

## I Construction Documents

- Include project information # plan, requirement, role & responsibility, and result
- Facilitate project management # alignment, communication, information transfer, and record

## I Limitations of Manual Analysis on Construction Documents

- Cost high # time, money, effort
- Vulnerable to human error # misunderstanding, subjective opinion, omitting



*Necessity of Natural Language Processing*

# Introduction



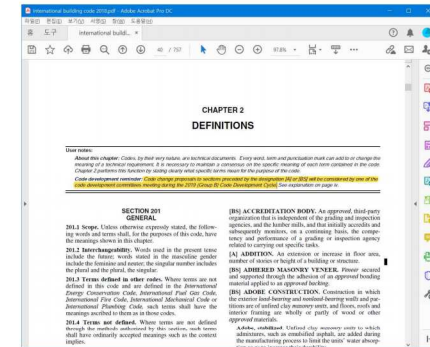
## Machine Translation



## Information Retrieval



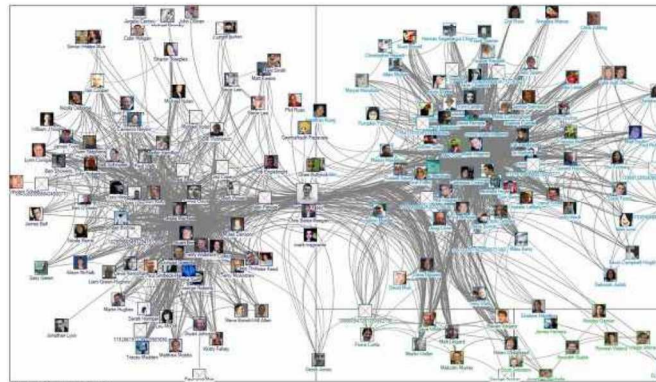
## Optical Character Recognition (OCR)



## Question Answering



## Network Analysis



## Natural Language Generation



한화, KIA에 10회말 끝내기 승리  
(2016-10-08, KIA 5對6 한화, 대전)

8일 대전구장에서 열린 2016 타이거뱅크 KBO리그 한화와 KIA의 경기에서 한화가 정근우의 짜릿한 끝내기 2루타에 힘입어 승리했다. 정근우는 5:5로 팽팽한 경기 중이던 10회말 2사 2루에서 극적인 2루타를 터트리며 대전구장을 열광에 빠뜨렸다. 정근우는 현재 시즌 575타수 178안타 18홈런 60볼넷 88타점 121득점을 기록하고 있다.... 더 보기

좋아요 댓글 달기 공유하기

## Sentiment Analysis



## Speech Recognition

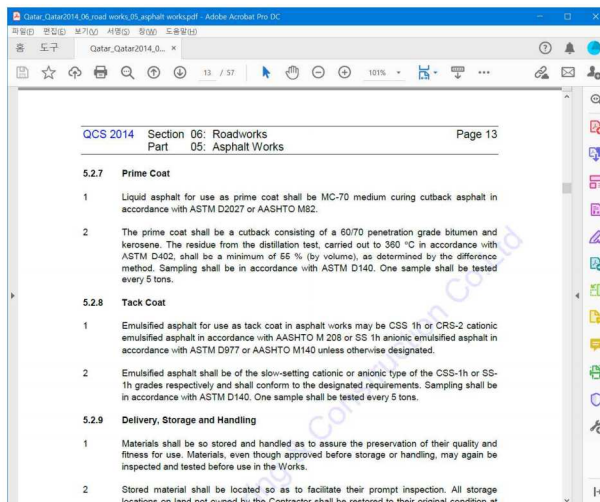


안녕하세요? 무엇을 도와드릴까요?

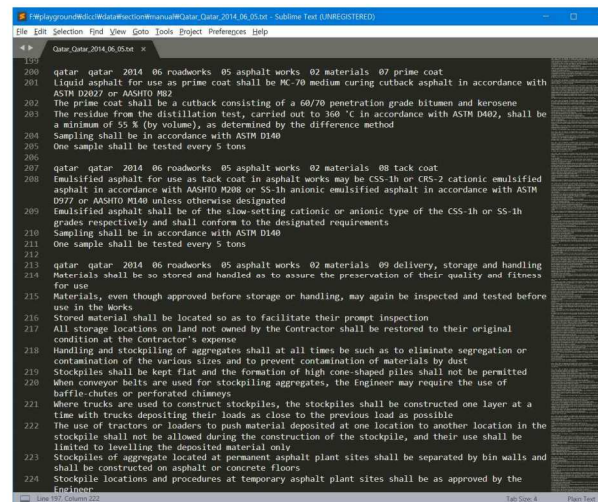


## I Natural Language Processing

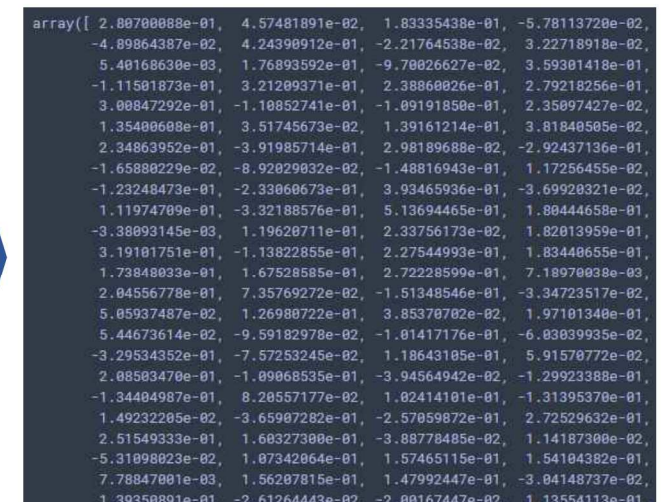
- To convert *text data (natural language)* into *computer-understandable (numeric vectors)* forms
  - Word, sentence, paragraph, documents, ...
- To structuralize the unstructured data (i.e., extract features and assign attributes)



Document  
(Natural Language)



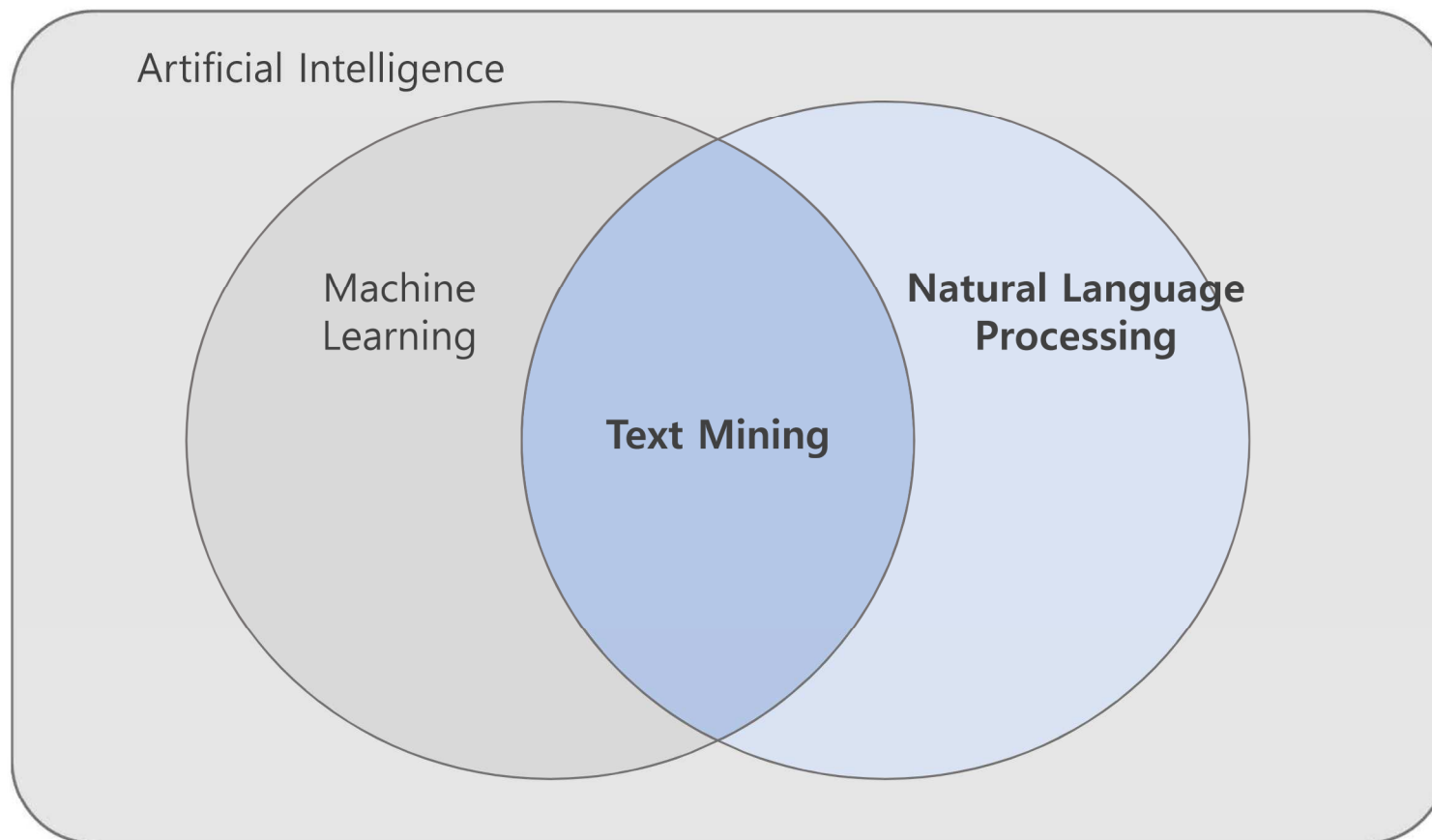
Text Data



Numeric Vector

## I Text Mining

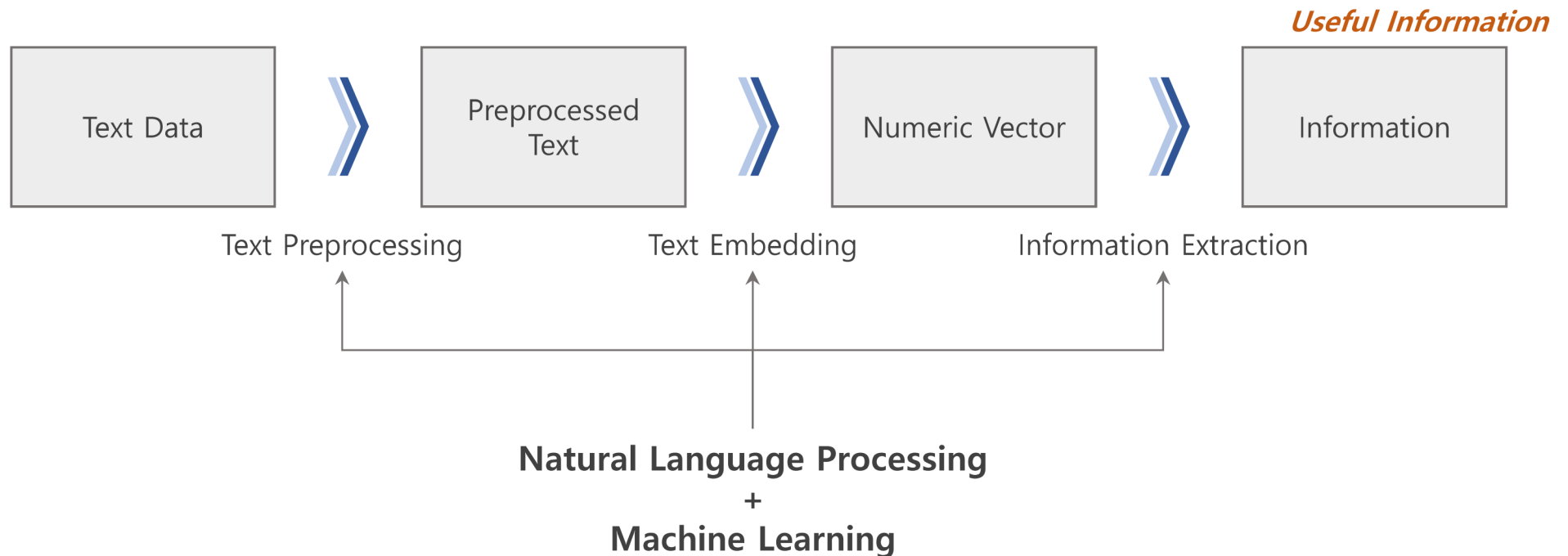
- A subfield of *artificial intelligence* that utilize *machine learning* approaches in *natural language processing* to retrieve *user-needed information* from *text data*





## I Text Mining

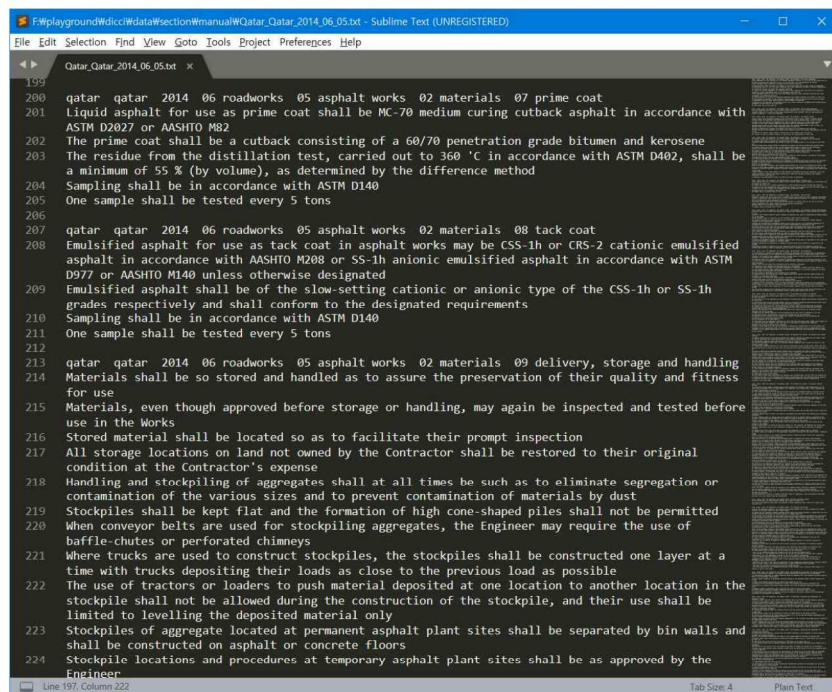
- A subfield of *artificial intelligence* that utilize *machine learning* approaches in *natural language processing* to retrieve *user-needed information* from *text data*



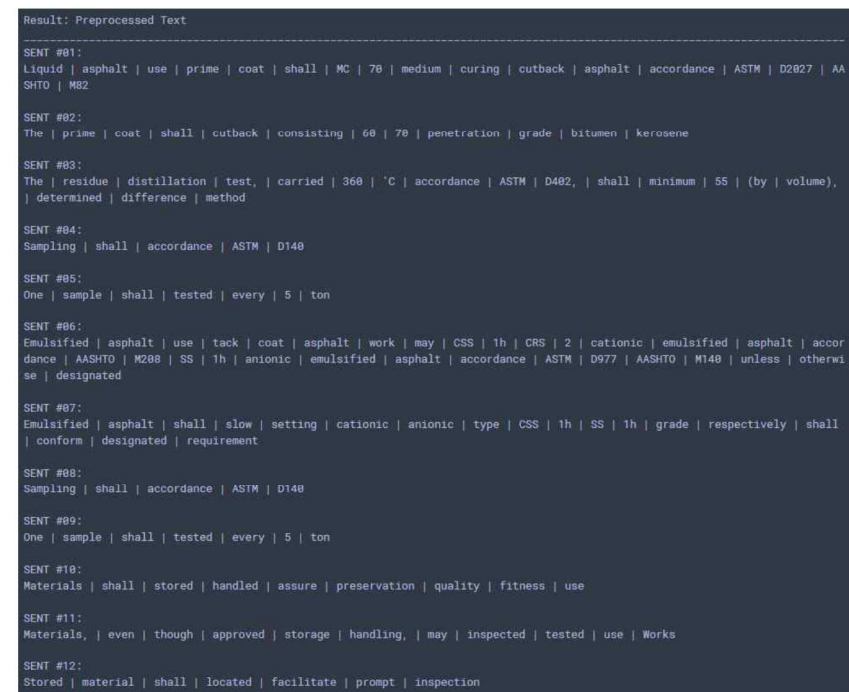
# 2 Text Preprocessing

## Text Preprocessing

- A process of *cleaning the text* to enable the data for *better representation of information*



Text Data



Preprocessed Text

## I Text Preprocessing Methods

Preprocessing Step	Description
Normalization	Remove unnecessary characters, symbols, or spaces based on predefined rules
Tokenization	Parse text into a minimum informative unit
Stopword Removal	Remove words that frequently occurred but less meaningful
Stemming / Lemmatization	Convert each word to its basic form (i.e., stem or lemma)
Part-of-Speech (PoS) Tagging	Markup every word with its PoS tag
Thesaurus	A dictionary of relationships between words

→ *Utilize a combination of text preprocessing approaches depending on data, information, and research objective*



## I Normalization

- Remove unnecessary characters, symbols, or spaces based on predefined rules

### Normalization Rules

- 문장의 모든 글자를 소문자로 변환
- 문장부호(, . ! ? 등) 제거
- 숫자와 함께 등장하는 "\$"는 "dollars"로 변환
- 조동사(do, can, should 등)와 붙어있는 "n't"는 "not"으로 변환
- 인칭대명사(I, he, she, they 등)에 붙어있는 "'m", "'s", "'re"은 "be"로 변환
- 문장에 동사가 있고, 명사에 "'s"가 붙어있다면, 소유격이므로 제거
- "usa"는 "united states"로 변환

...

### Text Data

Peter is from the USA.

Peter's job is a construction manager in Korea, but he can't speak Korean.

He has paid \$200 every month to learn Korean.



### Normalized Text

peter is from the united states

peter job is a construction manager in korea but he can not speak korean

He has paid 200 dollars every month to learn korean

## I Tokenization

- Parse text into a *minimum informative unit* (=token)
- In English, a *space* is commonly used as a separator

Normalized Text

he has paid 200 dollars every month to learn korean

Tokenized Text

he has paid 200 dollars every month to learn korean

- In Korean, tokenization should be based on *morpheme* (형태소)

Normalized Text

그는 한국어를 배우기 위해 매달 200달러를 지불했습니다

Tokenized Text

그 는 한국 어 를 배우 기 위해 매 달 ... 니다

지불 해 드립니다

## Tokenization (N-gram)

- The token is not just a single word, but the minimum informative unit of text
- *Several neighboring words* might have a special meaning → tokenize them into one token
  - Bi-gram: Consider maximum 2 words
  - Tri-gram: Consider maximum 3 words

Tokenized Text

he has paid 200 dollars every month to learn korean

Bi-gram Text

he has paid 200 dollars every month to learn korean

Tokenized Text

그 는 한국 어 를 배우 기 위해 매 달 ... 니다

Bi-gram Text

그 는 한국어 를 배우 기 위해 매 달 ... 니다

*3-gram: 지불했 습니다*

## I Stopword Removal

- Stopword: *frequently occurred* but *less informative* word
- *Remove stopwords* for efficiency and accuracy of analysis
  - Efficiency: reduce time and computing cost
  - Accuracy: enhance robustness against to bias of non-informative but frequent tokens
- The stopwords list should be developed by researchers *according to the analysis objective*
  - NLTK<sup>1)</sup> provides basic stopwords lists for several languages (English, Spanish, French, German, Italian, ...)
  - No standard stopwords list for Korean, but the list of Ranks IN might be useful

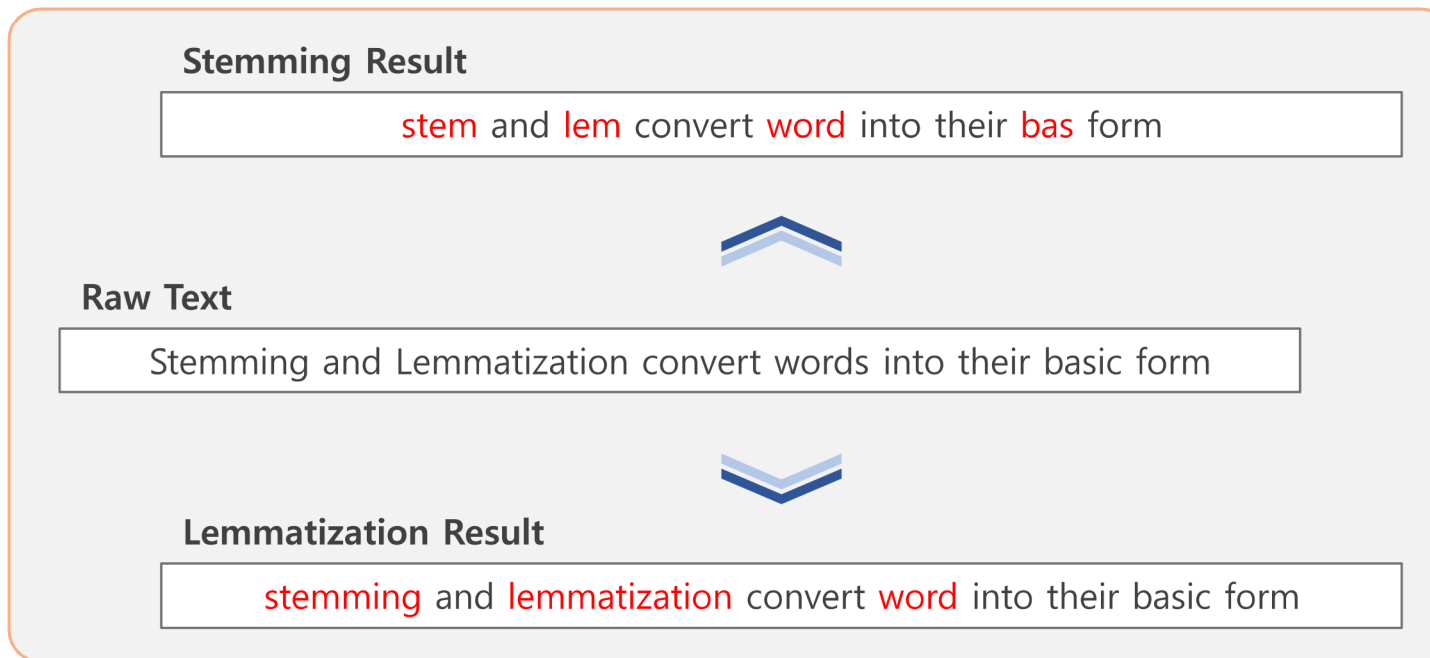
Language	Example of Stopword	Provider	URL
English	am, be, and, a, the, ...	NLTK-3.4.5	<a href="https://www.nltk.org/modules/nltk/corpus.html">https://www.nltk.org/modules/nltk/corpus.html</a>
Korean	그, 이, 저, 무엇, 어느, ...	Ranks IN	<a href="https://www.ranks.nl/stopwords/korean">https://www.ranks.nl/stopwords/korean</a>

1) NLTK: a python module for NLP (<https://www.nltk.org/>)



## I Stemming / Lemmatization

- Convert each word to its *basic form* (i.e., stem or lemma) to reduce the complexity of text data
  - Stem(어간): a base or root form of a word
  - Lemma(표제어/사전어): a canonical(원형) form of a word

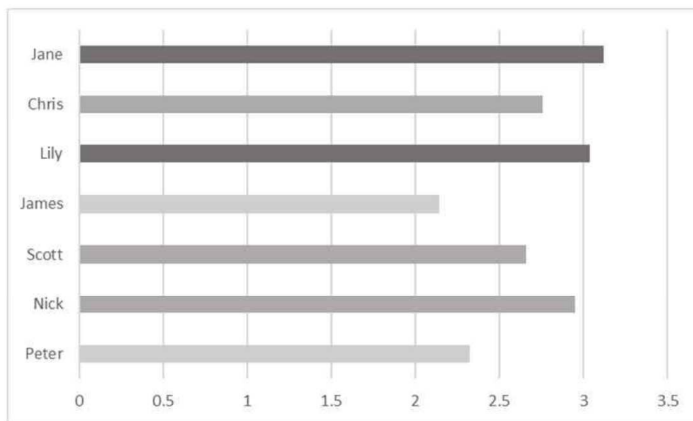


# 3 **Text Embedding**

---

## What is Embedding?

- A process of *mapping data to vector space* so that specific information is well represented
- Similar to concepts such as visualization, projection, mapping  
→ Embedding considers *which information to represent*



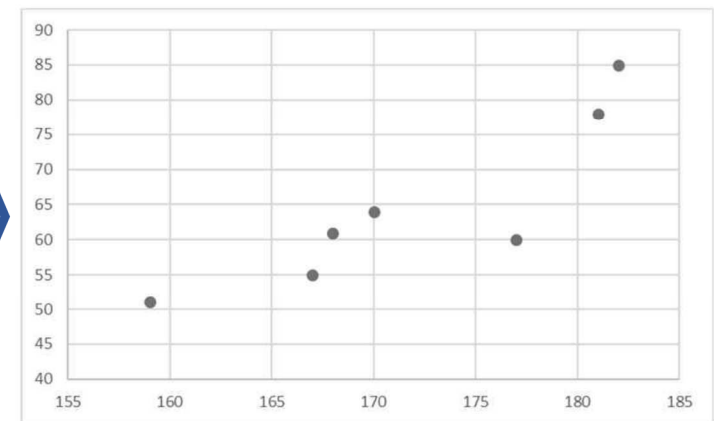
Obesity(비만) ratio



Name	Height	Weight
Peter	181	78
Nick	177	60
Scott	170	64
James	182	85
Lily	167	55
Chris	168	61
Jane	159	51



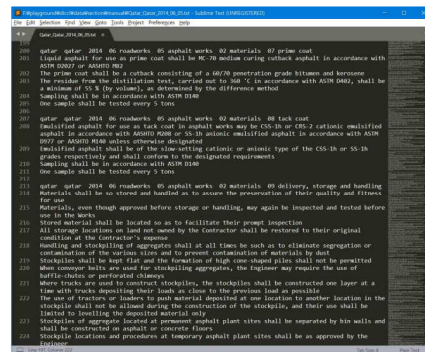
Height and weight



Correlation

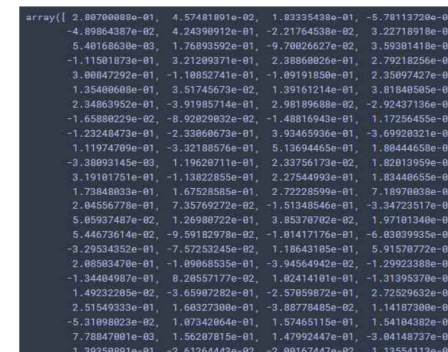
## Text Embedding

- A set of *language modelling* and *feature learning* techniques that maps text to numeric vector space
  - Text: human readable natural language
  - Numeric vector: computer comprehensible form
- To process text data for *better representation of specific information* that meets analysis objective
  - Text data: alphabet, word, sentence, paragraph, documents, etc.
  - Information: occurrence of each word, frequency of each word, length of sentence, etc.



100 qatar qatar 2014 00 roadworks 05 asphalt works 02 materials 02 prime coat  
101 Liquid asphalt for use as prime coat shall be MC-70 medium curing cutback asphalt in accordance with  
102 ASTM D957 or ASTM D959  
103 The prime coat shall be a surface consisting of a 60/70 penetration grade bitumen and kerosene  
104 The residue from the distillation test, carried out to 300 °C in accordance with ASTM D957, shall be  
105 a minimum of 75 % by mass, as determined by the difference method  
106 Sampling shall be in accordance with ASTM D190  
107 One sample shall be tested every 5 tons  
108  
109 qatar qatar 2014 00 roadworks 05 asphalt works 02 materials 02 tack coat  
110 Emulsified asphalt for use as tack coat in asphalt works may be CSS-1h or CMS-2 cationic emulsified  
111 asphalt in accordance with ASTM D960 or 55-1h anionic emulsified asphalt in accordance with ASTM  
112 D977 or ASTM D960 unless otherwise designated  
113 Emulsified asphalt shall be of the slow setting cationic or anionic type of the CSS-1h or 55-1h  
114 grades respectively and shall conform to the designated requirements  
115 Sampling shall be in accordance with ASTM D190  
116 One sample shall be tested every 5 tons  
117  
118 qatar qatar 2014 00 roadworks 05 asphalt works 02 materials 02 delivery, storage and handling  
119 Materials shall be so stored and handled as to secure the preservation of their quality and fitness  
120 for use  
121 Materials, even though approved before storage or handling, may again be inspected and tested before  
122 use in the works  
123 Stored material shall be located so as to facilitate their prompt inspection  
124 All storage locations on and not owned by the Contractor shall be restored to their original  
125 condition at the Contractor's expense  
126 Loading and unloading of aggregate shall at all times be such as to eliminate segregation or  
127 contamination of the various sizes and to prevent contamination of materials by dust  
128 Stockpiles shall be kept clear and the formation of high conical piles shall not be permitted  
129 When conveyor belts are used for stockpiling aggregate, the Engineer may require the use of  
130 baffles, chutes or perforated chutes  
131 Where trucks are used to construct stockpiles, the stockpiles shall be constructed one layer at a  
132 time with trucks depositing their loads as close to the previous load as possible  
133 The use of tractors or loaders to push material deposited at one location to another location in the  
134 stockpile shall not be allowed during the construction of the stockpile, and their use shall be  
135 limited to levelling the deposited material only  
136 Stockpiles of aggregate located at permanent asphalt plant sites shall be operated by his walls and  
137 shall be constructed on asphalt or concrete floors  
138 Stockpile locations and procedures at temporary asphalt plant sites shall be as approved by the  
139 Engineer

Text Data



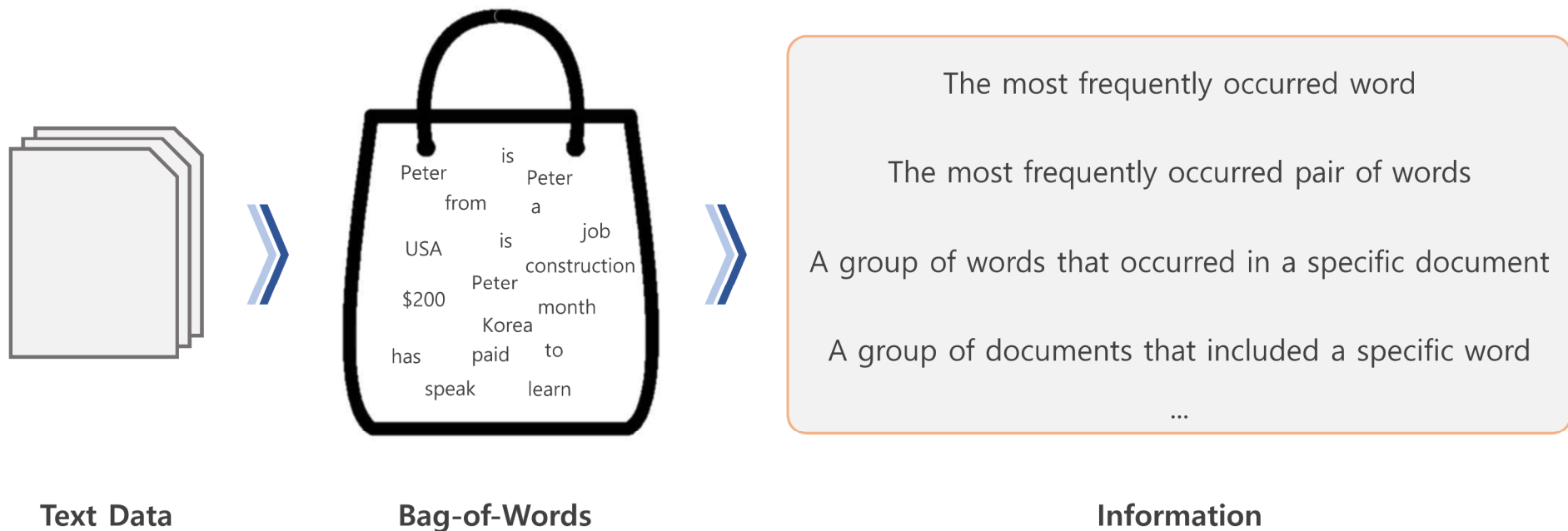
```
array([ 2.86780088e-01,  4.57481801e-02,  1.83235438e-01, -5.78113720e-02,  
       -4.89864387e-02,  4.24390912e-01, -2.21764538e-02,  3.22718918e-02,  
       5.48168630e-03,  1.76893592e-01, -9.78826627e-02,  3.59381418e-01,  
       -1.11581873e-01,  3.21289371e-01,  2.38868826e-01,  2.79218256e-01,  
       3.08847292e-01, -1.10832741e-01, -1.09191850e-01,  2.35897427e-02,  
       1.35406680e-01,  3.57745673e-02,  1.39161214e-01,  3.81848385e-02,  
       2.34883952e-01, -3.91085714e-01,  2.98189688e-02, -2.92437135e-01,  
       -1.65888229e-02, -8.92829832e-02, -1.48816943e-01,  1.17256455e-02,  
       -1.23248473e-01, -2.33868673e-01,  3.93465936e-01, -3.69928321e-02,  
       1.11974709e-01, -3.32188576e-01,  5.13694465e-01,  1.88444658e-01,  
       -3.38893145e-03,  1.19629711e-01,  2.3756173e-02,  1.82013959e-01,  
       3.19181751e-01, -1.13822855e-01,  2.27544939e-01,  1.83448655e-01,  
       1.73848833e-01,  1.67528585e-01,  2.7228599e-01,  7.18978838e-03,  
       2.84556778e-01,  7.35769272e-02, -1.51348546e-01, -3.34723517e-02,  
       5.85937487e-02,  1.26988722e-01,  3.85378782e-02,  1.97181348e-01,  
       5.44673614e-02, -9.59182978e-02, -1.81417176e-01, -6.83839935e-02,  
       -3.29343452e-01, -7.57233245e-02,  1.18643189e-01,  5.91570772e-02,  
       2.88583478e-01, -1.89868835e-01, -3.94656942e-02, -1.29923380e-01,  
       -1.24484987e-01,  8.28857177e-02,  1.82414181e-01, -1.31398378e-01,  
       1.49232285e-02, -3.65887282e-01, -2.57859872e-01,  2.72528632e-01,  
       2.51549333e-01,  1.69327380e-01, -3.88778485e-02,  1.14187380e-02,  
       -5.31898823e-02,  1.87342864e-01,  1.57465115e-01,  1.54184382e-01,  
       7.78847801e-03,  1.56287815e-01,  1.47992447e-01, -3.84148737e-02,  
       1.39358891e-01, -2.61264443e-02, -2.08167447e-02,  1.13554113e-01])
```

Numeric Vector



## I Bag-of-Words Model

- A language model that is to understand text data based on a bag-of-words
- The model keeps shape and frequency of each word → Simple, fast, easy to interpret
  - Occurrence: One-hot Encoding
  - Frequency: TF, DF, TF-IDF



## I One-hot Encoding

- Organize a *document-term matrix of 0/1* based on occurrence
- The most simple and intuitive embedding technique
  - Word embedding: “which document that the word appeared?”
  - Document embedding: “which terms that the document included?”

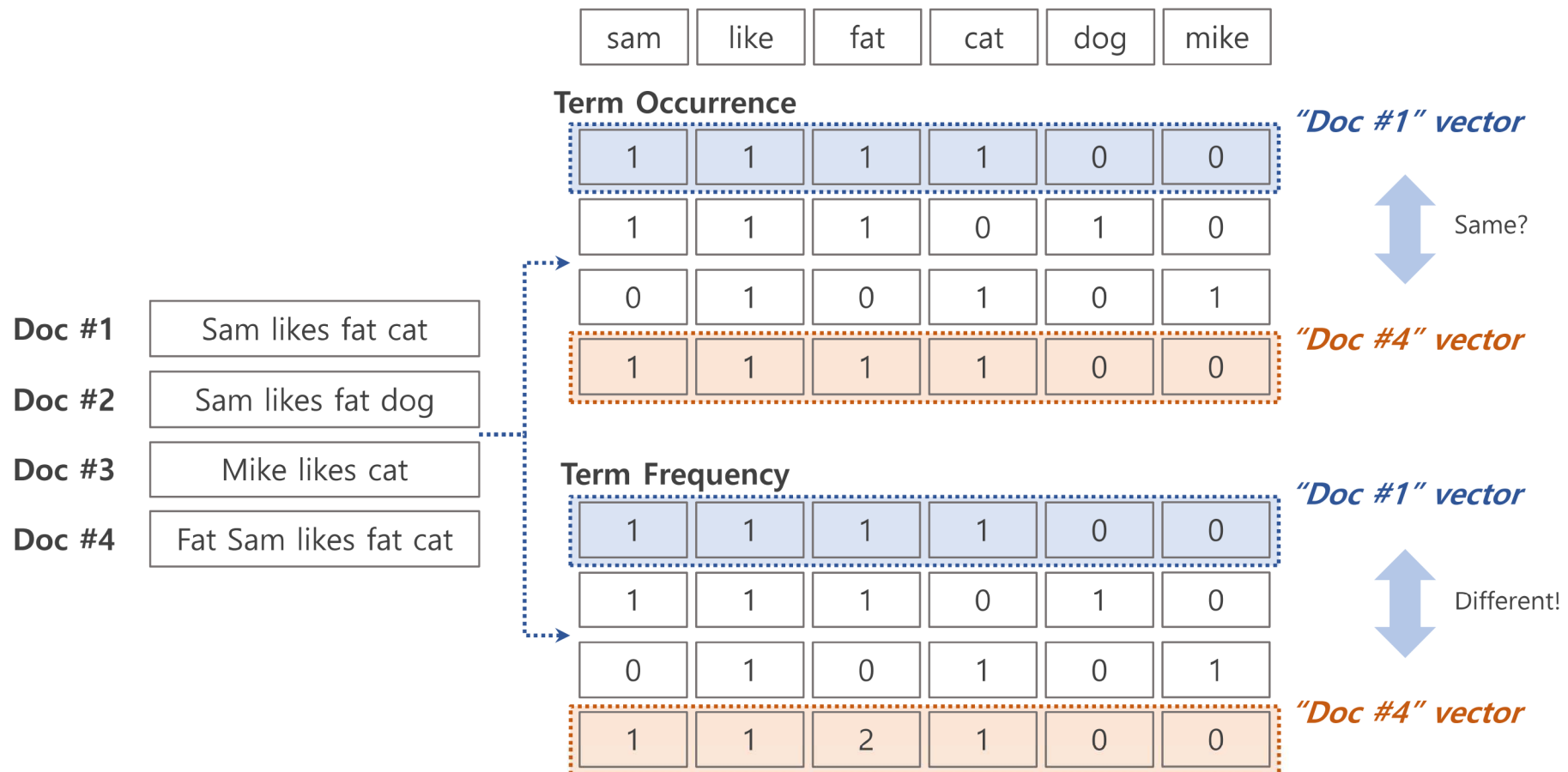
		sam	like	fat	cat	dog	mike
		Term Occurrence					
Doc #1	Sam likes fat cat	1	1	1	1	0	0
Doc #2	Sam likes fat dog	1	1	1	0	1	0
Doc #3	Mike likes cat	0	1	0	1	0	1

*“Doc #2” vector*

*“cat” vector*

## I Term Frequency (TF)

- Organize a *document-term matrix* based on *term frequency*
  - Better representation compared to One-hot encoding




## I Term Frequency (TF)

- Terms that *occurred in many documents* have low informative power  
→ they cannot be used to *distinguish one document from others*
  - E.g.) “a,” “the,” “is,” or “list” in the example
- Need to *normalize* the weight of *less informative terms*

		sam	like	fat	cat	dog	mike
		Term Frequency					
Doc #1	Sam likes fat cat	1	1	1	1	0	0
Doc #2	Sam likes fat dog	1	1	1	0	1	0
Doc #3	Mike likes cat	0	1	0	1	0	1
Doc #4	Fat Sam likes fat cat	1	1	2	1	0	0

No Information



Less Important



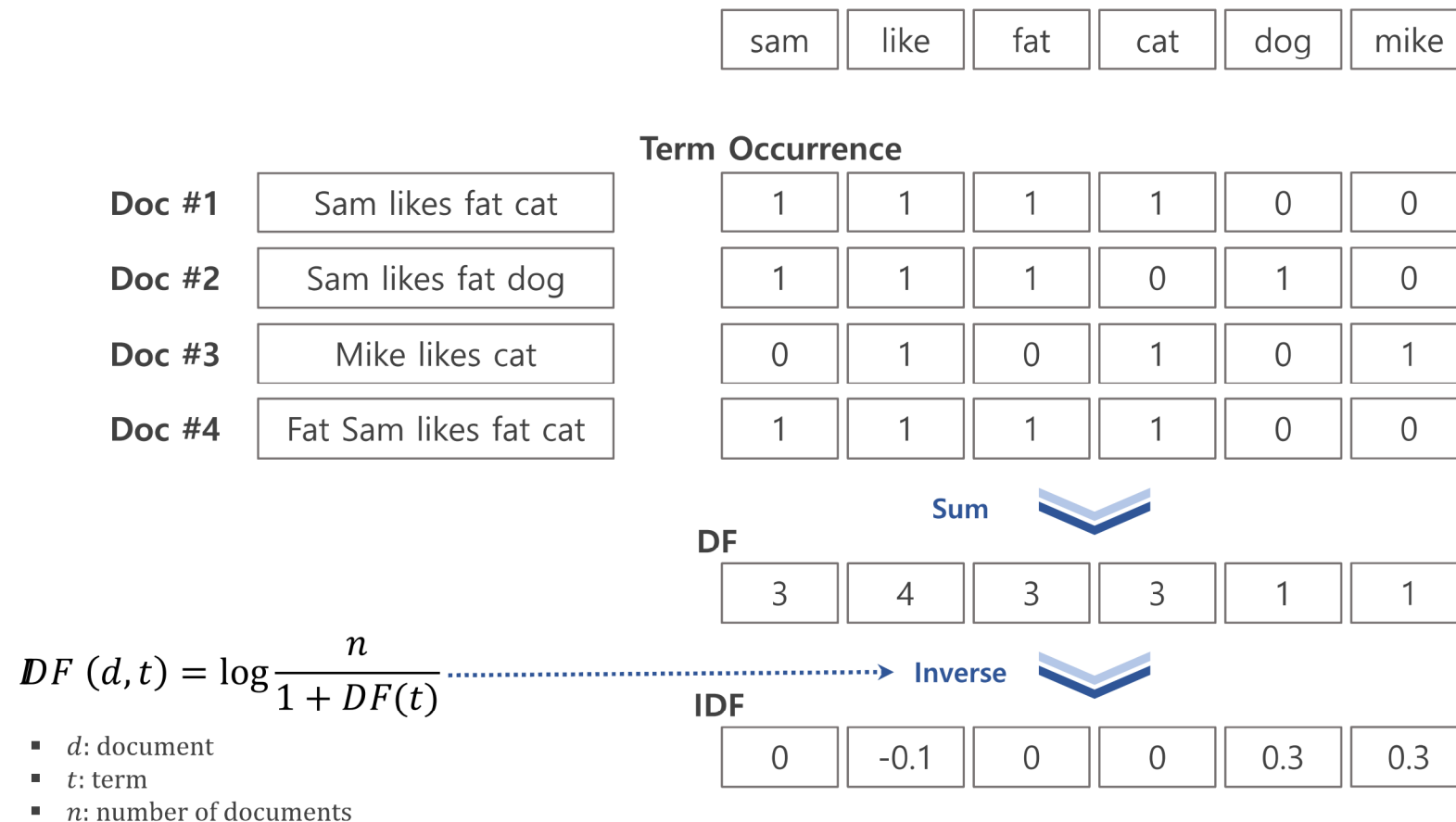
## I Document Frequency (DF)

- The *number of documents* that *include the term*
- DF considers the occurrence of each term rather than frequency, since the target information is “*how the term prevalently used* among documents?”

		sam	like	fat	cat	dog	mike
		Term Occurrence					
Doc #1	Sam likes fat cat	1	1	1	1	0	0
Doc #2	Sam likes fat dog	1	1	1	0	1	0
Doc #3	Mike likes cat	0	1	0	1	0	1
Doc #4	Fat Sam likes fat cat	1	1	1	1	0	0
		Sum					
		DF					
		3	4	3	3	1	1

## I Inverse Document Frequency (IDF)

- *Inverse value of DF*, which indicates the *informative power of the term* within the corpus



# Text Embedding



		Term Frequency					
Doc #1	Sam likes fat cat	1	1	1	1	0	0
Doc #2	Sam likes fat dog	1	1	1	0	1	0
Doc #3	Mike likes cat	0	1	0	1	0	1
Doc #4	Fat Sam likes fat cat	1	1	2	1	0	0

$$TF(t, d) = 0.5 + \frac{0.5 \times f(t, d)}{\max\{f(w, d) : w \in d\}}$$

- $d$ : document
- $t$ : term
- $w$ : word

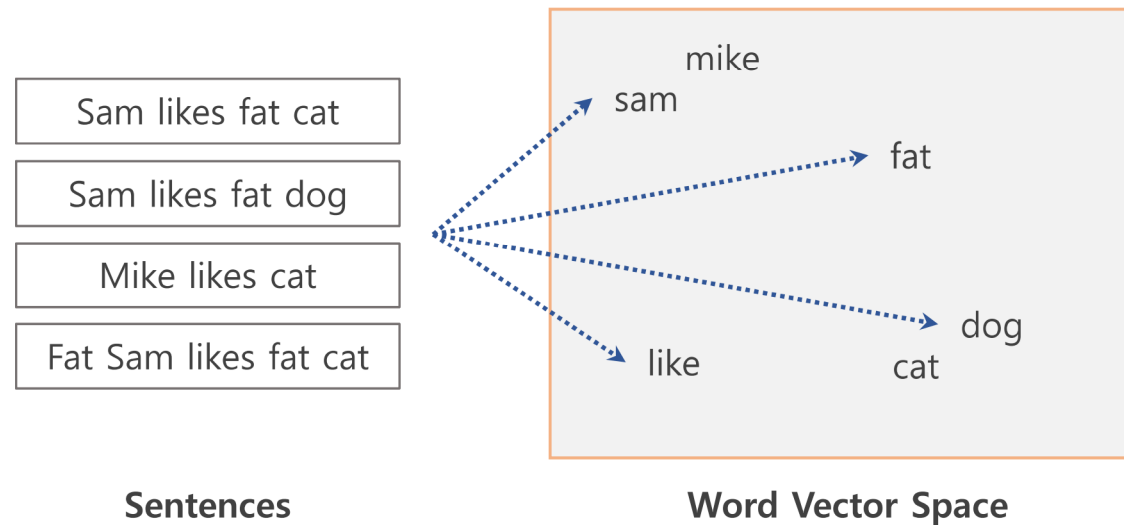
## Term Frequency-Inverse Document Frequency (TF-IDF)

		TF						
			sam	like	fat	cat	dog	mike
Doc #1	Sam likes fat cat	TF	1	1	1	1	0.5	0.5
	Sam likes fat dog		1	1	1	0.5	1	0.5
	Mike likes cat		0.5	1	0.5	1	0.5	1
	Fat Sam likes fat cat		0.75	0.75	1	0.75	0.5	0.5
Doc #2	Sam likes fat cat	IDF	0	-0.1	0	0	0.3	0.3
	Sam likes fat dog							
	Mike likes cat							
	Fat Sam likes fat cat							
Doc #3	Sam likes fat cat	TF-IDF	0	-0.1	0	0	0.15	0.15
	Sam likes fat dog		0	-0.1	0	0	0.3	0.15
	Mike likes cat		0	-0.1	0	0	0.15	0.3
	Fat Sam likes fat cat		0	-0.075	0	0	0.15	0.15
Doc #4	Sam likes fat cat							
	Sam likes fat dog							
	Mike likes cat							
	Fat Sam likes fat cat							

"Doc #1" vector

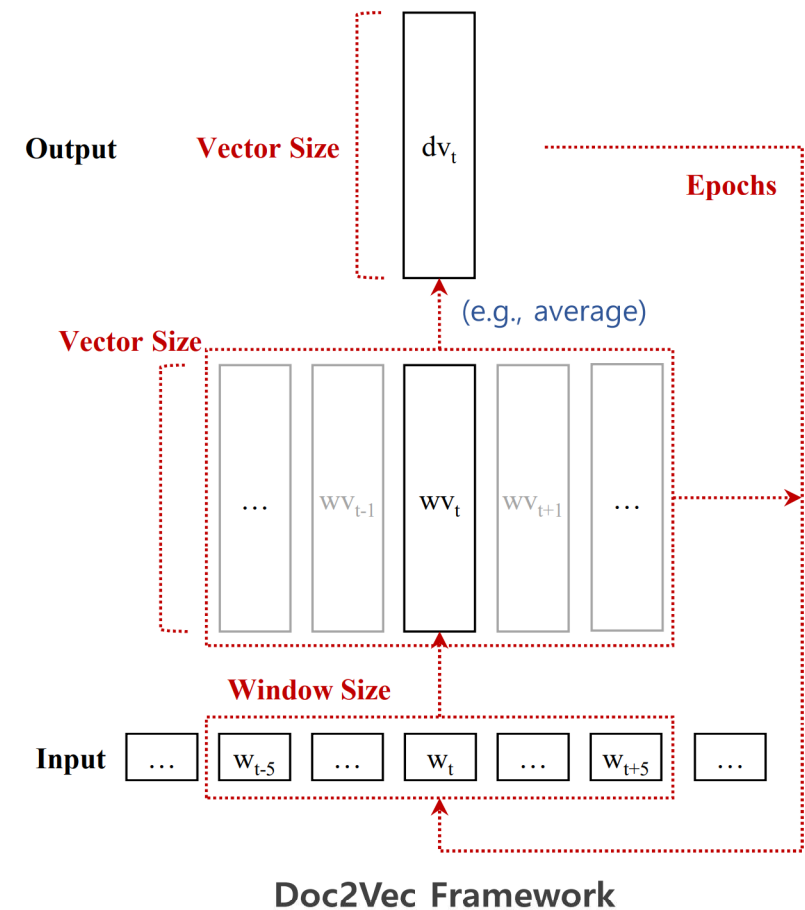
## I Distributed Language Model

- To represent *text data in dense matrix* considering the *distributed representations* of text
  - Word2Vec, Doc2Vec, GloVe, ELMo, BERT, ...
- Words of which *neighbor words are similar* would be *mapped close* in vector space



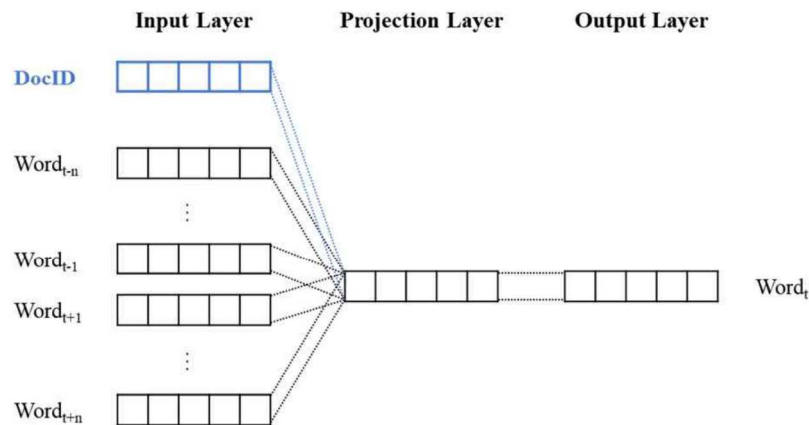
## I Doc2Vec

- To *vectorize the document* (i.e., a list of sequential words) based on the *word vectors from Word2Vec*
- Model training process is following:
  - 1) Consider the document vector as one of the word vectors
  - 2) Initialize the vectors
  - 3) Learn the vectors



## Doc2Vec Architecture

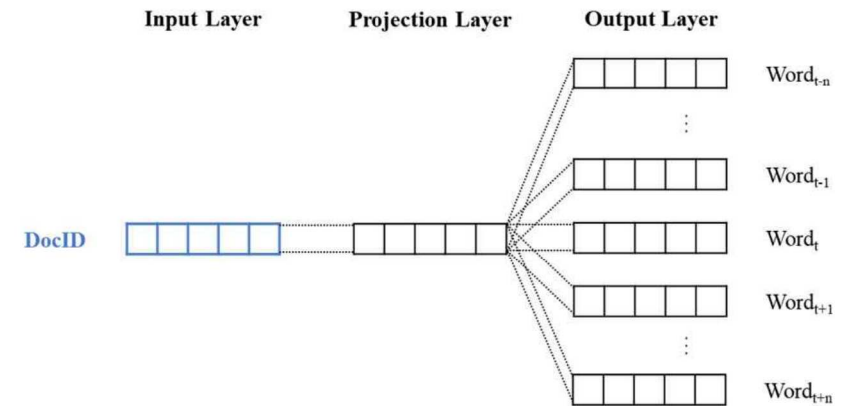
Architecture	Input	Output	Note
PV-DM (Paragraph Vector with Distributed Memory)	neighbor vectors	current vector	Intuitive
PV-DBOW (Paragraph Vector with Distributed Bag-of-Words)	current vector	neighbor vectors	Fast



PV-DM



문서정보뿐만 아니라  
그 안에 포함된 각 단어의  
정보도 함께 활용  
→ 설명력 높다

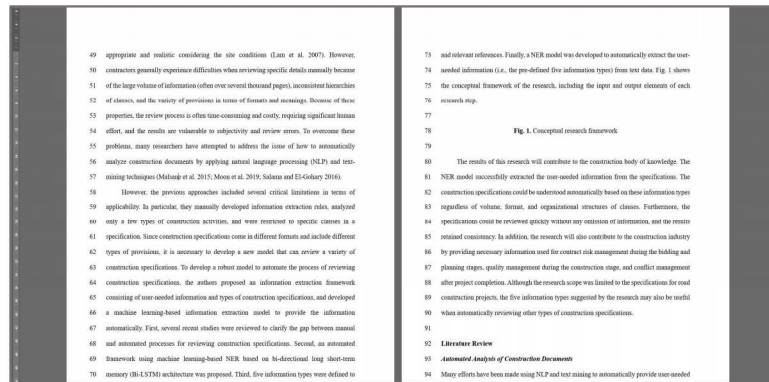


PV-DBOW



## Doc2Vec Results

- Text data: sentences from a journal article (Moon et al. 2020)



Parameter	Value
Vector Size	100
Window	5
min_count	10
Architecture	PV-DM
Epochs	100

Text Data

Parameters

Doc2Vec Results

Input Text

the objective is to map the text to numeric vector space

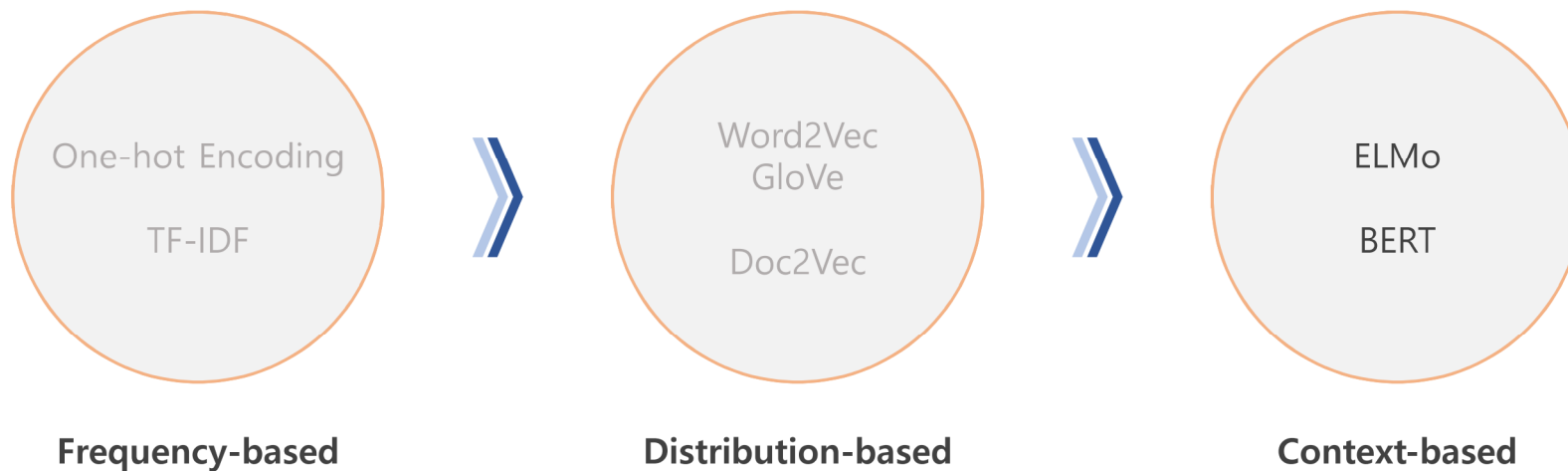
Output Results

['following', 'word', 'tokenized', 'sentence', 'converted', 'embedded', 'word', 'vector', ...]  
 ['word2vec', 'method', 'represents', 'word', 'numerical', 'vector', 'based', ...]  
 ['every', 'word', 'text', 'sentence', 'must', 'converted', 'numerical', 'data', ...]

...

## ■ State-of-the-Art Embedding Techniques

- Natural language uses the *same words* for *different meaning* in *different context*  
→ Text embedding techniques have continuously improved to capture the *context information*

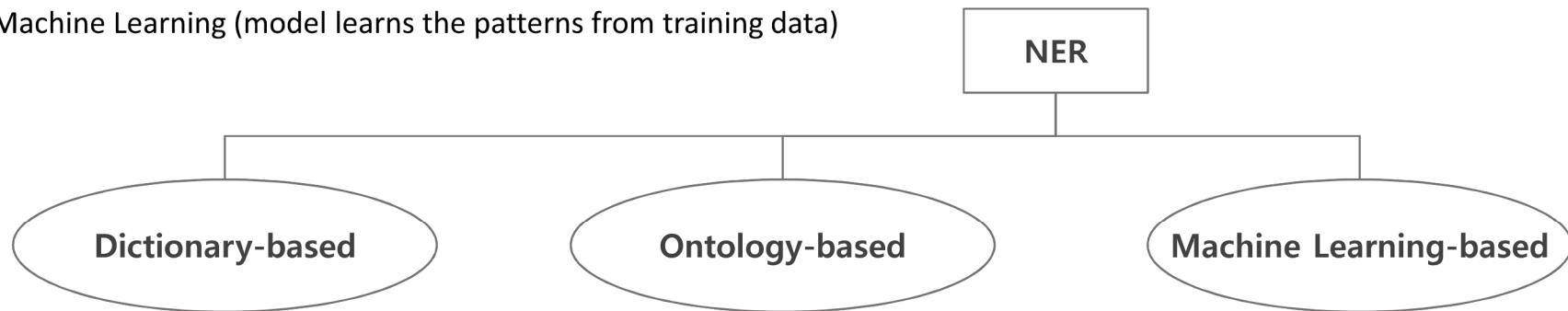




# 4 Information Extraction

## I Named Entity Recognition

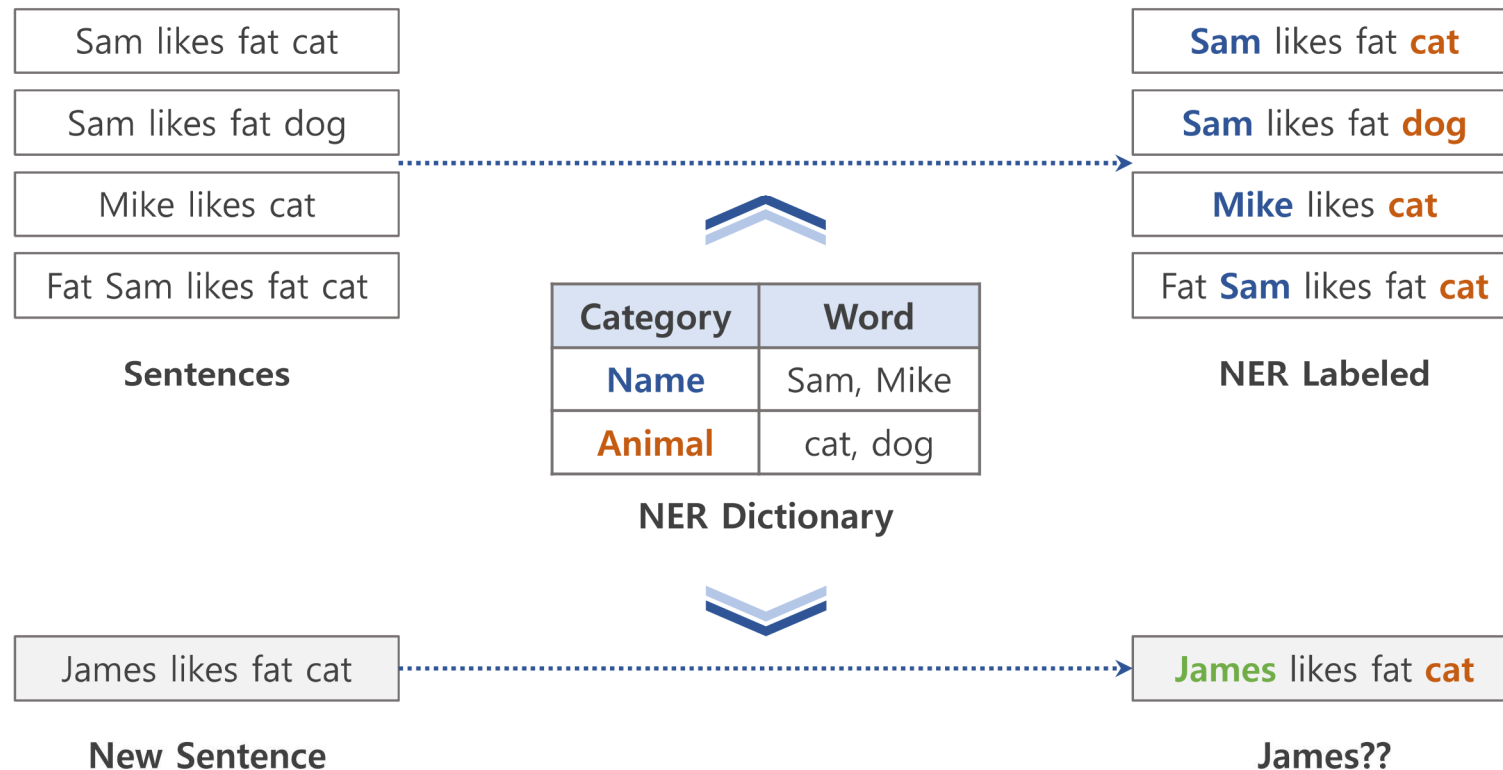
- NER approaches are divided by *how the classification rules are defined*
  - Dictionary (word and class)
  - Ontology (user-defined rules of words)
  - Machine Learning (model learns the patterns from training data)



Approach	User-defined Dictionary	User-defined rules	Computer-learned patterns
Strength	<ul style="list-style-type: none"> <li>Easy to interpret results</li> <li>Easy to apply user intention (precision 100%)</li> </ul>	<ul style="list-style-type: none"> <li>Easy to interpret results (imitate the way that human understands text)</li> <li>Relatively expandable</li> </ul>	<ul style="list-style-type: none"> <li>Machine learning model trains the human-labeled data</li> <li>Expandable (robust to new input)</li> </ul>
Weakness	<ul style="list-style-type: none"> <li>Every word is required to be assigned with proper label</li> <li>Results are restricted in the predetermined dictionary</li> </ul>	<ul style="list-style-type: none"> <li>Every rule is required to be defined</li> <li>Results are restricted in the predetermined rules</li> </ul>	<ul style="list-style-type: none"> <li>Cost to prepare the training data set</li> <li>Difficult to apply user intention</li> </ul>

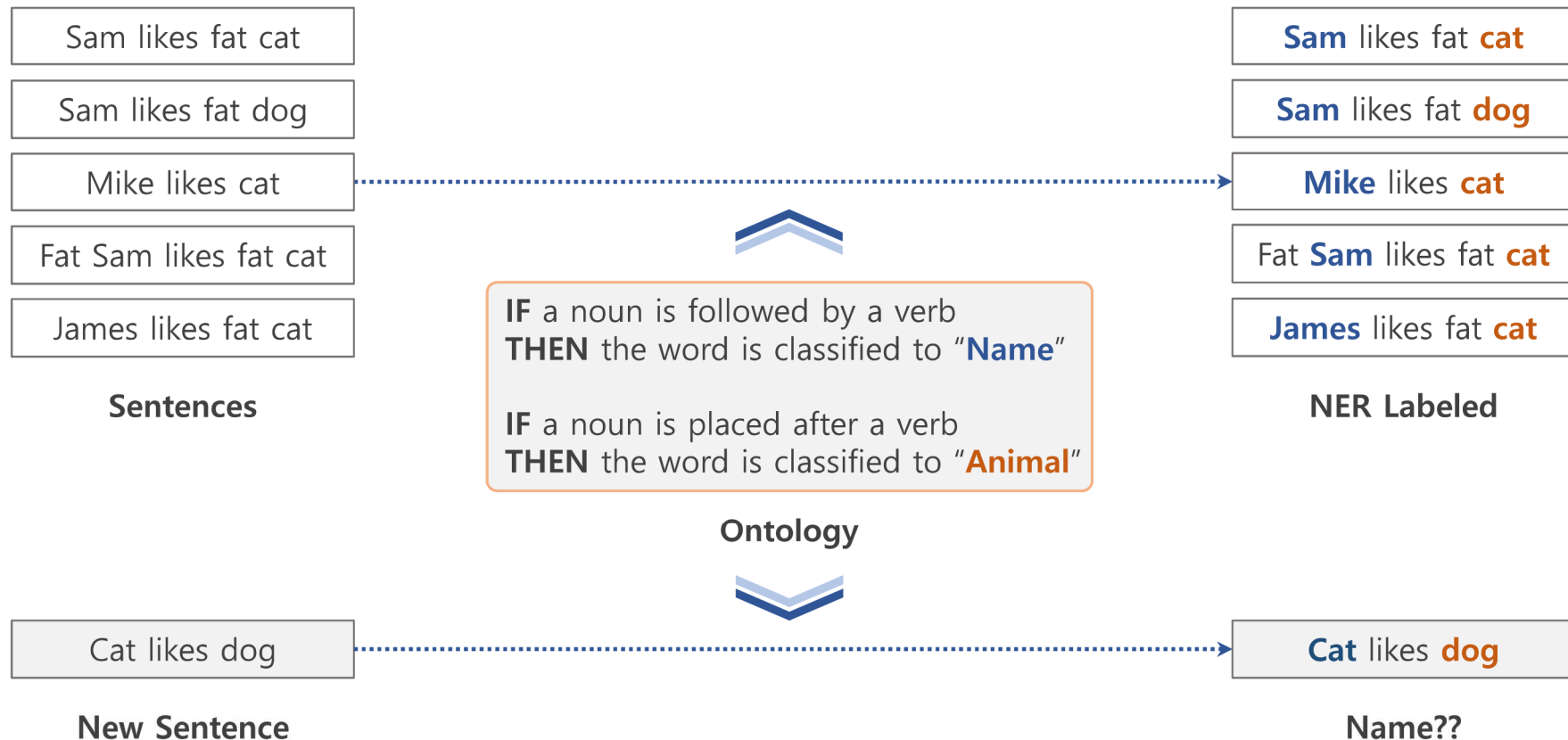
## Dictionary-based NER

- To perform NER based on a *dictionary* that links *every term to user-defined categories*
  - Strength: Every word in the dictionary can be recognized correctly (precision 100%)
  - Weakness: Every word should be listed in the dictionary → difficult to apply to new data



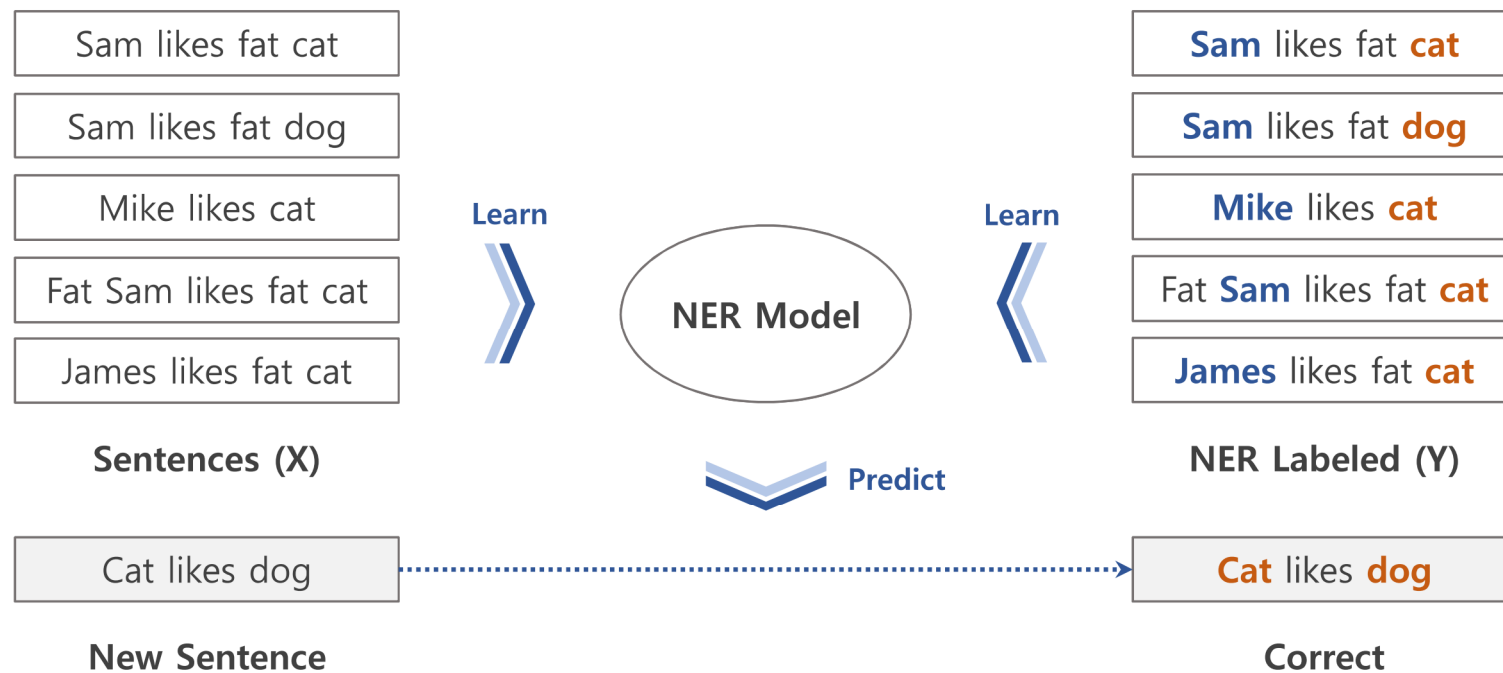
## I Ontology-based NER

- To perform NER based on *user-defined rules* (i.e., ontology)
  - Strength: Easy to interpret since the ontology reflects *the way that human understands text*
  - Weakness: Every pattern should be listed in the ontology → difficult to apply to new data



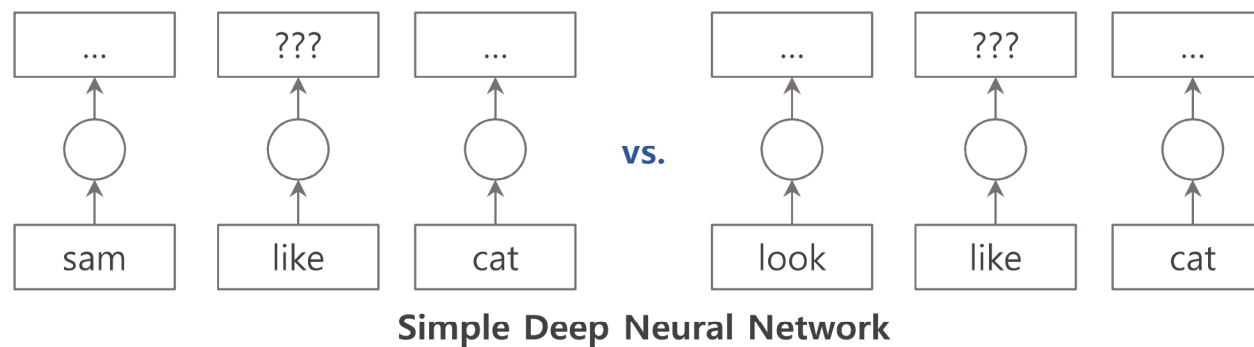
## I Machine Learning-based NER

- To perform NER based on *machine learning classification models*
  - Strength: Robust and expandable to new data → *proper to industrial application*
  - Weakness: Cost to prepare the training data set and difficult to apply user intention directly

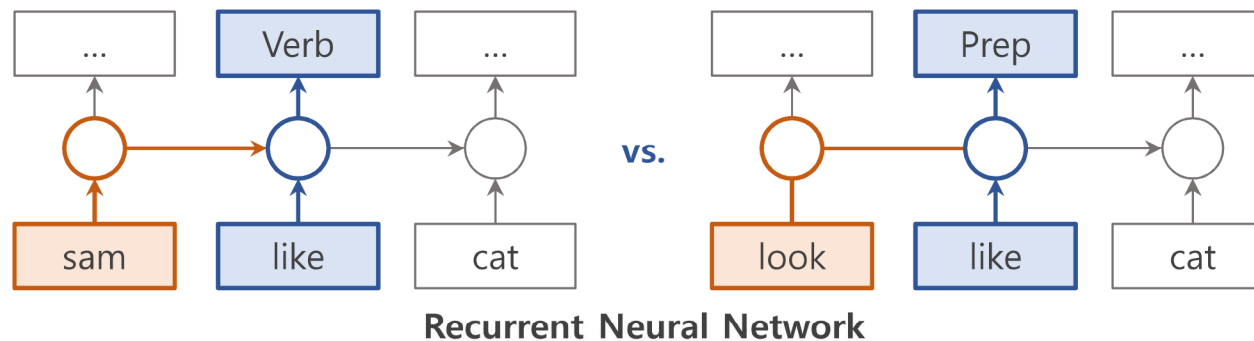


## NER Architecture – RNN

- Machine learning models that are based on simple deep neural network *do not consider context information*
  - Same term would be classified to same category regardless of neighboring words

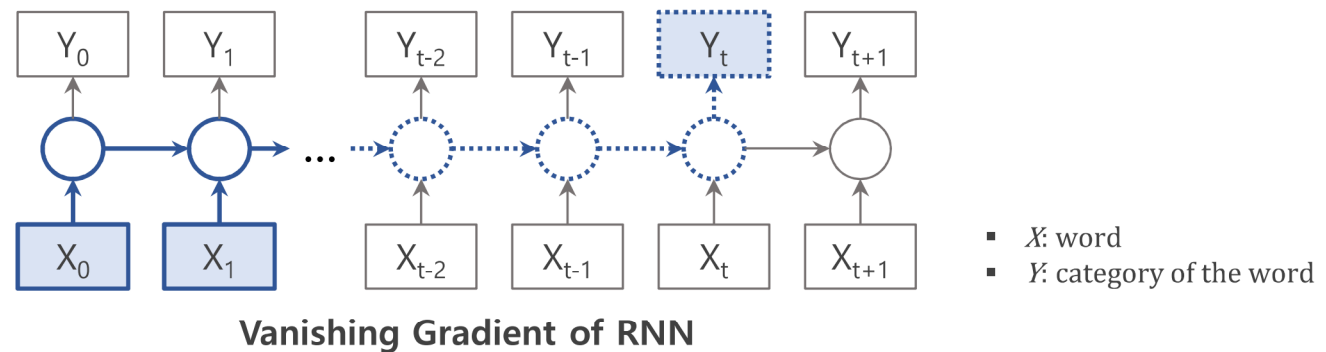


- Recurrent Neural Network (RNN) can classify the words *based on the previous states*

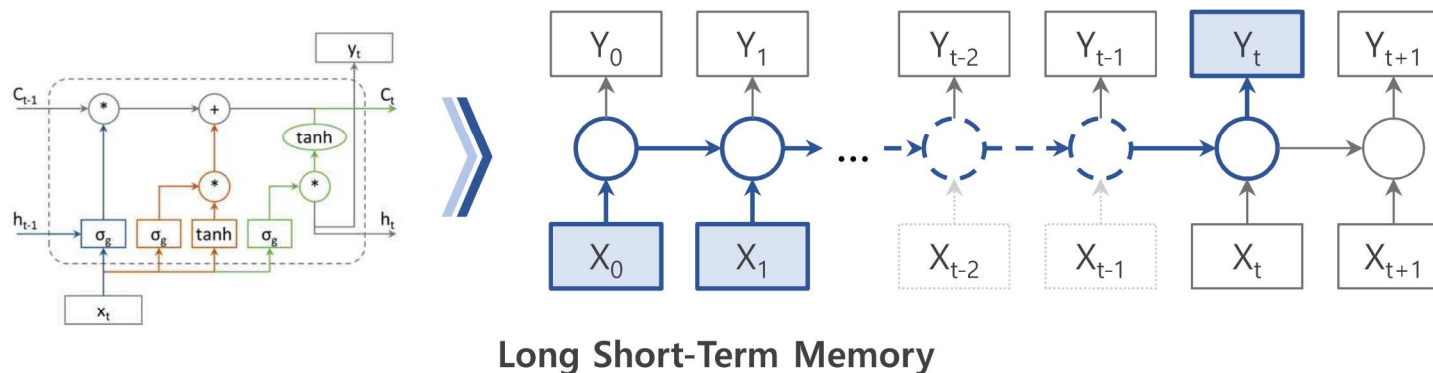


## NER Architecture – LSTM

- RNN architecture shows *vanishing gradient problem*
  - The longer the network serializes, the harder past information becomes to be delivered → learning ability decreases

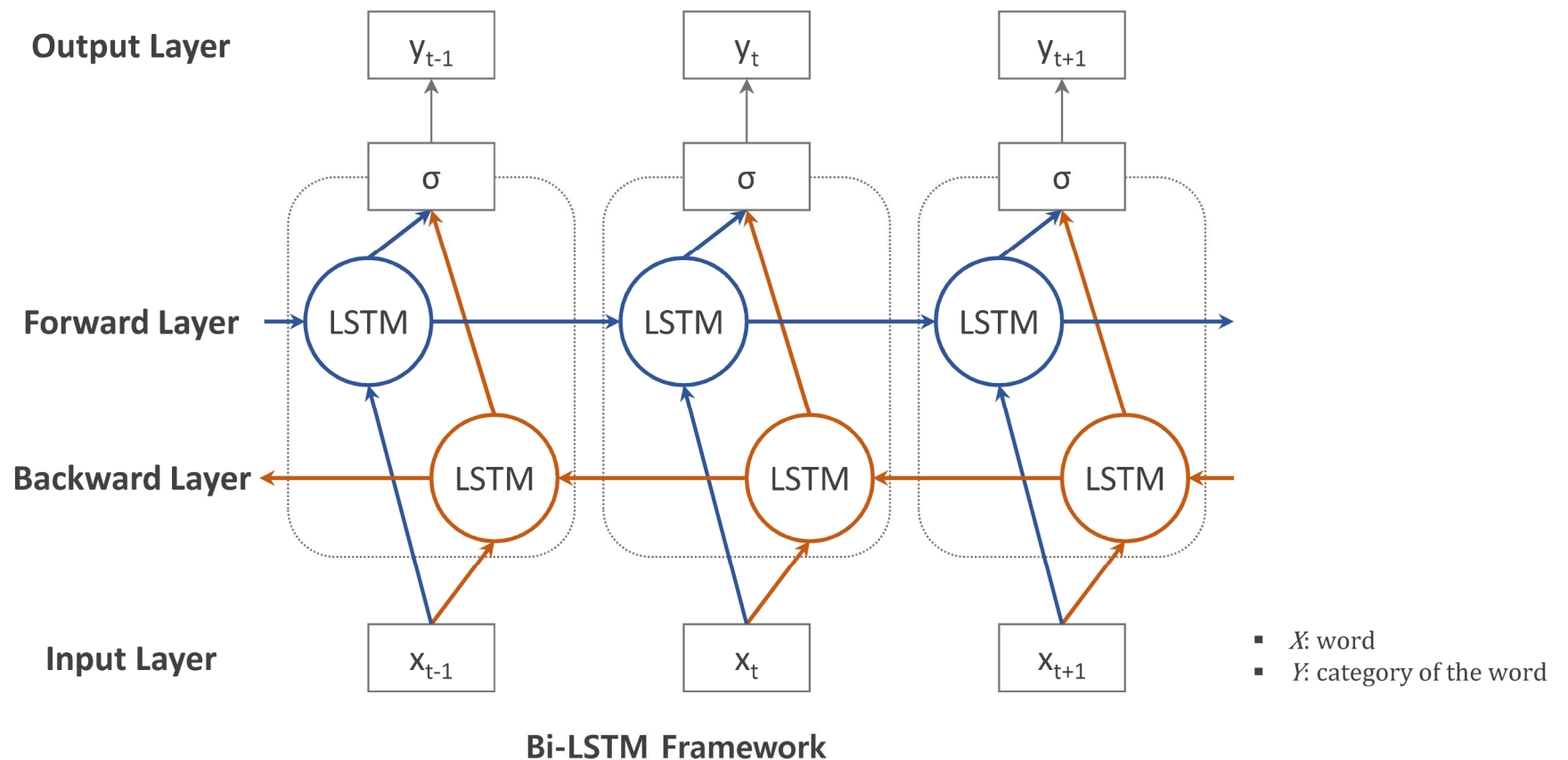


- Long Short-Term Memory (LSTM) addresses the problem with *special architectures in hidden states*



## NER Architecture – Bi-LSTM

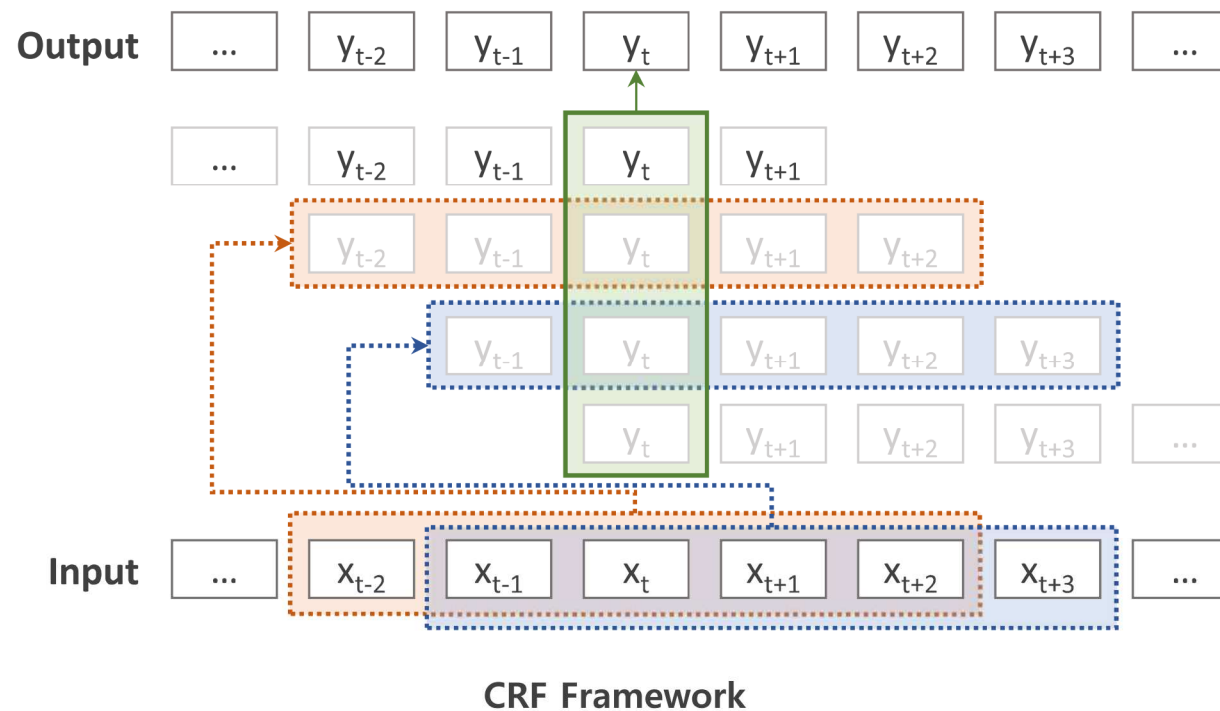
- Bi-directional LSTM considers the *sequential information* of words from *opposite side*



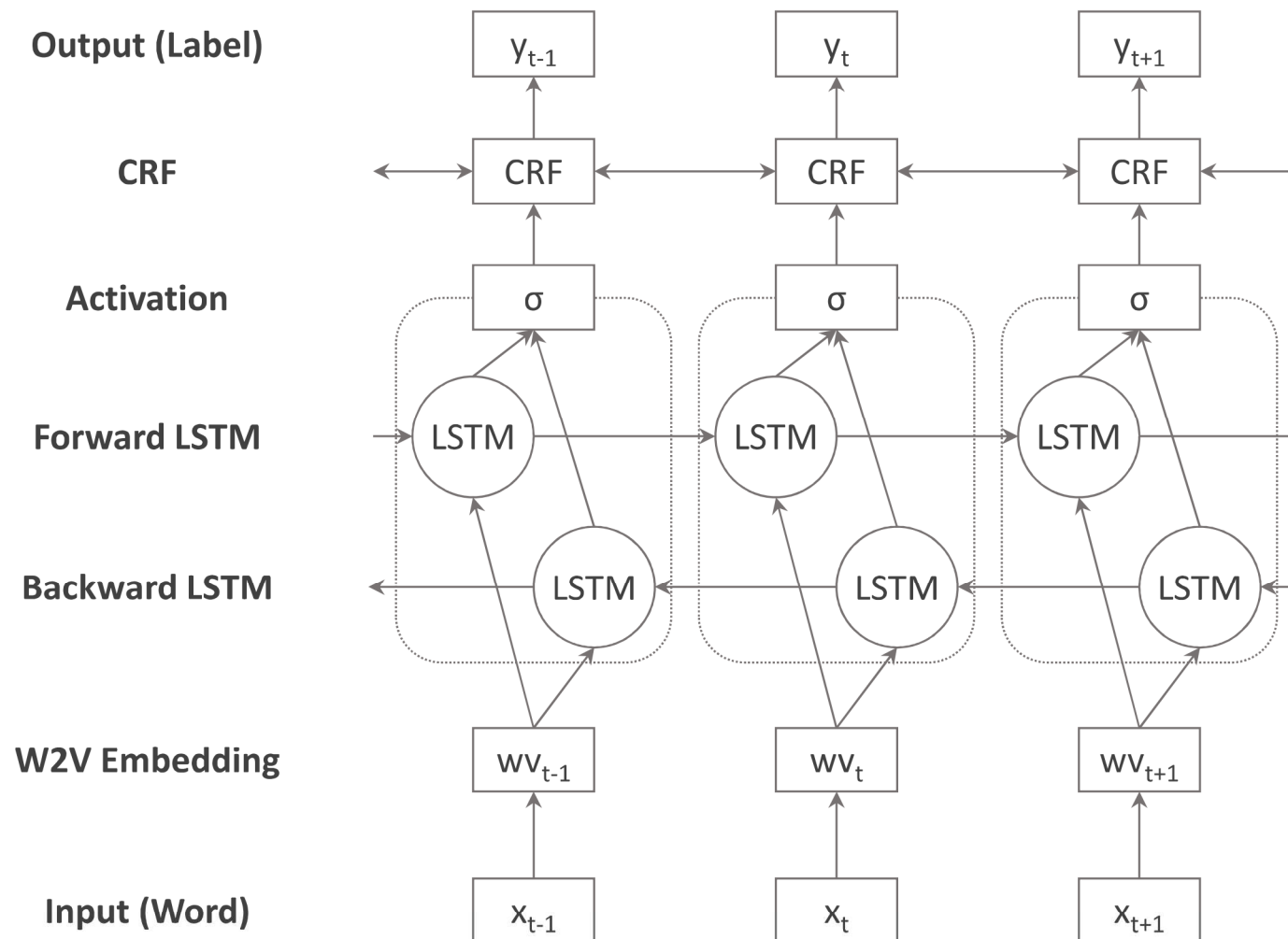


## NER Architecture – CRF

- Conditional Random Field (CRF) *looks around and adjust the output* (i.e., word category)
- To select a category that *maximizes the conditional probability* among *all possible sequences of labels*



## Final NER Architecture



- $x_t$ : word
- $y_t$ : NER category
- $wv$ : word vector

## I NER Results

- Data: 4,659 sentences (Construction Specification)
  - Training: 3,261 sentences (70%)
  - Validation: 1,398 sentences (30%)

NER Labels

Information Type	Category	Examples
Organization	ORG	contractor, engineer, designer
Action	ACT	must submit, have to approve, shall test
Element	ELM	formula, certification, design value
Standard	STD	one month, a week, 38 mm
Reference	REF	AASHTO, ASTM, BS EN ISO
None	NON	and, to, for

## NER Results

- Data: 4,659 sentences (Construction Specification)
  - Training: 3,261 sentences (70%)
  - Validation: 1,398 sentences (30%)

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + True\ Negative}$$

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

Predicted Class

Actual Data		NON	ORG	ACT	ELM	STD	REF	TOTAL
	NON	10,360	10	382	527	63	26	11,368
	ORG	14	571	0	28	0	2	615
	ACT	366	0	4,409	82	15	1	4,873
	ELM	694	12	86	9,273	138	22	10,225
	STD	62	0	25	135	1,764	13	1,999
	REF	10	0	2	36	0	981	1,029
	TOTAL	11,506	593	4,904	10,081	1,980	1,045	30,109

Class	Precision	Recall	F1 Score
NON	0.900	0.911	0.906
ORG	0.963	0.928	0.945
ACT	0.899	0.905	0.902
ELM	0.920	0.907	0.913
STD	0.891	0.882	0.887
REF	0.939	0.953	0.946
<b>AVG</b>	<b>0.919</b>	<b>0.914</b>	<b>0.917</b>

## NER Results

### Original Text

The design and quality control of ACHM surface course mix shall be according to Section 404. Design Requirements for Asphalt Concrete Hot Mix Surface Course (1/2inch [12.5 mm]). Fines to asphalt ratio shall be defined as the percent materials passing the No. 200 (0.075 mm) sieve (expressed as a percent of total aggregate weight) divided by the effective asphalt binder content. (a) Mineral aggregate will be measured by the ton (metric ton). Additives for liquid asphalt, when required or permitted, shall meet the requirements of Subsection 702.08.



NER Model

### NER Results

The<sub>[NON]</sub> design<sub>[ELM]</sub> and<sub>[NON]</sub> quality<sub>[ELM]</sub> control<sub>[ELM]</sub> of<sub>[NON]</sub> ACHM<sub>[REF]</sub> surface<sub>[ELM]</sub> course<sub>[ELM]</sub> mix<sub>[ELM]</sub> shall<sub>[ACT]</sub> be<sub>[ACT]</sub> according<sub>[NON]</sub> to<sub>[NON]</sub> Section<sub>[REF]</sub> 404<sub>[REF]</sub>. Design<sub>[ELM]</sub> Requirements<sub>[ELM]</sub> for<sub>[NON]</sub> Asphalt<sub>[ELM]</sub> Concrete<sub>[ELM]</sub> Hot<sub>[ELM]</sub> Mix<sub>[ELM]</sub> Surface<sub>[ELM]</sub> Course<sub>[ELM]</sub> (1/2inch<sub>[STD]</sub> [12.5 mm])<sub>[STD]</sub>. Fines<sub>[ELM]</sub> to<sub>[NON]</sub> asphalt<sub>[ELM]</sub> ratio<sub>[ELM]</sub> shall<sub>[ACT]</sub> be<sub>[ACT]</sub> defined<sub>[ACT]</sub> as<sub>[NON]</sub> the<sub>[NON]</sub> percent<sub>[ELM]</sub> materials<sub>[ELM]</sub> passing<sub>[ACT]</sub> the<sub>[NON]</sub> No.<sub>[STD]</sub> 200<sub>[STD]</sub> (0.075<sub>[STD]</sub> mm)<sub>[STD]</sub> sieve<sub>[STD]</sub> (expressed<sub>[NON]</sub> as<sub>[NON]</sub> a<sub>[NON]</sub> percent<sub>[ELM]</sub> of<sub>[NON]</sub> total<sub>[NON]</sub> aggregate<sub>[ELM]</sub> weight)<sub>[NON]</sub> divided<sub>[NON]</sub> by<sub>[NON]</sub> the<sub>[NON]</sub> effective<sub>[ELM]</sub> asphalt<sub>[ELM]</sub> binder<sub>[ELM]</sub> content<sub>[ELM]</sub>. (a)<sub>[NON]</sub> Mineral<sub>[ELM]</sub> aggregate<sub>[ELM]</sub> will<sub>[ACT]</sub> be<sub>[ACT]</sub> measured<sub>[ACT]</sub> by<sub>[NON]</sub> the<sub>[NON]</sub> ton<sub>[ELM]</sub> (metric<sub>[ELM]</sub> ton)<sub>[ELM]</sub>. Additives<sub>[ELM]</sub> for<sub>[NON]</sub> liquid<sub>[ELM]</sub> asphalt<sub>[ELM]</sub>, when<sub>[NON]</sub> required<sub>[NON]</sub> or<sub>[NON]</sub> permitted<sub>[ACT]</sub>, shall<sub>[ACT]</sub> meet<sub>[ACT]</sub> the<sub>[NON]</sub> requirements<sub>[NON]</sub> of<sub>[NON]</sub> Subsection<sub>[REF]</sub> 702.08<sub>[REF]</sub>.

**LET'S DO IT!**

# 5 NLP Tutorial

---