

NOVEMBER 2021

# PROJECT REPORT

## A KNOWLEDGE-BASED RECOMMENDATION SYSTEM FOR PROJECT RISK MANAGEMENT: ONTOLOGY LEARNING APPROACH

Prepared by: Haddad Hamed Sofien  
Ikram Ben Charrada  
Mohamed Aymen Barhoumi  
Essia Houaoui  
Chiheb Jamazi  
Sarra Abderrazak  
Ichrak Houidi

# 02

# SUMMARY

1. NATURAL LANGUAGE PROCESSING
2. DATA EXTRACTION
3. PROCESS EXTRACTION
4. TEXT PROCESSING
5. PREPROCESSING TECHNIQUES
  - 5.1 CASE CHANGING
  - 5.2 REMOVAL OF PUNCTUATION
  - 5.3 NUMBER CONVERSION TO LETTERS
  - 5.4 REMOVING STOPWORDS
  - 5.5 TOKENIZATION
  - 5.6 LEXICON NORMALIZATION
  - 5.7. PART OF SPEECH (POS) TAGGING
  - 5.8. CHUNKING
  - 5.9. EXTRACTING N-GRAMS
    - 5.9.1 EXTRACTION BI-GRAMS
    - 5.9.2. EXTRACTING tri-GRAMS
  - 5.10. VECTORIZATION FEATURE ENGINEERING (TF-IDF)
  - 5.11. NAMED ENTITY EXTRACTION
  - 5.12. RDF TRIPLES
  - 5.13. IS A / SUCH AS RELATION EXTRACTION
  - 5.14. HAVE / INCLUDE RELATION EXTRACTION
6. OWL ONTOLOGY
  - 6.1 ONTOLOGY EVALUATION
7. Ontology MATCHING
8. Deployment

**03**

# 1. NATURAL LANGUAGE PROCESSING

In any Machine learning task, cleaning or preprocessing the data is as important as model building. Text data is one of the most unstructured forms of available data and when comes to deal with Human language then it's too complex. Have you ever wondered how Alexa, Siri, Google assistant can understand, process, and respond in Human language.

NLP is a technology that works behind it where before any response lots of text preprocessing takes place. This tutorial will study the main text preprocessing techniques that you must know to work with any text data.

Natural Language Processing is a branch of Artificial Intelligence that analyzes, processes, and efficiently retrieves information text data. By utilizing the power of NLP one can solve a huge range of real-world problems which include summarizing documents, title generator, caption generator, fraud detection, speech recognition, recommendation system, machine translation, etc.

## 04

## 2. DATA EXTRACTION

Using the PyPDF2 library, we transformed the pdf file into text. In addition, the import has been limited to 45 PAGE FROM 309 to 354, and we have transformed it as a list. then cleaned up the text created by remove items such as numbering and headings. Then the text was split into sentences in a data frame for further editing.

```
['9 - PROJECT HUMAN RESOURCE MANAGEMENT\n',  
 'Examples of interpersonal skills that a project manager uses most often include:\n',  
 '•\t Leadership. Successful projects require strong leadership skills. Leadership is important through all \n',  
 'phases of the project life cycle. There are multiple leadership theories defining leadership styles that \n',  
 'should be used as needed for each situation or team. It is especially important to communicate the vision \n',  
 'and inspire the project team to achieve high performance.\n',  
 '•\t Influencing. Because project managers often have little or no direct authority over team members in a \n',  
 'matrix environment, their ability to influence stakeholders on a timely basis is critical to project success. \n',  
 'Key influencing skills include:\n',  
 'o Ability to be persuasive and clearly articulate points and positions;\n',  
 'o High levels of active and effective listening skills;\n',  
 'o Awareness of, and consideration for, the various perspectives in any situation; and\n',  
 'o Gathering relevant and critical information to address important issues and reach agreements \n',  
 'while maintaining mutual trust.\n',  
 '•\t Effective decision making. This involves the ability to negotiate and influence the organization and the \n',  
 'project management team. Some guidelines for decision making include:\n',  
 'o Focus on goals to be served,\n',  
 'o Follow a decision-making process,\n',  
 'o Study the environmental factors,\n',  
 'o Analyze available information,\n',  
 'o Develop personal qualities of the team members,\n',  
 'o Stimulate team creativity, and\n',  
 'o Manage risk.\n',  
 '9.4.3 Manage Project team: outputs\n',  
 '9.4.3.1 change requests\n',  
 'Staffing changes, whether by choice or by uncontrollable events, can affect the rest of the project management \n',
```

## 05

## 3. PROCESS EXTRACTION

After converting the pdf file into text, we went through our text and extracted the main 6 processes of project risk management, and to do so we used the regular expression, we defined patterns for the process, the subprocess, and the content of each subprocess.

We combined the output into one data frame that we will be working on within the next phases.

	process	type	sub process	details
60	control risks	Inputs	Project Management Plan	project management plan includes risk manageme...
61	control risks	Inputs	risk register	risk register has key inputs include identifie...
62	control risks	Inputs	Work Performance data	work performance data related various performa...
63	control risks	Inputs	Work Performance reports	work performance reports take information perf...
64	control risks	tools and techniques	risk reassessment	control risks often results identification new...
65	control risks	tools and techniques	risk Audits	risk audits examine document effectiveness ris...
66	control risks	tools and techniques	Variance and trend Analysis	many control processes employ variance analysi...
67	control risks	tools and techniques	technical Performance Measurement	technical performance measurement compares tec...
68	control risks	tools and techniques	reserve Analysis	throughout execution project some risks may oc...
69	control risks	tools and techniques	Meetings	project risk management should be agenda item ...
70	control risks	outputs	Work Performance Information	work performance information as a control risk...
71	control risks	outputs	change requests	implementing contingency plans workarounds som...
72	control risks	outputs	Project Management Plan updates	approved change requests have effect risk mana...
73	control risks	outputs	Project documents updates	project documents may be updated as a result c...
74	control risks	outputs	organizational Process Assets updates	risk management processes produce information ...

05

## 4. TEXT PREPROCESSING

“

*Text preprocessing is a method to clean the text data and make it ready to feed data to the model. Text data contains noise in various forms like emotions, punctuation, text in a different cases. When we talk about Human Language then, there are different ways to say the same thing, And this is only the main problem we have to deal with because machines will not understand words, they need numbers so we need to convert text to numbers in an efficient manner.*

Libraries used to deal with NLP problems

There are many libraries and algorithms used to deal with NLP-based problems. A regular expression(re) is mostly used library for text cleaning. NLTK(Natural language toolkit) and spacy are the next level library used for performing Natural language tasks like removing stopwords, named entity recognition, part of speech tagging, phrase matching, etc.

NLTK is an old library used for practicing NLP techniques by beginners. Spacy is the latest released library with the most advanced techniques and is mostly used in the production environment so I would like to encourage you to learn both the libraries and experience its power.



# 06

## 5. PREPROCESSING TECHNIQUES

### 5.1. CASE CHANGING

It is a common approach to lowercase everything for the sake of simplicity. It helps to maintain the consistency flow during the NLP tasks and text mining.

'9 - project human resource management\ninclude:\n• leadership. successful pro  
11 \nphases of the project life cycle. t  
hould be used as needed for each situat:  
spire the project team to achieve high p  
or no direct authority over team member:  
timely basis is critical to project suc  
learly articulate points and positions;  
f, and consideration for, the various pe

### 5.2. REMOVAL OF PUNCTUATION

It is important to remove all noise from the data including all the punctuation marks.

'9 project human resource management\nexamples  
hip skills leadership is important through all  
d for each situation or team it is especially i  
agers often have little or no direct authority  
s \nkey influencing skills include\nno ability  
of and consideration for the various perspectiv

# 07

## 5.3. NUMBER CONVERSION TO LETTERS

From the previous step, it's apparent that our text contains unwanted characters other than spaces and punctuations, which are numbers that we have to convert into letters.

“

nine project human resource management examples of integrated skills leadership is important through all phases of the ch situation or team it is especially important to communicate little or no direct authority over team members in a g skills include o ability to be persuasive and clearly r the various perspectives in any situation and o gather

”

```
{ 'a',  
  'about',  
  'above',  
  'after',  
  'again',  
  'against',  
  'ain',  
  'all',  
  'am',  
  'an',  
  'and',  
  'any',  
  'are',  
  'aren',  
  "aren't",
```

## 5.4. REMOVING STOPWORDS

Most text data that we work with is going to contain a lot of words that don't look useful to us. These words, called stopwords, are useful in human speech, but they don't have much to contribute to data analysis. Removing stopwords helps us eliminate noise and distraction from our text data, and also speeds up the time analysis takes (since there are fewer words to process).

```
text= re.sub("\s*\d\s*", " ", text).strip()  
text =re.sub('https?://\S+|www\.\S+', ' ', text)  
text=re.sub("\d+", " ", text)
```

# 08

## 5.5.TOKENIZATION

Tokenization is breaking the raw text into small chunks. Tokenization breaks the raw text into words, sentences called tokens. These tokens help in understanding the context or developing the model for the NLP. The tokenization helps in interpreting the meaning of the text by analyzing the sequence of the words.

	sentences	Tokens
0	nine - project human resource management example	[nine, -, project, human, resource, management...]
1	successful projects require strong leadership	[successful, projects, require, strong, leader...]
2	leadership is important phases project life cycle	[leadership, is, important, phases, project, l...]
3	are multiple leadership theories defining leadership	[are, multiple, leadership, theories, defining...]
4	is especially important communicate vision instead	[is, especially, important, communicate, visio...]
...	...	...
457	a guide project management body knowledge (pmbok)	[a, guide, project, management, body, knowledg...]
458	key benefit process is enables project manager	[key, benefit, process, is, enables, project, ...]
459	inputs, tools techniques, outputs process are	[inputs, , tools, techniques, , outputs, pro...]
460	inputs tools & techniques outputs risk management	[inputs, tools, &, techniques, outputs, risk, ...]
461	a guide project management body knowledge (pmbok)	[a, guide, project, management, body, knowledg...]
462	rows x 2 columns	

## 5.6.LEXICON NORMALIZATION

Lexicon normalization is another step in the text data cleaning process. In the big picture, normalization converts high dimensional features into low dimensional features which are appropriate for any machine learning model. For our purposes here, we're only going to look at lemmatization, a way of processing words that reduces them to their roots.

Lemmatization uses the context in which the word is being used.

	sentences	Tokens	Lemmatization
0	nine - project human resource management examp...	[nine, -, project, human, resource, management...]	[nine, -, project, human, resource, management...]
1	successful projects require strong leadership ...	[successful, projects, require, strong, leader...]	[successful, project, require, strong, leaders...]
2	leadership is important phases project life cy...	[leadership, is, important, phases, project, I...]	[leadership, is, important, phase, project, li...]
3	are multiple leadership theories defining lead...	[are, multiple, leadership, theories, defining...]	[are, multiple, leadership, theory, defining, ...]
4	is especially important communicate vision ins...	[is, especially, important, communicate, visio...]	[is, especially, important, communicate, visio...]
...	...	...	...
457	a guide project management body knowledge (pmb...	[a, guide, project, management, body, knowledg...]	[a, guide, project, management, body, knowledg...]
458	key benefit process is enables project manager...	[key, benefit, process, is, enables, project, ...]	[key, benefit, process, is, enables, project, ...]
459	inputs, tools techniques, outputs process are ...	[inputs, ,, tools, techniques, ,, outputs, pro...]	[input, ,, tool, technique, ,, output, process...]
460	inputs tools & techniques outputs risk managem...	[inputs, tools, &, techniques, outputs, risk, ...]	[input, tool, &, technique, output, risk, mana...]
461	a guide project management body knowledge (pmb...	[a, guide, project, management, body, knowledg...]	[a, guide, project, management, body, knowledg...]

462 rows × 3 columns

# 10

Cleaned_sentences
[nine, -, human, interpersonal, a, often, :, .]
[successful, strong, .]
[important, .]
[multiple, should, as, needed, .]
[especially, important, achieve, high, .]
...
[a, (, pmbok®, ), fifth, three, twenty-seven,
...
[key, high-priority, .]
[..., figure, .]
[&, baseline, environmental, organizational,
e...
[a, (, pmbok®, ), fifth, to, :, sanchez, pmi, ...]

## 5.7.PART OF SPEECH (POS) TAGGING

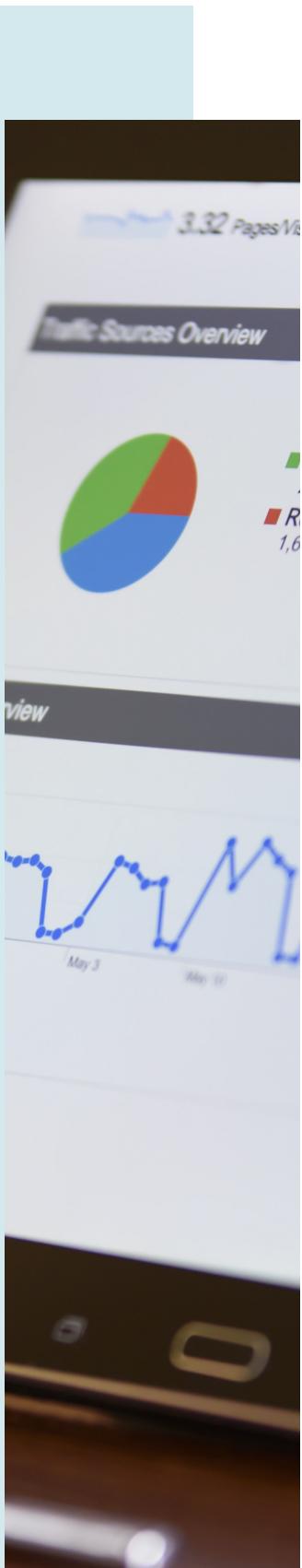
A word's part of speech defines its function within a sentence. A noun, for example, identifies an object. An adjective describes an object. A verb describes an action. Identifying and tagging each word's part of speech in the context of a sentence is called Part-of-Speech Tagging, or POS Tagging.

## 5.8.CHUNKING

Chunking is a process of extracting phrases from unstructured text, which means analyzing a sentence to identify the constituents(Noun Groups, Verbs, verb groups, etc.) However, it does not specify their internal structure, nor their role in the main sentence. It works on top of POS tagging.

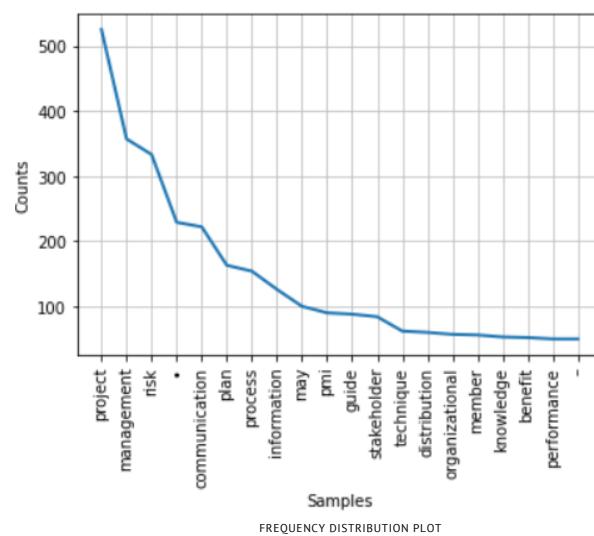
```
(Chunk plan/NN updated/VBN include,/NNS)
are/VBP
to,/NN
resource/NN
management/NN
plan./NN
project/NN
(Chunk documents/NNS updates/VBZ documents/NNS)
updated/VBN
(Chunk include,/NNS are/VBP to:/NNS)
(Chunk issue/NN log,/VBZ roles/NNS)
description,/VBP
staff/NN
assignments./NN
enterprise/NN
factors/NNS
(Chunk updates/NNS enterprise/VBP factors/NNS)
updates/NNS
result/NN
manage/NN
project/NN
team/NN
```



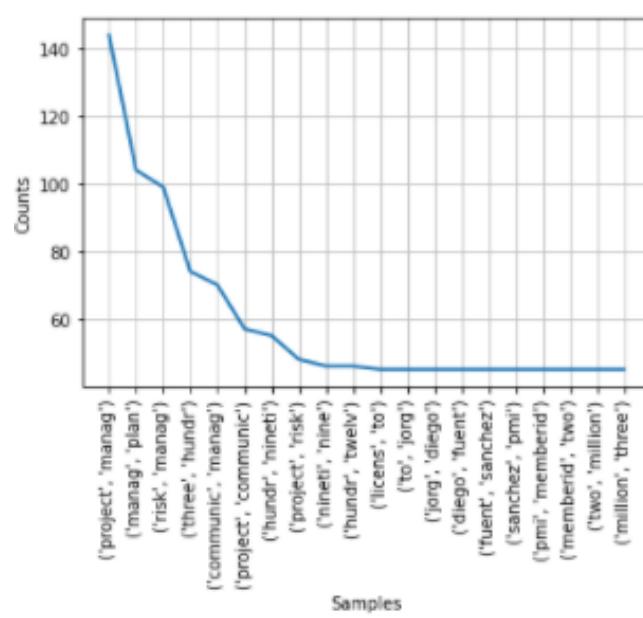


## 5.9. EXTRACTING N-GRAMS

N-grams of texts are extensively used in text mining and natural language processing tasks. They are basically a set of co-occurring words within a given window and when computing the n-grams you typically move one word forward.



### 5.9.1 EXTRACTING BI-GRAMS

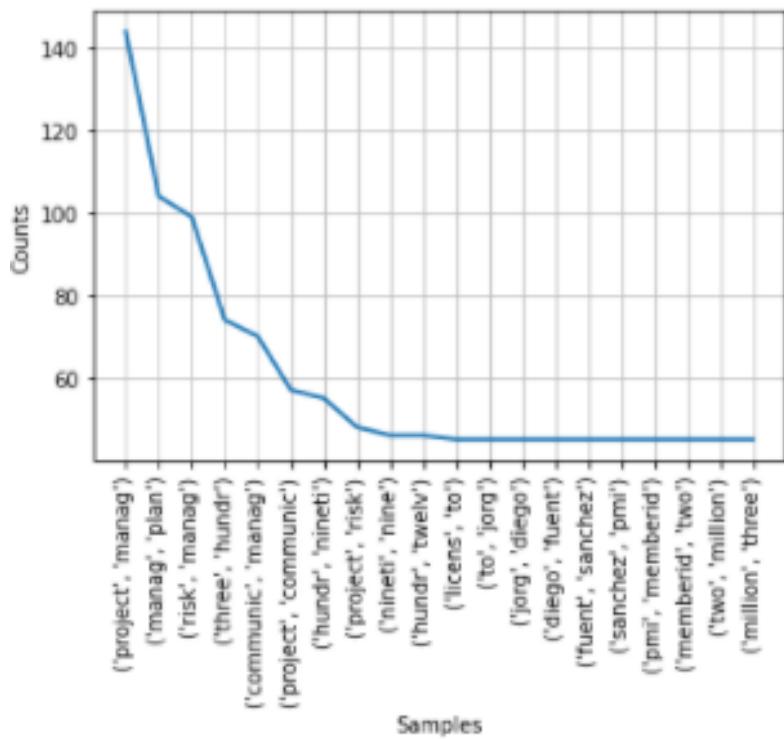


12



“

## 5.9.2. EXTRACTING TRI-GRAMS



”

# 13

## 5.10. VECTORIZATION FEATURE ENGINEERING (TF-IDF)

A statistical measure that evaluates how relevant a word is to a document in a collection of documents.

This is done by multiplying two metrics: how many times a word appears in a document, and the inverse document frequency of the word across a set of documents.

```
'establish': 0.18911656301824686,  
'impact': 0.13075890445485192,  
'may': 0.07885032882853209,  
'perspect': 0.2368286884902131,  
'posit': 0.17662268935741687,  
'similar': 0.206725688923815,  
'tabl': 0.19703468762300935},  
  
{'(similar \ntable': 0.14640212037019382,  
'10\n10.1.2.5 mee': 0.05463548604450035,  
'10\n10.2.1.2 wor': 0.0867527526042024,  
'10\ninputs tools': 0.06154568906831955,  
'10.1.1 plan com': 0.08260640679088622,  
'10.1.1.3 enterp': 0.07371539232663146,  
'10.1.1.4 organi': 0.11956273972146853,  
'10.1.2 plan com': 0.0918476267110474,  
'10.1.2.2 commun': 0.10070537465711162,  
'10.1.2.3 commun': 0.11582754419999255,  
'10.1.2.4 commun': 0.14374961186756666,  
'10.1.3.2 projec': 0.10126817059018299,
```



14



“

## 5.11. NAMED ENTITY EXTRACTION:

Using the vocabulary of spacy, we extracted from sentences the named entities which are concepts

	named entity	output	named entity	output	
0	CARDINAL	nine	225	GPE	optimal
1	CARDINAL	two hundred	404	CARDINAL	three hundred
2	CARDINAL	eighty-four			
3	ORDINAL	fifth			
4	PERSON	jorge diego			
...	...	...			
515	PERSON	pmi memberid		concept	individual
516	CARDINAL	two million	60	org	diagram control communications
517	CARDINAL	three hundred	12	cardinal	three hundred eight
518	CARDINAL	ninety nine thousand	23	cardinal	twenty-one

”

# 15

## 5.12. RDF TRIPLES:

Out[92]:

	Concepts	Relation	Instance
0	plan	risk	management
1	process	is	ensures
2	management	are	commensurate
3	communicate	obtain	agreement
4	stakeholder	ensure	risk
...	...	...	...
1145	control project project	develop	project work management
1146	variance trend analysis	require	use performance information
1147	project management plan	described	section project management
1148	work performance data	described	section work performance
1149	work performance report	take	information performance measurement

1150 rows × 3 columns

## 5.13. IS A / SUCH AS RELATION EXTRACTION:

	Y	Relation	X	noun_process
0	risk management plan	is a	component	Project Management Plan
1	risk management context	is a	combination	Analytical techniques
2	techniques	such as	use strategic risk scoring	Analytical techniques
3	area	such as	senior management project stakeholders	Expert Judgment
4	terms	such as	levels risk	Meetings
5	risk management plan	is a	component	risk Management Plan
6	probability impact matrix	is a	grid	risk Management Plan
7	area	such as	senior management project stakeholders	Expert Judgment
8	risk management plan	is a	component	risk Management Plan
9	probability impact matrix	is a	grid	risk Management Plan
10	objective	such as	schedule cost quality	risk Probability and Impact Assessment
11	risk quality assessment	is a	technique	risk data Quality Assessment
12	area	such as	senior management project stakeholders	Expert Judgment
13	risk management plan	is a	component	risk Management Plan
14	probability impact matrix	is a	grid	risk Management Plan
15	values	such as	durations schedule activities	data Gathering and representation techniques
16	bounds	such as	early concept stage distribution	data Gathering and representation techniques
17	use type analysis	is a	decision	Quantitative risk Analysis and Modeling techn...

# 16

## 5.14. HAVE / INCLUDE RELATION EXTRACTION:

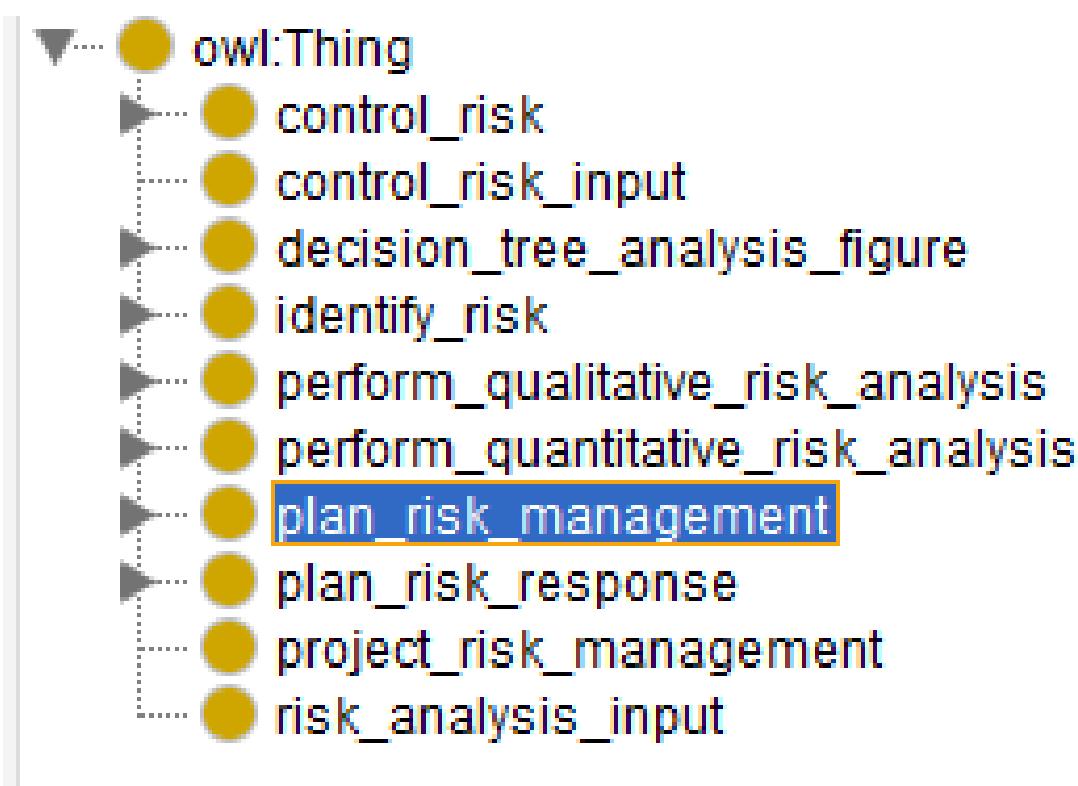
```
[ ] df_Object_Propety
```

	Y	Relation	X	noun_process
0	risks	have	a impact	Probability and Impact Matrix
1	example risks	have	a impact	Probability and Impact Matrix
2	risks	have	a impact objectives	Probability and Impact Matrix
3	example risks	have	a impact objectives	Probability and Impact Matrix
4	variables	have	a degree	Quantitative risk Analysis and Modeling techn...
5	impact variables	have	a degree	Quantitative risk Analysis and Modeling techn...
6	importance impact variables	have	a degree	Quantitative risk Analysis and Modeling techn...
7	variables	have	a degree uncertainty	Quantitative risk Analysis and Modeling techn...
8	impact variables	have	a degree uncertainty	Quantitative risk Analysis and Modeling techn...
9	importance impact variables	have	a degree uncertainty	Quantitative risk Analysis and Modeling techn...
10	actions	include	forming risk	Strategies for Positive risks or opportunities

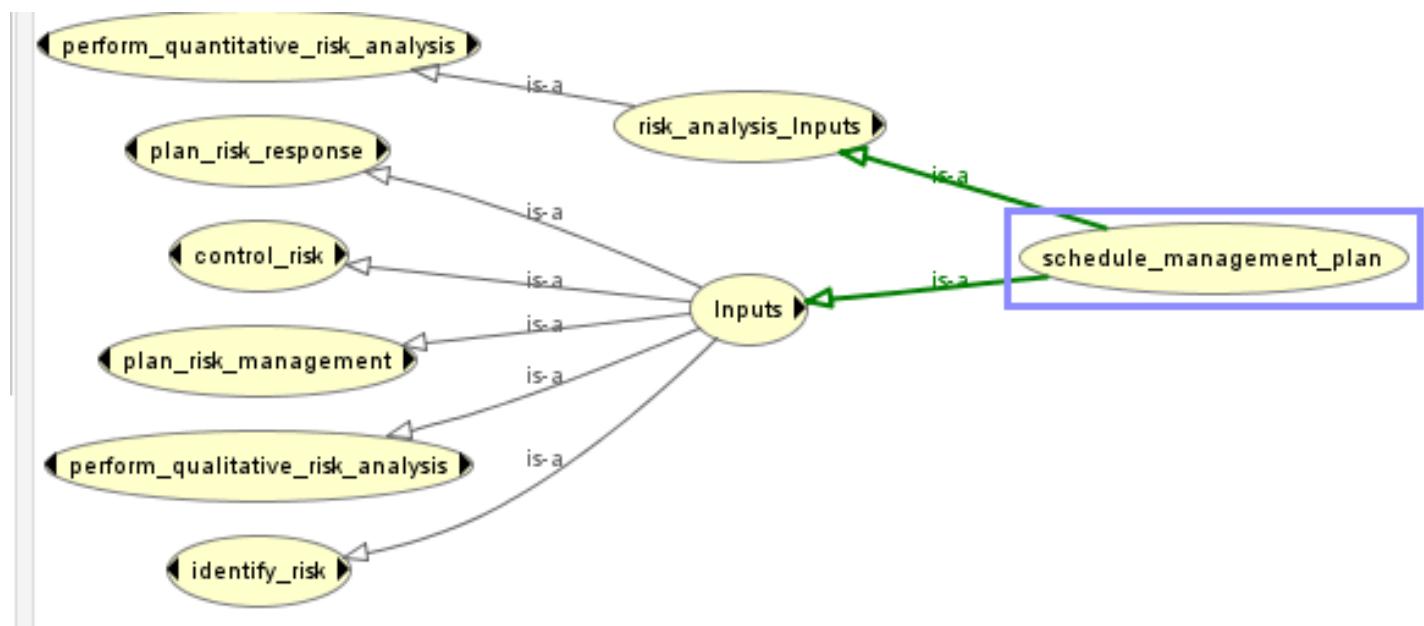
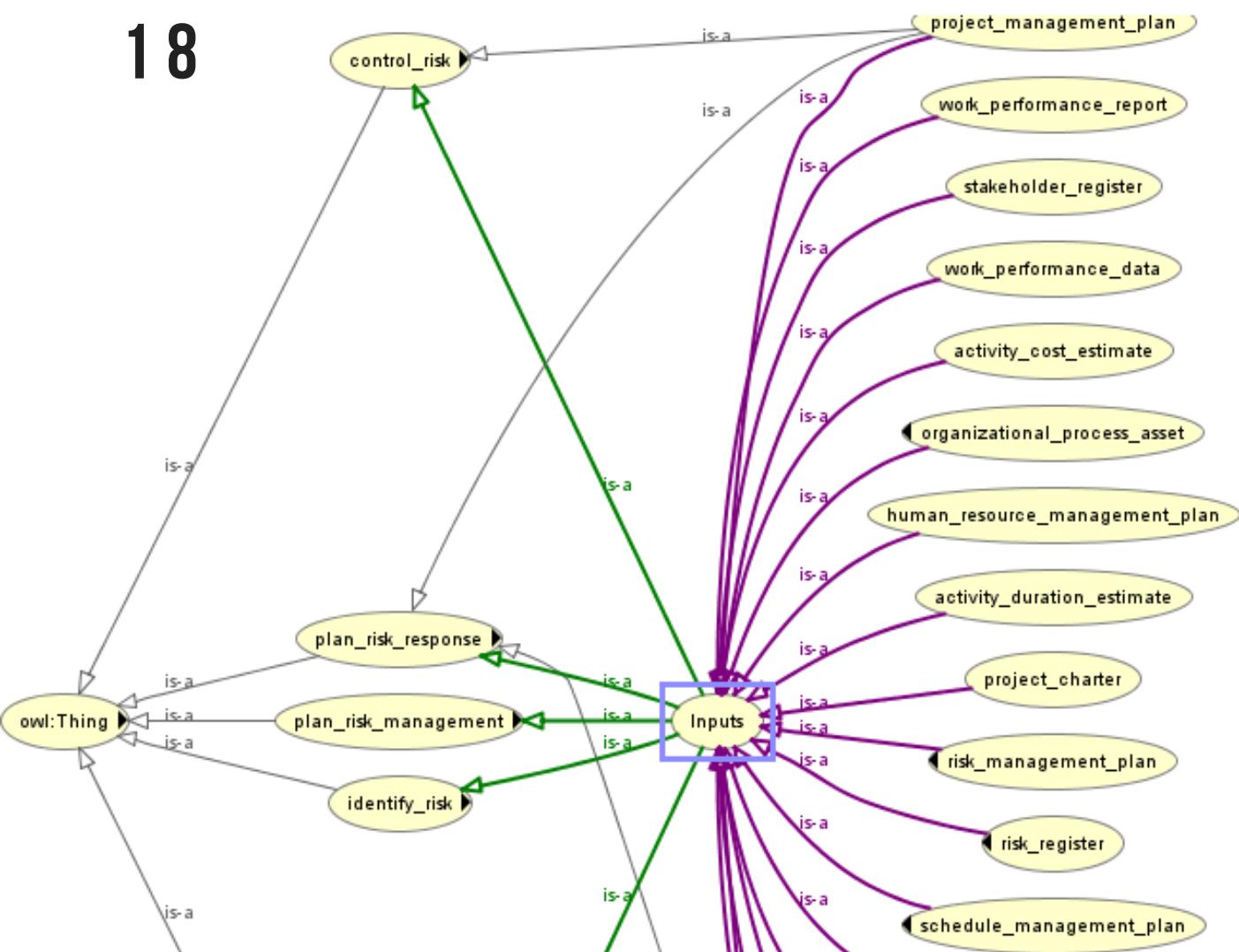
## 6. OWL ONTOLOGY

The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies. Ontologies are a formal way to describe taxonomies and classification networks, essentially defining the structure of knowledge for various domains: the nouns representing classes of objects and the verbs representing relations between the objects.

Ontologies resemble class hierarchies in object-oriented programming but there are several critical differences. Class hierarchies are meant to represent structures used in source code that evolve fairly slowly (perhaps with monthly revisions) whereas ontologies are meant to represent information on the Internet and are expected to be evolving almost constantly.



18



# 19

## 6.1 ONTOLOGY EVALUATION

Measuring the results of our model outputs gets a lot more complex when we're dealing with language.

**Recall:** The recall counts the number of overlapping n-grams found in both the model output and reference – then divides this number by the total number of n-grams in the reference.

**Precision:**

To avoid this we use the **precision** metric – which is calculated in almost the exact same way, but rather than dividing by the **reference** n-gram count, we divide by the **model** n-gram count.

**F1-Score**

Now that we both the recall and precision values, we can use them to calculate our ROUGE F1 score like so:

$$2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$



```
rouge = Rouge()
score=rouge.get_scores(txtEva, final, avg=True)
print('Precision = '+str(score['rouge-1']['p']))
print('Recall = ' +str(score['rouge-1']['r']))
print('f-measure = '+str(score['rouge-1']['f']))
```

```
▷ Precision = 0.838859825620389
    Recall = 0.7992121812030412
    f-measure = 0.8142331288343558
```

20

## 7. ONTOLOGY MATCHING

In the areas of Semantic Web and data integration, ontology matching is one of the important steps to resolve semantic heterogeneity. Manual ontology matching is very labor-intensive, time-consuming and prone to errors. So development of automatic or semi-automatic ontology matching methods and tools is quite important

Ontology alignment, or ontology matching, is the process of determining correspondences between concepts in ontologies. A set of correspondences is also called an alignment. The phrase takes on a slightly different meaning, in computer science, cognitive science or philosophy.

```
▶ response_requete(' Plan risk Management') #1er test validé : requête contient le nom d'un processus
```

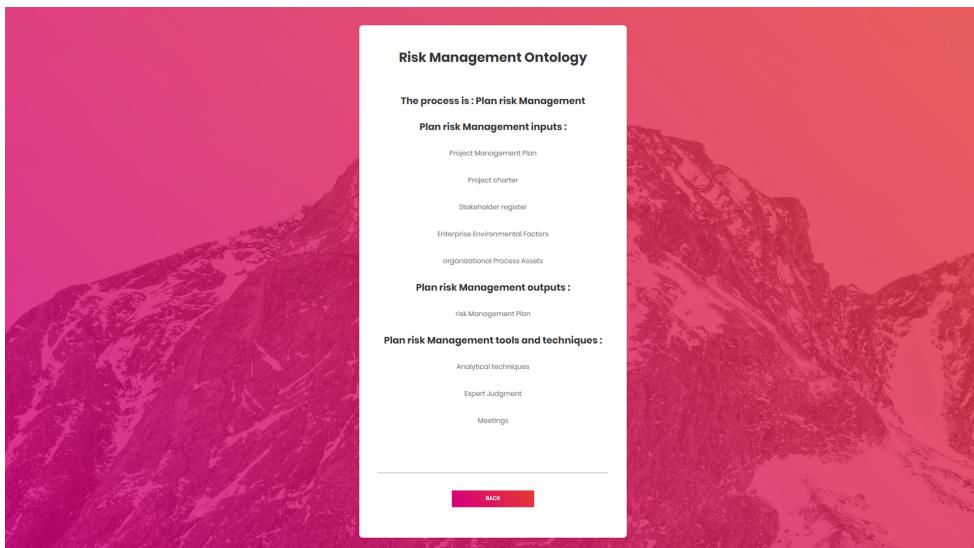
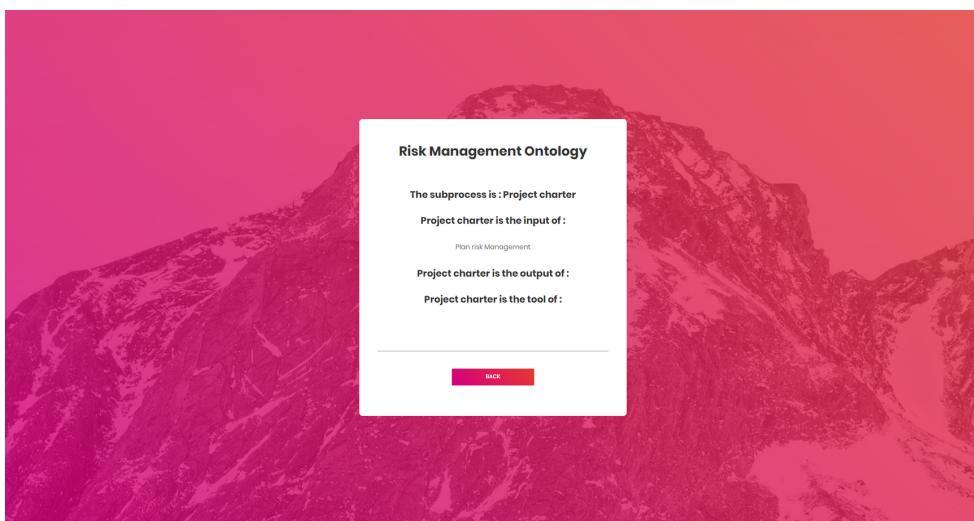
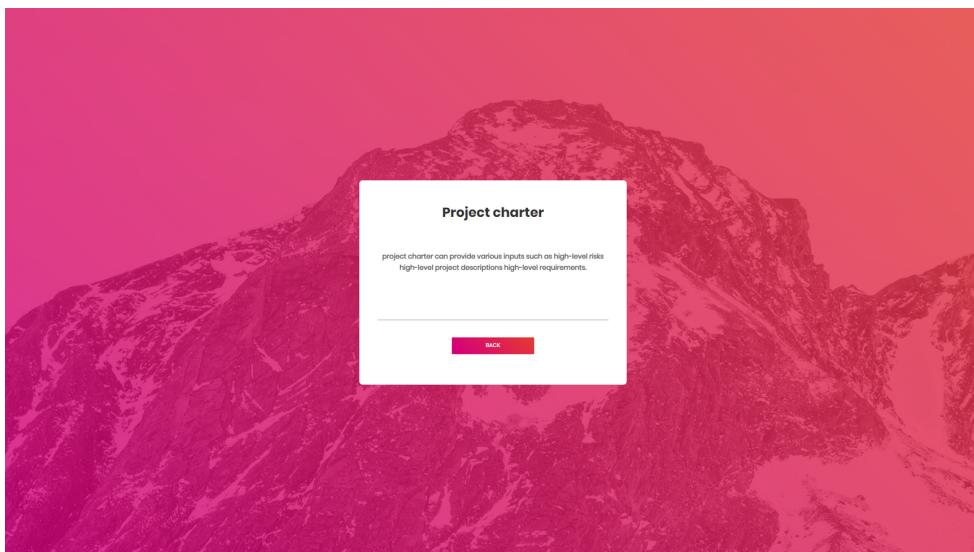
```
('The process is :',
 ' Plan risk Management',
 ' Plan risk Management inputs :',
 [' Project Management Plan',
  ' Project charter',
  ' Stakeholder register',
  ' Enterprise Environmental Factors',
  ' organizational Process Assets'],
 ' Plan risk Management outputs :',
 [' risk Management Plan'],
 ' Plan risk Management tools and techniques :',
 [' Analytical techniques', ' Expert Judgment', ' Meetings'])
```

```
response_requete('Project Charter ') #3ème test validé
```

```
'The subprocess is : Project charter',
 ' Project charter is the input of : ',
 [' Plan risk Management'],
 ' Project charter is the output of : ',
 [],
 ' Project charter is the tool of : ',
 [])
```

21

## 8. DEPLOYMENT



# 22

## 9. CONCLUSION

Managing ontology evolution in a dynamic, multi-actor and distributed environment on the Web is one of the main challenges environment is one of the main challenges that knowledge engineering researchers are currently trying to address, a challenge that has arisen because of the use of ontologies to build the virtual environment of tomorrow, the Semantic Web. In truth, it is unrealistic to want to build the semantic Web, a dynamic and distributed environment whose architecture is based on the semantic referencing of Web objects, without considering the evolution of an ontology that serves as a semantic repository. This conclusion quickly spread to the research community, They are engaged in new research with the object of study being the management of the process of evolution. However, this research, far from being completed, sometimes brings more questions than answers, for example the question of the definition of a complete and coherent set of methodological elements to ensure the evolution of the ontology in a consistent way, or the question of how to carry out a relevant analysis of the effects of changes on the semantic referencing of semantic referencing of Web objects.