hong Kong | PhD, Postdoctoral researcher | CS & Machine Learning | >20 years |

# Appendix II: Raw Materials, Details and Intermediate Products of the Synthetic Approach

|  |  |
| --- | --- |
| ***Name of Sheets*** | ***Data included*** |
| 1. 759 Apps | * App attributes |
| 2. Feature words for sentiment analysis | * Adjectives, adverbs as sentiment feature words |
| 3. 2063 Comments & Sentiments | * Contents of all comments * Sentiments results from 3 models for each comment (Text blob, Google, and Experts) |
| 4. 362 Apps having comments for Correlation Analysis | * Apps’ attributes * Apps’ sentiment score |
| 5. Problems in 928 Negative Comments | * Contents of all negative comments * Problems complained in each negative comments |

The specific contents can be seen in the attached Excel file or downloaded from ‘https://github.com/0AnonymousSite0/A-Domain-Knowledge-Incorporated-Text-Mining-Approach-for-Capturing-User-Needs-on-BIM-Applications’

# Appendix III: Labelling work in NVivo



Types of problems

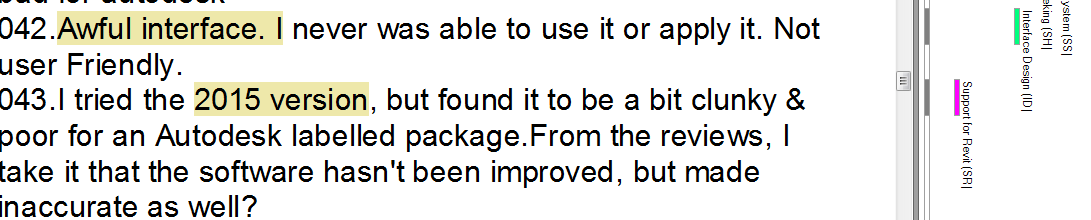
Labels in

each comment

Contents of

comments

Labels for user problems in NVivo (Screen shot)



Labels for user problems in NVivo (Screenshot)

# Appendix IV. Pseudocode for sentiment analysis

|  |  |
| --- | --- |
| Sentiment Analysis for user comments | |
| **Input:** *CommentArray, SentimentArray, SentimentFeatureWords* | |
| **Output:** *SAmodel, SAModelAccuracy* | |
| 1 | **Import** NumPy // Provide functions of Array calculation, RandomChoose**,** Remove. |
| 2 | **Import** TextBlob // Provide functions of Tokenize, RemoveStopwords. |
| 3 | **Import** Google Natural Language **//** Provide functions ofIdentifyLanguage, TranslateLanguage. |
| 4 | *TrainingSetComment*, *TrainingSetSentiment*, *TestingSetComemnt*, *TrainingSetSentiment* ← **RandomChoose** (*CommentArray*, *SentimentArray*) |
| 5 | **Function** **TextProcessing** (*Array*) |
| 6 | **For** *Comment* **in** *Array* **Do** |
| 8 | **if** *English=* **IdentifyLanguage** *Comment* **then** |
| 9 | *Comment* ← **TranslateLanguage** (*Comment*) |
| 10 | **End if** |
| 11 | *WordArray* ← **Tokenize (***Comment*) |
| 12 | *WordArray* ← **RemoveStopwords** (*WordArray*) |
| 13 | *WordArray* ← **ExtractSentimentFeatureWords** (*WordArray*) |
| 14 | *Comment* ← *WordArray* |
| 15 | **End for** |
| 16 | **Return** *Array* |
| 17 | **End Function** |
| 18 | **Function** **ExtractSentimentFeatureWords** (*Array*) |
| 19 | **For** *Word* **in** *Array* **Do** |
| 20 | **if** *Word* **not in** *SentimentFeatureWords then* |
| 21 | *Array* ←**Remove** (*Array, Word*) |
| 22 | **End if** |
| 23 | **End for** |
| 24 | **Return** *Array* |
| 25 | **End Function** |
| 26 | *TrainingSet* ← **TextProcessing** (*TrainingSetComment*) |
| 27 | *TestingSet* ← **TextProcessing** (*TestingSetComment*) |
| 28 | *SAModel* ← **TraniningClassifier** (*Training Set, TrainingSetSentiment*) |
| 29 | *SAModelAccuracy* ← **SAModel** (*Testing Set, TrainingSetSentiment*) |

# Appendix V. Pseudocode for topic modelling

|  |  |
| --- | --- |
| Topic model for identifying user problems | |
| **Input:** *CommentArray, TopicArray, ProblemFeatureWords* | |
| **Output:** *TopicModel, TopicModelAccuracy* | |
| 1 | **Import** NumPy // Provide functions of Array, RandomChoose**,** Remove. |
| 2 | **Import** TextBlob // Provide functions of Tokenize, RemoveStopwords. |
| 3 | **Import** Google Natural Language **//** Provide functions ofIdentifyLanguage, TranslateLanguage. |
| 4 | **Import** Stanford Topic Modelling Toolbox // Provided SLDA algorithm |
| 5 | *TrainingSetComment*, *TrainingSetSentiment*, *TestingSetComemnt*, *TrainingSetSentiment* ← **RandomChoose** (*CommentArray*, *TopicArray*) |
| 6 | **Function** **TextProcessing** (*Array*) |
| 8 | **For** *Comment* **in** *Array* **Do** |
| 9 | **if** *English=* **IdentifyLanguage** *Comment* **then** |
| 10 | *Comment* ← **TranslateLanguage** (*Comment*) |
| 11 | **End if** |
| 12 | *WordArray* ← **Tokenize (***Comment*) |
| 14 | *WordArray* ← **RemoveStopwords** (*WordArray*) |
| 15 | *Comment* ← *WordArray* |
| 16 | **End for** |
| 17 | **Return** *Array* |
| 18 | **End Function** |
| 19 | *TrainingSet* ← **TextProcessing** (*TrainingSetComment*) |
| 20 | *TestingSet* ← **TextProcessing** (*TestingSetComment*) |
| 21 | *TopicModel* ← **TraniningClassifier** (*Training Set, TrainingSetSentiment*) |
| 22 | *TopicModelAccuracy* ← **TopicModel** (*Testing Set, TrainingSetSentiment*) |

# Appendix VI. Comparison of ‘word-problem distribution’ in topic models

|  |  |  |
| --- | --- | --- |
| ***Models***  ***Problems*** | ***Typical Feature words in topic models*** | |
| ***Topic model in this case study*** | ***Topic modelling without domain knowledge***  ***(e.g.*** *Stanford Topic Modelling Toolbox****)*** |
| Problem 1 | Update, work, new, 2017, version  (support for Revit) | File, point, excel, export, project |
| Problem 2 | Import, file, model, imported, exporting  (import & export) | IFC, model, file, data, import |
| Problem 3 | Google, web, earth, site, CAD  (external linkage) | Get, help, autodesk, work, interface |
| Problem 4 | Windows, 64, W7, computer, software  (support for computer system) | Software, file, new, computer, program |
| Problem 5 | Views, interface, click, resolution, ribbon  (interface design) | Site, web, import, google, ribbon |
| Problem 6 | Trail, free, pay, license, buy  (purchase & pay) | Time, project, run, lost, try |
| Problem 7 | Contact, email, problem, message, reply  (specification & help-seeking) | Work, parameters, product, views, support |

*Note: The topic model without domain knowledge is provided by* *Stanford Topic Modelling Toolbox, and the Unsupervised-Latent Dirichlet Allocation based topic models from different sources (e.g. Google, Natural Language Toolkit, and Stanford Topic Modelling Toolbox) are similar.*