Bauhaus-Universität Weimar
Faculty of Media
Degree Programme Medieninformatik

Linking Complex Concepts in Argumentation Knowledge Graphs

Bachelor's Thesis

Lucky Chandrautama

1. Referee: Prof. Dr. Benno Stein
2. Referee:

Submission date: June 17, 2021

Declaration

Unless otherwise indicated in the text or references, this thesis is entirely the product of my own scholarly work.

Weimar, June 17, 2021

. . . .

Lucky Chandrautama

Abstract

A main challenge for construction knowledge bases (KBs) is *concept linking*, i.e., mapping the different mentions of a concept into a single unique entry. While linking "simple" types of concepts such as entities (e.g., "language") has been tackled in literature, few works target linking complex concepts (e.g., "adoption of universal language"). The thesis at hand proposes a simple method for linking complex concepts based on three main steps: concept representation using multiple components of the concepts (e.g., their entities), concept clustering based on semantic similarity, and concept linking based on bidirectional and unidirectional textual entailment. We apply our method to the argumentative knowledge graph (AKG) of Al-Khatib et al., and the results show that with appropriate clustering and entailment threshold, textual entailment can be used to linking concepts that are semantically similar as well as entailed. As such, we are able to not only remove duplicates among the concept instances but also to uncover new implicit knowledge relations in graphs.

Acknowledgements

I would like to give my gratitude first to Khalid for his patience and guidance during the writing of this thesis. There are a lot I learnt from him in the area of NLP and it has been a pleasure to work with him. I want to thank my friend and comrade-in-arms Quang, whom I went through years after years of study with and I share a lot of hardship with. I thank my classmates: Theresa, Anton, Feras, Xiao, Henrik, Jenny, Valentina, Philip T, Phillip M and many others who I haven't met for a while. Without them, my time in Bauhaus would be unbearably difficult. I want to thank Adrian for his support and motivation. His passion and knowledge are truly inspiring. I also want to thank the lecturers, especially Andreas Jakoby. Thanks to him, Quang and I are able to study again after being dropped out from our previous study. To my housemates whom I can trust in taking care of the house, who keep the volume down at my request and whose drinking invitation I have always refused for the past few months, I say thank you so much. Thank you to Liana, Indira, Adna, Halim, Edo, Henry, Idham, Fahri, Suci, Nita, Christanto, Dipa, Hery and Yogi for simply existing. You all bring laughter to my days. Of course, I want to thank my family back in Indonesia for their patience and kindness and for allowing me to pursue my study abroad. To them, I will always be coming back. And God Almighty for His Blessing.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Given the unstructured nature of the information found in digital form, the role of a Knowledge Graph (KG) in storing, processing, and reasoning over this information has become increasingly more prominent. A KG is a data structure that comprises entities of interest as nodes, the relationship between the nodes as edges, and the properties of the entities/relations as attributes [21, 51]. KGs often integrate a knowledge base or certain ontology as a mean to organize the set of possible entities and relations [51], allowing semantic and lexical inference methods to facilitate reasoning capabilities such as acquisition of new knowledge.

Basically, KGs can be utilized in many use cases and applications. In the family of search engine application, for example, KG has been implemented and used by various companies to optimize searching capability (beyond the lexical representation of queries), including Google's KG, DBPedia, Microsoft's Satori and Facebook's entity graph. Moreover, some KGs are applied to incorporate many datasets into a more reliable and intergrated knowledge source [19, 20] and are used to assist learning and enhance the understanding in some scientific fields such as biomedical science [5, 56], business and commerce [59, 66], and juridical domain [10, 17].

The construction of KGs requires addressing several tasks as different challenges may be faced. One of these challenges is that the entities (extracted from unstructured texts) in the graph are not necessarily unique; two related and semantically similar entities may exist in the graph as two separated, unrelated nodes. Entity Linking (EL) is the task of resolve the disambiguation of the different mentions of an entity by linking it to a single unique entity *fingerprint* (e.g., a knowledge data entry in Wikipedia). For instance, mentions of *"auto"* and *"car"* can be linked to *"Car"* on Wikipedia. That being said, EL is a fundamental step in a successful KG construction as it reduces the amount of storage and, more importantly, it increases the efficiency of processing and understanding the KG by eliminating the duplication of nodes and uncover implicit edges.

In most of the constructed KGs in literature, a node denotes an entity. However, several KGs can be built based on semantics that require more complex representation of nodes; nodes with more complex linguistic structure that may consist of multiple entities and different interactions between them. Such *complex concepts* can represent events, principal, and ideas, in addition to entities in a form of phrases or short sentence. For example, *"take player away from academic"* and *"chance for a stable government would significantly decrease after usa withdrawal from iraq"*. Since existing EL tools deal mainly with entity representation of nodes, they may fail in linking *complex concepts*.

The experiments done in this thesis utilizes a type of KG with *complex concepts*, namely argumentation graph. Argumentation graph (AG) is a special type of KGs, where each node represents an argumentative concept and the edge between them is a directed edge that represents a cause-and-effect relation. Although it is quite a new topic of research with high complexity, research on argumentation graph has gained increasing popularity due to its usefulness in various argumentation task such as argumentation question answering [47], sentiment analysis [11], argumentatiive conflict resolution [23, 35] and acquisition of new information [3, 13]. It is also worth to mention that usage of datastructure like argumentation graph improves one's ability to think critically [16, 50].

The thesis at hand investigates ways to link semantically similar complex concepts in argumentation graphs. As mentioned above, EL cannot be used solely for such task. To this end, grouping techniques such as clustering and textual entailment (TE) for complex concepts are used in conjuction with EL. A key contribution of linking argumentative concepts in AG is uncovering implicit relations by creating a path in the graph that is previously non-existent. As a concrete example, consider three concept instances namely *"serious injuries"*, *"physical abuse"* and *"society"* from the same AG. By TE, *"serious injuries"* entails *"physical abuse"*. Since *"physical abuse"* impacts negatively to *"society"* according to the given AG, it can also be implied that *"serious injuries"* impacts negatively to *"society"*, which make sense by human interpretation. By such implication, a new cause effect relation can be therefore acquired between the two concepts that are previously unrelated (Figure 1.1).

The challenge of this task is that *complex concepts* (e.g., argumentative concepts) are often either phrases or short sentences made of several entities and other words as modifiers. Therefore, Natural Language Processing (NLP) techniques are important to identify the semantic of the concepts. On this basis, this thesis employs various NLP and Machine learning (ML) methods for developing a new proposed linking approach. This approach is based on the following steps:

1. Extract the concepts from the graph.

2. Remove their duplicates

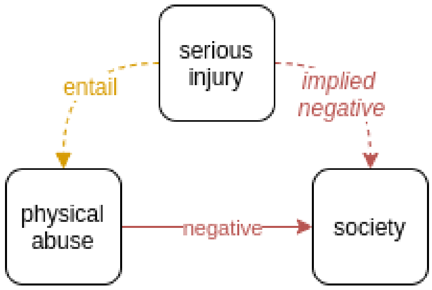


Figure 1.1: A new relation (dotted red line) implied by the entailment and existing relation

3. Obtain various concept properties to be used as part of their representations
4. Perform K-Means clustering from 5 to 1500 for each concept representation
5. Find the most promising configuration (k - clustering and concept representation)
6. Run textual entailment for each cluster of the chosen configuration
7. Link or merge the concepts based on their entailment
8. Create and finally output a new graph.

Applying these steps on the Argumentation Knowledge graph of Al-Khatib et al. [2020], the results show that TE can be used to imply a new argumentative relation out of the existing concept instances and their corresponding argumentative relation with good result. This makes use the nature of bidirectional and unidirectional TE.

The remaining contents of this thesis is structured into 5 Chapters. In Chapter 2, I explain some related works on the topic as inspirations of this work and how their researches might be applicable to this thesis. Then followed by Chapter 2, the overview of the entire workflow for the technique I used is explained. Each major components of the workflow namely concept representation phase and concept linking phase will be explained in detail separately in follow-up sections. Concept representation phase which consists of the preprocessing of the graph and extraction of various forms of concept representations will be explained in Section 3.1. Section 3.2 contains the steps to determine candidates concepts for linking, which include clustering and textual entailment. Chapter 4 addresses the evaluation and finally Chapter 5 is the conclusion.

Chapter 2

Related Works

I look into some related researches in the field of entity linking, argument mining and argument graph construction to get better understanding on the argument instances that needs to be merged and their characteristic as sentences in natural language. I also look into some clustering techniques used for short text and textual entailment as the possible helping tools for merging these identical concepts.

2.1 Entity Linking

Entity linking (EL) is a process in Natural Language Processing (NLP) to determine which entity is mentioned from a source textual document [14]. This is done by detecting the surface form of such entity (how the entity is written) in the source and link each of them to an appropriate instance in a knowledge base (what it is mainly referred as) [22, 43]. EL is important task because some word may have ambiguous meaning or used differently from one source to the others due to the open and decentralized nature of the Web [57]. Such problem of word ambiguity give rise to 2 different challenges that an EL tools needs to tackle [27]: synonymy (different spans of text referring to the same entity) and homonymy (the same name being shared by multiple entities). Different EL techniques can be tailored specifically for either short or long text [34, 38] or for different purposes such as web identifications [58, 63], knowledge extraction [26, 28] or queries [8, 67].

For this thesis, I only consider EL techniques for short text since concepts in argumetation knowledge graph may vary in length which often times are in the form of short texts. As explained by Chen et al. [2018], short texts like tweets and search query provide much less number of words and thus less contextual information. Additionally, short texts are composed differently from a normal sentences used in literature and document since they lack in proper grammatical and/or linguistic structure [33]. Thus, EL tools for short texts is different from other as it requires more than finding the entity with the most probable matching but also contextual meaning of the text itself [43]. Hence, it should be noted that not all EL tools are applicable to this research.

TagMe is a probabilistic approach developed by Ferragina and Scaiella [2010] to resolve the entity-linking problem for short text fragment. It links the short text input to the appropriate Wikipedia pages. TagMe computes the score of relatedness or confidence score between fragments in a short text to their respective Wikipedia pages by collective agreement. A Wikipedia page that corresponds to the highest confidence score is said to be the best match to the fragment.

Another interesting technique is done by Guo et al. [2011] using a graph based method. They make use the Wikipedia graph connectivity by exploiting the name nodes to provide context to the candidate article node and thereby they are able to select the most likely entity to be linked with the mention. Although the idea is a sound one, the argumentative knowledge graph differs from the general knowledge graph like Wikipedia. Firstly, the nodes in argumentative knowledge graph are argumentative concepts which contain multiple mention and therefore there exist also multiple candidate entities. Secondly, the relations between the nodes provide more context in argumentative knowledge graph since the graph itself store cause-effect relation. Therefore, the technique used in this work can be much simplified by concatenating the neighboring node with the concerned node to provide more context to the ambiguous mentions.

It is also important to mention that EL for the purpose of merging identical surface form may be computationally expensive task for large datasets. Elmagarmid et al. discuss clustering as the possible approach to optimize the EL task by reducing the number record comparison. Arasu et al. [2009] also feature clustering as the initial stage of their EL process. Formally speaking, it is defined if entity A finds no similar entities in the cluster, there exist no similar entities to A in the entire graph or dataset. As a result, only a set of few entities are examined at any given moment rather than the entire dataset.

2.2 Clustering for textual data

In general, clustering is a process of grouping together data entries according to their similarity distance. For a textual corpus, a vector is calculated for each sentence based on the occurrence of each word that they have. Such calculation is called feature extraction. It is commonly done by using either

1. **BoW** (Bag of Words): each sentence or concepts is represented as a vector of 0s and 1s based on which words out of the entire corpus exist in the sentence,

2. **tf-idf** (term frequency-inverse document frequency): the number of occurrence each word appears in a sentence is counted and subtracted by the frequency of each word appears in the entire corpus, or

3. **embedding** (or in this case, sentence level embedding): the collection of words and each sentence unique identification in the entire corpus is mapped into feature vectors by making use of a pretrained model.

Both the BoW and tf-idf method have a downside in which both cannot capture the relation between the words and the sentence and preserve their semantic meaning [45]. Therefore, embedding is used in this thesis to create the vectors of concepts for the purpose of clustering. There are two models that commonly used in this manner, namely Doc2Vec [41] and BERT [18]. According to Lau and Baldwin [2016], Doc2Vec performs well in the task of duplicate question detection and the similarity of a pair of sentences prediction. The more advanced one, BERT model has seen its usage in K-Means which leads to a positive results [65] [55]. Both of this model is used in this thesis and will be compared.

The resulting vectors are then measured against one another to determine their similarity distance from one another. One can choose between euclidean distance and cosine similarity. Although it is simpler to use Euclidean distance, the cosine similarity distance yield better results for clustering textual data [60]. After similarity distance are calculated, clustering is then performed.

One of the most commonly used clustering techniques for textual corpus is K-Means clustering. K-Means is a flat clustering technique that groups the record entries into k number of non-hierarchical cluster [36]. With K-Means, k-number of central data points or centroid are chosen (either arbitrarily or deliberately from the record) during the initialization step. Next, each entry of the record is then assigned to a cluster according to the nearest centroid (assignment step) and these centroids are then recalculated by taking the mean value of all the entries assigned to each current centroid (update step). The assignment and update step are performed in alternating manner at each iteration until the centroids do not change significantly.

However, K-Means is not the only commonly used clustering technique for textual corpus. Agglomerative clustering is hierarchical clustering technique [69]. It starts by assigning each data input as a singleton cluster. Unlike K-Means that divides the data simultaneously, agglomerative clustering merges the data in succession. At the beginning each entry is considered as its own cluster. Then at every step after, pairs of clusters are merged based on its closeness or similarity until all clusters have been merged into one big cluster containing all object or until a specified number of cluster is reached. Unlike K-Means that needs access to the feature space or vectors, it only a pairwise comparison of two inputs. Agglomerative clustering is used with some success for short question-answering text [64] and document by making use the similarity between their keywords [48].

Regardless the approach to textual clustering one may take, the results are groups of related textual objects. In addition, relatedness between two concepts seems to give indication if they are semantically similar [52].

2.3 Textual entailment

Since the goal of this thesis is to find a possible way to handle multiple complex concept instances that imply the same meaning, it is helpful to look into paraphrasing and Textual Entailment (TE) task more closely due to the nature of the argumentative concepts that consist of multiple entities of various topics. TE is a unidirectional relation between two input texts, from a premise p to a hypothesis h [15]. The values of TE correspond to *entailment*, *contradiction* and *neutral* relation score between the input sentences. Denoted by $p \Rightarrow h$, a premise p entails a hypothesis h if the truth given by p follows that by h and therefore its *entailment* score is higher than both *contradiction* and *neutral* score. Different groups have different ways to formulate this unidirectional relation. According to Glickman and Dagan [2005], $p \Rightarrow h$ iff $P(h|p) > P(h)$ while Mihalcea et al. [2009] work by the definition that $p \Rightarrow h$ iff h is not informative with respect to p.

While TE is unidirectional relation, paraphrasing or semantic equivalence is bidirectional TE between two input texts since a concept needs to imply the other and vice-versa [4, 61]. Berant et al. [2010] argues that strong or mutual entailment between the two concepts is required for them to qualify as semantically equal/similar as it shows that the two concepts confirms the truth implied from each other ($p \Leftrightarrow h$).

Also related to this thesis, Adler et al. [2012] shows that textual entailment can be used to explore a textual corpus to find the cause-effect relation between the concepts and represent them more meaningfully in a hierarchical entailment graph. Although the research is applied on health-care area, the approach seems to be applicable to a more general corpus as well. The entailment graph is based on the research by Berant et al. [2010] in which inference between two concepts can be easily obtained. Berant et al. demonstrate that entailment have transitive property and entailment graph is useful to structure prepositions hierarchically. Their work shows that the entailment hierarchy have a specific - general relation between the parent node $\mathcal{A}$ and child node $\mathcal{B}$ where $\mathcal{A}$ is a hypothesis to the entailing $\mathcal{B}$ ($\mathcal{B} \Rightarrow \mathcal{A} = p \Rightarrow h$). Although the prepositions used in their research are constrained to the same topic and made of an SPO structure, the relation between the premise and hypothesis can be useful to infer a new causality. Considering that argumentative graph expresses a cause-effect relation between the concept, usage of textual entailment may lead to a more accurate argumentative graph.

Another useful property of textual entailment is that it has a strong correlation with semantic relatedness. Vo and Popescu [2016] indicates that high relatedness results generally in entailment and low relatedness in neutral. Although there is no conclusive result about the contradiction, there is a strong notion that contradiction may also follow such relation. Regardless, we can thereby determine that candidates for semantic equivalents are all related. Thus, semantically equal or similar concepts will always be found together in the same cluster of related concepts which aligns with the assumption made by Arasu et al. [2009].

Despite the huge potential of textual entailment in many areas, the research in the area of textual entailment is relatively new at the moment. However, the implementation developed by Gardner et al. has shown encouraging results and steadily maintained since its first public release. The variant of their textual entailment using RoBERTA model [42] in particular performs well [68].

2.4 Argument Mining

Since this study is dealing with argumentation linking, it is also beneficial to investigate how such argumentation is extracted and how ambiguity on argumentative concept may occur. The process of extracting argumentative concept from a body of text is often called argument mining or automated argument analysis. Argument mining is used to identify and extract automatically the structure and components of arguments found in sources [53] [46]. This task is a follow-up to the more expert dependent task of argument analysis as the ever-increasing source of argumentation data makes it even harder and more tedious to be extracted manually even with the help and supervision of human expert [46].

Lawrence and Reed [2020] break argument mining process into several interrelated tasks which has different level of complexity. Therefore, argument mining only needs to be performed according to the purpose of the resulting argumentation graph. For example, it is sufficient to tackle the problem on identifying the argument component to inspect the range of argument in an essay [49] and verify the stance of such essay [54]. The more challenging goals like reconstruction of enthymemes [24] or finding out the relation between argument components [9] require identification of clausal and relational property which significantly more difficult.

However, many of the techniques used for argument mining rely heavily on structural, syntactic, lexical and pragmatic features of the source like cue words found in common argument scheme [24], the debate structure that is specific to the debate's page [2] or the similarity and relations between their functional argument components [30]. Studying these approaches, I believe that there is a need to finetune the mined argumentative concepts since their semantic features are not taken into account and therefore there could be some duplicates especially when the argumentative concepts are harvested and compiled from different sources.

2.5 Argumentation Graph

Argumentation knowledge graph (or argumentation graph) is a specific kind of knowledge graph that stores the argumentative concept entities in the nodes and the causality between them in their relation. Arguments around a topic or concept can be obtained by traversing the graph. Al-Khatib et al. [2020] model an argumentation graph through argumentative concept instances as its nodes and their effect relation as its edges. In order to legitimize the argument depicted in the graph, they also store public sentiment of the concept and the entity (surface forms and grounding) contained in the concept as the attributes of the node. Moreover, concept consequences or effect from one concept to the other are treated as attributes of the edge. This model is similar to the ideal model for teaching suggested by Davies et al. [2019] in a sense that argument around and related to a concept and type of causality between concepts can be obtained by traversing the graph and edges. The biggest difference between these two models is that Khatib et al. simplify the concept instances in which they make no distinction whether they are a claim, reasoning or conclusion.

Similarly, Cayrol and Lagasquie-Schiex [2013] focuses their study on the relations between the concepts in bipolar argumentation framework. In their research, bipolarity refers to support and attack relation towards an argument. However, what makes this framework differs from that of Al-Khatib et al. is that a node is byitself an argument (complex concept) instead of a more general concept.

Chapter 3

Approach

At first, EL seemed a very promising procedure to acquire mergable and linkable concepts. However, it is found that sole usage of EL in this case was not as reliable as presumed because:

1. Complex concepts are mostly associated with multiple entity due to the existence of different mentions within the phrase or sentence.

2. Phrases and sentences does not contain only mentions but also other words such auxiliary verbs or adjective

3. Limitation of the EL tools and its knowledge base to recognize mention in the concepts.

Therefore, a more comprehensive workflow needs to be developed for the purpose of finding semantic similarity in conjunction with EL.

Inspired by Vilariño et al. [2012], TE can be applied to find similar concepts candidates. This aligns with the statement made by Berant et al. [2010] and Adler et al. [2012] that two concepts p and h are semantically equal if they confirm the truth from each other by entailment ($p \Leftrightarrow h$). In addition, unidirectional nature of TE may be able to be used to imply causality between concepts in AG as the premise p seems to be more informative [44] and more specific [1] than the hypothesis h.

However, TE is very time-consuming and hardware-demanding. In addition, the state-of-the-art approach by Gardner et al. [2018] involves a pairwise unidirectional comparison between the concept entries in the dataset making the process for high number of entries extremely long. Since a corpus or a concept graph in most cases includes thousands of concept entries, it is very helpful to group the entries together. For this I follow the assumption made by Arasu et al. [2009] that entries are always grouped together properly based on its semantic relatedness so that the similarity concept candidates only exists within the same cluster. This approach aligns with the work by Pedersen et al. [2007] who recognize that semantic similarity is a subset of relatedness where semantic equivalence is the highest degree of similarity. In this thesis, I make no distinction between semantic similarity and equivalence. Since TE is a unidirectional relation and entailment is a subset of relatedness [62], the relationship between these semantic classes can be summarized as Venn diagram in Figure 3.1.

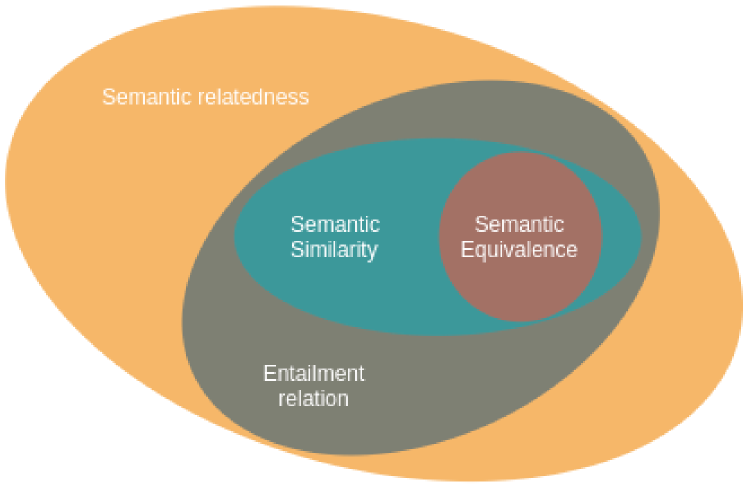


Figure 3.1: Semantic Class Membership

Recognizing the relationship among these semantic classes, clustering needs to be done first before TE for the purpose of grouping the concepts based on their semantic relatedness. To this end, K-Means clustering technique with cosine similarity is used. The resulting clusters of concepts are configured based on how concepts are represented as input. As mentioned above, concepts can be represented in various ways. This is where both EL and WIKIPEDIA article category scraping play a role in this research. Thus using these tools a concept is represented for the purpose of clustering as:

(a) its original textual content

(b) concatenation of its mentions

(c) concatenation of its grounding

(d) concatenation of its text and groundings (*e*) concatenation of its text and the WIKIPEDIA article page category based on its grounding

(*f*) concatenation of its text, the WIKIPEDIA article page category based on each of its groundings and its groundings themselves.

The concepts text as well as the concept itself and their relation is obtained from the input graph. Denoted by $s \rightarrow t$ for positive relation and $s \nrightarrow t$ for negative relation, argumentative relation from source concept s to the target concept t refers to how s affect t (either positively or negatively) The input graph used for this thesis is provided by Al-Khatib et al. [2020] as part of their research. It consists of some metadata, 5016 concept instances as nodes and 17229 relationships as links. Each concept and relationship have their own properties as described in their paper.

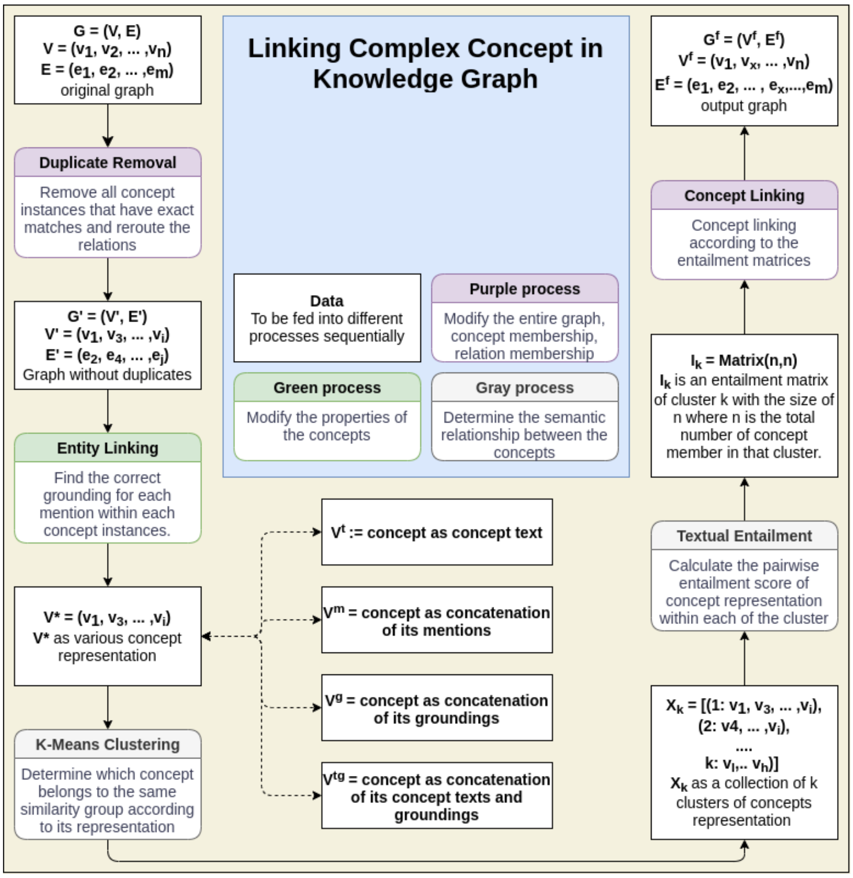


Figure 3.2: Concept Linking Workflow

To summarize, the entire workflow run in the following order (as illustrated in Figure 3.2):

1. Extract the concepts from the graph.
2. Remove their duplicates
3. Obtain various concept properties to be used as part of their representations
4. Perform K-Means clustering from 5 to 1500 for each concept representation
5. Find the most promising k - clustering
6. Run textual entailment for each cluster
7. Link or merge the concepts based on their entailment
8. Create and finally output a new graph.

The step 1 - 3 belong to the Concept Representation Phase (Section 3.1), where the representations of each concept instances are formed using EL (Subsection 3.1.2) with prior Duplicate Removal. The step 4 - 5 belong to the Concept Grouping Phase (Section 3.2), where the concepts representations are grouped using K - Means clustering technique. Finally, TE is performed to the member of each cluster for step 6 - 8 (Section 3.3).

3.1 Concept Representation

The goal of this phase is twofold:

1. Preprocess the graph by removing obvious duplication
2. Extract various concept representation from each concept nodes

From the graph, a concept instance is defined as a node with the following attributes:

- concept text: the original textual content of the concept
- ID: unique identification key
- value: concept consequence, a commonly agreed good or bad sentiment on the concept

Through this phase, some concept instances will be removed and remaining ones will be updated to contain additional attributes. The additional attributes are shown in Table 3.1 with ID, concept text and value remains unchanged. The newly added attributes will be used to form various concept representations (Table 3.3) as follows:

(a) its original textual content

(b) concatenation of its mentions

(c) concatenation of its grounding

(d) concatenation of its text and groundings

ID	n1-426
concept text	national renewable energy
value	n/a
mention	national, renewable energy renewable, energy renewable energy standard
grounding	nation, renewable energy, renewable portfolio standard

Table 3.1: The attributes of concept instance $n1 - 426$ after EL

3.1.1 Duplicate Removal: Preprocessing the graph

Duplicate removal works by removing any clones of concept texts that exist in the graph with different ID. After they are removed, all relation associated with these concepts are rerouted to a remaining clone while original relations between clones are simply removed. This is done first before all other processes because it is very quick and effective to simplify graph. I considered employing a near-duplicate instead of exact match to yield more distinct set of concept instance and compact graph. However, I opt for a safer approach and use the exact match instead since some concepts instances from the input KG are not only phrases or short sentences but also acronyms and one word-long texts. As such, slight differences between concept text can refer to completely different meaning. Thus, a near-duplicate may potentially remove concepts that is not supposedly regarded as a clone of another. The result of this process is a list of concept instances whose text is uniquely different from one another.

3.1.2 Entity Linking

Since a complex concept holds so much information, it can be represented in many ways. By performing EL on each of these concepts, key information can be extracted in the form of their mentions and grounding which then are used to represent these concepts beyond its original textual content. For this, TagMe[1] is used as an EL tool (Algorithm 3.1). The results are list of mentions and groundings for each concept instance. They are stored as additional properties of each concept.

[1] `https://sobigdata.d4science.org/web/tagme/tagme-help`

Algorithm 3.1: singleEntityLinking(c, A)

```
/* EL with only single concept.                    */
```
Input : $x :=$ A concept instance from graph
$A :=$ EL tool
Output: A modified concept instance with only single EL results
$e = A.annotate(c.text)$
$c.entities = e$
return c

Even though TagMe is developed as EL tools for short text, it performs poorly for some concepts. Such that, some mentions are referred to a grounding that is out of context considering the complete text and the neighboring concept nodes. In theory this will be more prominent for concept that has very short text. For example: a concept text *"screening"* is grounded to *Screening (medicine)*[2] with the confidence score 0.061 by using Algorithm 3.1: *Single EL*. Considering its adjacent concepts (Figure 3.3), this grounding is not relevant.

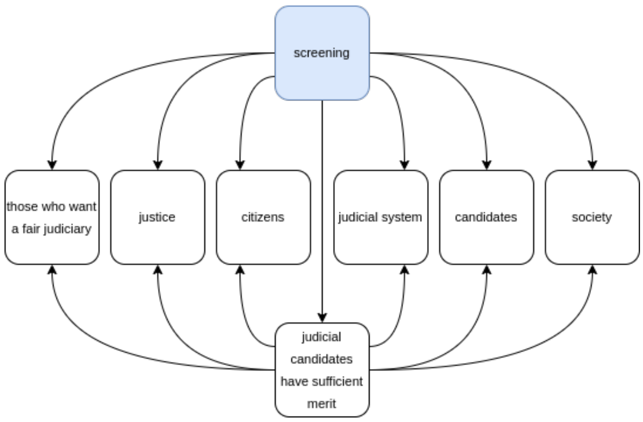


Figure 3.3: Concept instance *n1-80* (*"screening"* in blue) with its adjacent concept instance

To mitigate this problem, the subgraph where a concept belongs to needs to be considered while performing EL task (Algorithm 3.2: *Agreed EL*). This is done by finding all the concepts that directly connected (both inwards and outwards) to the original concept. Then each of the neighbors' text is concatenated with the original concept text forming a pair. Each pair creates a combined text then it is annotated by TagMe. As seen in Table 3.2, the various combined texts are created for *"screening"* using its neighbors, which lead to different groundings.

When pair annotations are completed, the agreed entities needs to be found. To do this, unique mentions and groundings are counted and the maximum is taken

[2]`https://en.wikipedia.org/wiki/Screening_(medicine)`

Combined texts	Grounding of *"screening"*	Score
screening candidates	*Sampling (statistics)*	0.0085
screening those who want a fair judiciary	*Sampling (statistics)*	0.0986
screening judicial system	*Sampling (statistics)*	0.1995
screening justice	*Genetic testing*	0.1802
screening society	*Screening (medicine)*	0.1765
screening citizens	*Sampling (statistics)*	0.1749
screening judicial candidates have sufficient merit	*Sampling (statistics)*	0.1399

Table 3.2: Combined text as input and grounding for *"screening"* using *Agreed El*

as the entities. More specifically, it aims to take the most dominant grounding for each original mention from the single annotation detected in each pairing. From the results seen in Table 3.2, *Sampling (statistics)*[3] is chosen as the grounding for *"screening"* since it is detected for 5 out of 7 pairings. In case multiple equally dominant groundings from the pairings exist, the one with the highest average confidence score is chosen. Using *Agreed EL* will hopefully yield a more context accurate results for not only concepts with short text but also ones with longer text, which holds several mentions, as the surrounding nodes can give some context to the original concept.

With these results, each concept instance has an additional attribute, namely entity list, where each element refers to the EL results from the Algorithm 3.2:*Agreed El*. In addition to the concept text, the following concept representations can be formed by using the agreed entities:

(*a*) its original concept text

(*b*) concatenation of its mentions

(*c*) concatenation of its grounding

(*d*) concatenation of its concept text and groundings

For example, the result for the concept *"national renewable energy"* can be represented in the form of *(a)* - *(d)* as seen in Table 3.3. Because *mentions* are essentially a subset of the original concept text, they hold even less information and do not provide any more context to the concept. Thus, any further use of mentions is omitted.

With these 4 forms of concept representation, concept instances are ready for the next stage of processing namely clustering by K-Means.

[3] `https://en.wikipedia.org/wiki/Sampling_(statistics)`

Algorithm 3.2: agreedEntityLinking(c, A)

```
/* EL with single concept and pairing with its
   neighbor.                                    */
Input  : c := A concept instance from graph
         A := EL tool
Output: A modified concept instance with both single and agreed EL
        results
e = A.annotate(c.text)
c.singleEntities = e
foreach n ∈ c.neighbors do
    p = c.text + n.text
    f = A.annotate(p)
    c.addPairResults(e)
end
g = findAgreement(c.getPairResults(), e)
c.agreedEntities = g
return c
```

(a)	national renewable energy
(b)	national renewable energy renewable energy renewable energy standard
(c)	nation renewable energy renewable portfolio standard
(d)	national renewable energy nation renewable energy renewable portfolio standard

Table 3.3: Concept representation of *"national renewable energy"*

3.2 Concept Grouping

In this section, concept instances will be grouped together according to the six forms of their concept representation. The goal of this stage is to group the concept instances based on their relatedness. This is done by K-Means cluster. It reduces the number of record comparison that needs to be done by textual entailment.

Taking the various concept representations resulted from the previous steps, concepts instances are clustered non-hierarchically using K-Means[4] and cosine distance [5] as the similarity metrics. To produce the feature space required to perform this clustering technique, I have the liberty to use and test two popular embed-

[4]https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

[5]https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine_similarity.html

ding models, Doc2Vec[6] and BERT[7]. Hence, there are 12 clustering configurations in total: two embedding models for each of the six concept representations.

Because domain knowledge (how many clusters is supposed to exist) is unknown, the concepts instances are incrementally clustered from k = 5 to 1500 (both extremes are included) with 5 units intervals. This amounts to 300 clustering results for each configuration. I opt to use 5 units intervals in order to get a much faster process. Although it is definitely helpful for better result, my assessment determine that an approximation of relatedness is sufficient considering the quality of each cluster. Generally, the scores do not differ from one cluster to next so significantly that it justifies the longer processing time for accuracy. Hence, I consider 5 unit intervals to be sufficient.

Based on manual observation with the help of intrinsic scoring, k around 300 to 400 seem to provide a sufficiently good results and thus, enclose a good number of linking candidates. This is important because higher k leads to faster pairwise textual entailment processing but reduce the accuracy because a number of linking candidates may potentially be located in different clusters. With this in mind, clusters for each configuration are chosen and undergone the next step of grouping, namely textual entailment.

3.3 Concept Linking

With the chosen cluster of the concept representations, TE is performed. The goal of this phase are twofold to find the candidate of semantic equivalences and to create new argumentative links based on the TE results. Consider the following cases:

(1) Bidirectional textual entailment (BTE): *are the group of concept instances which entail one another semantically equivalent?* (Subsection 3.3.1)

(2) Unidirectional textual entailment (UTE): *does the premise imply a new relation, given an existing argumentative relation?* (Subsection 3.3.2)

(3) Combination of UTE and BTE: *does implied relation due to UTE involving concept instances that are semantically equivalent make sense?*

The case (1) handles sematically equivalent concepts (Subsection 3.3.1) by using BTE. Groups of mutually entailed are merged into a *composite concept* then all original argumentative links or relations are re-routed into the newly created *composite concept* (Figure 3.4). This can potentially reveal new knowledge by bridging two

[6]`https://radimrehurek.com/gensim/models/doc2vec.html`
[7]`https://www.sbert.net/docs/pretrained_models.html`

concepts that previously have no relation (both directly or indirectly) and simplify the subgraph.

(a) Before merging A, B and C **(b)** After merging A, B and C into X

Figure 3.4: A, B and C are semantic equivalents. Merging them to X create a link between G and E and reveal a cycle with F

The case (2) refers to creation of new relation due to TE (Figure 3.5). Since UTE means that the truth in the premise concept is implied in the hypothesis concept, argumentative relation can also be formed to a third concept. This is divided into two subcases (Subsection 3.3.2): *(2a) entailed to the source* and *(2b) entailed to the target*. Both address whether the newly created relations make sense (dotted purple arrow).

(a) Blue arrow: X entails to Y **(b)** Dotted purple arrow: Implication

Figure 3.5: It is implied that Z affect X because X entail Y and Y is affected by Z

The case (3) emerges because approaches for handling case (1) and case (2) are carried out subsequently on the same argumentation graph. As such BTE and UTE may not be isolated case for *composite concept*. Case (3) is essentially similar to the case (2). The difference is that it handles specifically the *composite concept* instances whose components entails to the other concept instances.

The implementation of TE used in thesis is developed by AllenNLP[8] [29]. Described as a triplet of $entailment$ $\mathbb{E}$, $contradiction$ $\mathbb{C}$ and $neutral$ $\mathbb{N}$ score which total to 100%, TE score between a premise and a hypothesis can be illustrated as point in a triangular coordinate system as shown in Figure 3.6. Hence, tendency to any of the extremes indicates their type of unidirectional entailment relationship. For example, any point within the area near the entailment corner (shown in green) shows that the concept p entails the other h (denoted by $p \Rightarrow h$).

[8]`https://demo.allennlp.org/textual-entailment/roberta-snli`

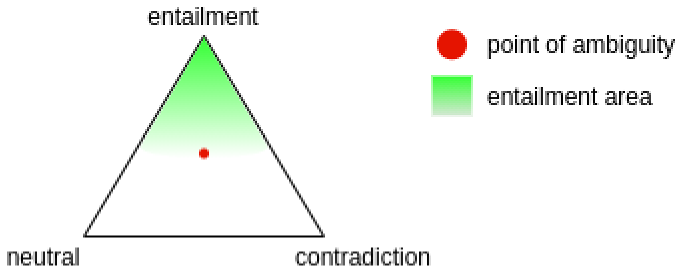


Figure 3.6: Textual entailment coordinate system

However, there is neither reference on how this *entailment area* that is defined nor a clear scoring threshold. What is clear is that ambiguity occurs when all three relations type are equal (shown as red point in the middle of the triangle). At this point, the ratio of the scores $\mathbb{E} : \mathbb{C} : \mathbb{N} = 1 : 1 : 1$. Using this ratio, I define the *entailment* ratio of $\mathbb{E} : \mathbb{C} : \mathbb{N} = \varepsilon : 1 : 1$, where ε is an *entailment factor*. For the purpose of starting assessment, initial entailment factor is defined as $\varepsilon_0 = 1.1$. Which means, $\mathbb{E}$ between p and h needs to be at least 10% greater than the other two scores for them to qualify as $p \Rightarrow h$. To give some clarity, consider the following cases of the entailment score $\mathbb{S}_{p,h}$, where p is the input premise and h is the input hypothesis with concept instance $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ as inputs:

- $\mathbb{S}_{A,B}$ = $(\mathbb{E}, \mathbb{C}, \mathbb{N}) = (40.00\%, 35.00\%, 25.00\%)$. Thus, $\mathcal{A} \Rightarrow \mathcal{B}$ since $\mathbb{E} > \varepsilon_0 \times \mathbb{C} \wedge \mathbb{E} > \varepsilon_0 \times \mathbb{N}$.
- $\mathbb{S}_{B,A}$ = $(\mathbb{E}, \mathbb{C}, \mathbb{N}) = (51.30\%, 48.69\%, 0.01\%)$. Therefore, $\mathcal{B} \nRightarrow \mathcal{A}$ and their entailment relation are considered ambiguous between *entailment* and *contradiction* since $\mathbb{E} < \varepsilon_0 \times \mathbb{C}$ even though $\mathbb{E} > \varepsilon_0 \times \mathbb{N}$.

The entailment factor ε are then applied pairwise bidirectionally to the concept instances. For this, *concept text* is used as the representation of the input pair of concept instances as to avoid a misinterpretation of the concepts and to capture their actual semantic meaning. Then, a matrix of entailment scores of all pairwise combination within each cluster is obtained. For example, a cluster number 178 has the following matrices: $\mathbb{E}$ as seen in Table 3.4, $\mathbb{C}$ in Table 3.5 and $\mathbb{N}$ in Table 3.6.

$\mathbb{E}_{178}$	n2-46	n2-1184	n1-764	n1-2217
n2-46	1.0000	0.0003	0.0002	0.0004
n2-1184	0.0005	1.0000	0.0008	0.0005
n1-764	0.0006	0.9677	1.0000	0.0005
n1-2217	0.0006	0.0003	0.0000	1.0000

Table 3.4: $Entailment$ $\mathbb{E}$ matrix of cluster 178

$\mathbb{C}_{178}$	n2-46	n2-1184	n1-764	n1-2217
n2-46	1.0000	0.9922	0.9921	0.0025
n2-1184	0.0137	1.0000	0.0007	0.0015
n1-764	0.0554	0.0005	1.0000	0.0016
n1-2217	0.9921	0.9966	0.9997	1.0000

Table 3.5: $Contradiction$ $\mathbb{C}$ matrix of cluster 178

$\mathbb{N}_{178}$	n2-46	n2-1184	n1-764	n1-2217
n2-46	1.0000	0.0075	0.0078	0.9971
n2-1184	0.9858	1.0000	0.9985	0.9980
n1-764	0.9440	0.0318	1.0000	0.9979
n1-2217	0.0073	0.0032	0.0003	1.0000

Table 3.6: $Neutral$ $\mathbb{N}$ matrix of cluster 178

This approach is carried out for every cluster. Because merging of semantic equvalent concepts into *composite concepts* due to BTE is carried out first, the linking due to UTE needs to take into account the TE relation of the components of *composite concepts*. If any of the components has UTE with other concepts or a component of another *composite concept*, linking approaches will be carried out.

The effectiveness of this heuristic and this phase overall will be assessed using a survey. The answers of the survey will become an initial indication for an ideal TE result (sensible merging and linking candidates). The value of entailment factor ε will then be adjusted as to match the ideal TE results as close as possible. The assessment and survey answers as well as the adjusted factor result are further explained in Section 4.3.

3.3.1 Case (1): Bidirectional Textual Entailment (BTE)

Because candidates of semantic equivalence might not be only two but rather several concept instances within any given cluster, a heuristic needs to be used (Lemma 3.3.1.2). According to Berant et al., TE has a transitive property.

Lemma 3.3.1.1 (Berant et al. [7]). *If* $a \Rightarrow b$ *and* $b \Rightarrow c$ *then* $a \Rightarrow c$

By extension, I consider bidirectional textual entailment (BTE) to have a transitive property as well.

Lemma 3.3.1.2. *If* $\mathcal{A} \Leftrightarrow \mathcal{B}$ *and* $\mathcal{B} \Leftrightarrow \mathcal{C}$ *then* $\mathcal{A} \Leftrightarrow \mathcal{C}$

Proof. Assume Lemma 3.3.1.2 is false. BTE can be defined as two UTEs in opposite direction. If transitivity on BTE cannot be held true then one of the two UTEs is false. If any of the two UTEs is false, Lemma 3.3.1.1 cannot be true. Hence, there is a contradiction and Lemma 3.3.1.2 must be true. □

Therefore, it is enough to defined concept instances $\mathcal{A}, \mathcal{B}, \mathcal{C}$ as semantic equivalence $\mathcal{A} \Leftrightarrow \mathcal{B} \Leftrightarrow \mathcal{C}$ when $\mathcal{A} \Leftrightarrow \mathcal{B}$ and $\mathcal{B} \Leftrightarrow \mathcal{C}$ are known. However, this assumes that *entailment* has always 100% confidence score which may not be realistic with the current state-of-the-art implementation. This may lead to somewhat inaccurate merging candidates. Examples of this are shown in Table 3.7.

New ID	Entailment results
nx-151114592	"kid" ⇔ "young person" ⇔ "nclb child" ⇔ "one child" ⇔ "youth" ⇔ "child"
nx-183468704	"health industry" ⇔ "healthcare worker" ⇔"health consciou" ⇔ "healthcare system" ⇔ "health care supplier" ⇔ "healthcare industry" ⇔ "health care cost" ⇔ "health care subscriber" ⇔ "health crisi" ⇔ "health risk" ⇔ "health information" ⇔ "spread health" ⇔ "health care" ⇔ "healthcare" ⇔ "health care system"

Table 3.7: Bidirectionally entailed concept instances with $\varepsilon = 1.1$. These are to be merged and hence, have a new ID

3.3.2 Case (2): Unidirectional Textual Entailment (UTE)

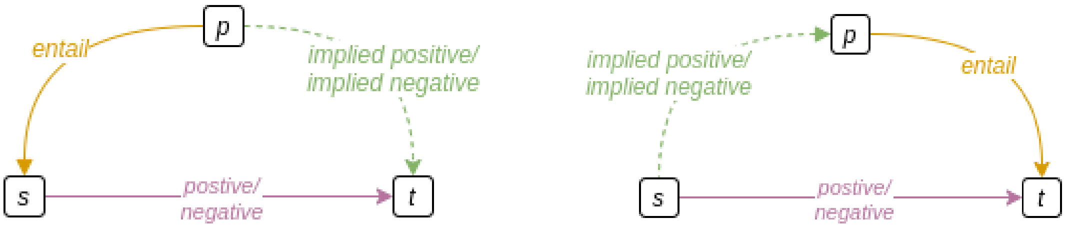
To handle concept linking for unidirectional textual entailment (UTE), I consider these following two cases (Figure 3.7): *(2a) entailed to the source* and *(2b) entailed to the target.* From the chosen $k-$ clustering result, triplets of concepts (p, s, t) are obtained based on their unidirectional TE, where $p \neq s \neq t$. The reasoning behind the approaches on case (2a) and (2b) is based on the definition of the entailment that premise $p \Rightarrow$ hypothesis h iff h is not informative with respect to p [44]. My observation seems to align with this statatement as p tends to be more specific and holds more information than h.

Case (2a): Implied relation due to entailed argumentative source

Given that there is an argumentative relation from a source concept (s) to a target concept (t), an implied argumentative relation from a premise concept (p) to t may emerge since $p \Rightarrow s$ as seen in Subfigure 3.7a.

Remark 3.3.2.1. *Given that* $s \rightarrow t$ *and* $p \Rightarrow s, p \rightarrow t$

The reasoning behind this approach is that since p shows more specific meaning to s due to UTE, p also affects t just like s does. Take example number 1 from



(a) Entailed to the source

(b) Entailed to the target

Figure 3.7: e is a premise concept and s and t concepts have an argumentative relation and are used as hypothesis

the Table 3.8, it is implied that *"producer of healthy food"* affect *"dieter"* positively ($p \rightarrow t$) since entails *"producer of healthy food"* to *"healthier menu"* ($p \Rightarrow s$) and *"healthier menu"* affect *"dieter"* positively ($s \rightarrow t$). Likewise, this applies to the negative argumentative relation ($s \nrightarrow t$).

No.	p	s	t	$arg(s,t)$	$\mathbb{E}(p,s)$
1	"producer of healthy food"	"healthier menu"	"dieter"	positive	0.9019
2	"human greenhouse ga emission"	"greenhouse ga emission"	"air pollution"	negative	0.9508
3	"terrorist attack"	"terrorism"	"person targeted by terrorism"	negative	0.9768

Table 3.8: Entailed to source triplet samples

Case (2b): Implied relation due to entailed argumentative target

The similar logic can be applied when premise p entails to target t instead of source s. Again given that $s \rightarrow t$, it is implied that $s \rightarrow p$ since $p \Rightarrow t$ as seen in Subfigure 3.7b.

Remark 3.3.2.2. *Given that $s \rightarrow t$ and $p \Rightarrow t$, $s \rightarrow p$*

This is because p seems to have the truth and specific meaning to t and thus s will also affect p. Take triplet number 1 from the Table 3.9 as an example, it is implied that *"needle exchange"* affect *"risk of pulmonary embolism"* negatively ($s \nrightarrow p$) since *"risk of pulmonary embolism"* entails *"infection"* ($p \Rightarrow s$) and *"needle exchange"* affect *"infection"* negatively ($s \nrightarrow t$). Likewise, this applies to the positive argumentative relation ($s \rightarrow t$).

No.	p	s	t	$arg(s,t)$	$\mathbb{E}(p,s)$
1	"risk of pulmonary embolism"	"needle exchange"	"infection"	negative	0.7233
2	"eu budget"	"independent kosovo the most viable"	"eu"	positive	0.9553
3	"develop the party platform"	"flag controversial issue for the public"	"party"	positive	0.8407

Table 3.9: Entailed to target triplet samples

3.3.3 Case (3): UTE involving semantically equivalents

This case is essentially similar to the previous UTE cases. The difference is that instead of basic concept instance, any of the triplet (p, s, t) is *composite concept*. Consider a premise p, and two concept instances a and b. Given that $a \Leftrightarrow b$, *composite concept* X is formed with (a, b) as the component. If $p \Rightarrow a$ is valid, $p \Rightarrow b$ is also valid.

Lemma 3.3.3.1. *Since* $a \Leftrightarrow b$, $p \Rightarrow a$ *when* $p \Rightarrow b$*.*

Proof. Assume $p \Rightarrow a$ is false. Because $a \Leftrightarrow b$, b hold the same semantic meaning as a and thus $b = a$. Since $p \Rightarrow b$ is valid, $p \Rightarrow a$ must also be valid. This is a contradiction on the assumption. Therefore, Lemma 3.3.3.1 must be valid. $\square$

Hypothetically, this can also be applied for the premises that are formed into *composite concept* considering TE has a transitive property. With this, the same logic as the subcases of UTE can be applied. Without the loss of generality, consider a triplet of *composite concepts* $(\mathcal{P}, \mathcal{S}, \mathcal{T})$ with arbitary number of components. Given that, $(p \in \mathcal{P}, s \in \mathcal{S}, t \in \mathcal{T})$, the following UTE cases is also valid (the negative argumentative relation also apply here):

(a) Implied relation due to entailed argumentative *composite* source:

Remark 3.3.3.1. $\mathcal{P} \rightarrow \mathcal{T}$ *since* $\mathcal{S} \rightarrow \mathcal{T}$ *and* $\mathcal{P} \Rightarrow \mathcal{S}$ *iff* $p \rightarrow t$, $s \rightarrow t$ *and* $p \Rightarrow s$

For example:
$\mathcal{P} = \{p_0\} = \{$"human greenhouse ga emission"$\}, \mathcal{S} = \{s_0, s_1, \ldots\} = \{$ "greenhouse ga emission", "release greenhouse gas", $\ldots\}, \mathcal{T} = \{t_0\} = \{$ "air pollution" $\}$. Because $p_0 \Rightarrow s_0$ and $s_0 \nrightarrow t_0$, it can be implied that $p_0 \nrightarrow t_0$ following Remark 3.3.2.1. Due to semantic equivalence, $\mathcal{S} \nrightarrow \mathcal{T}$ and $\mathcal{P} \Rightarrow \mathcal{S}$ are valid. Hence, $\mathcal{P} \nrightarrow \mathcal{T}$.

(b) Implied relation due to entailed argumentative *composite* target:

Remark 3.3.3.2. *$\mathcal{S} \rightarrow \mathcal{P}$ since $\mathcal{S} \rightarrow \mathcal{T}$ and $\mathcal{P} \Rightarrow \mathcal{T}$ iff $s \rightarrow p$, $s \rightarrow t$ and $p \Rightarrow t$*

For example:
$\mathcal{P} = \{p_0\} = \{$"develop the party platform"$\}, \mathcal{S} = \{s_0\} = \{$ "flag controversial issue for the public" $\}$, $\mathcal{T} = \{t_0, t_1, \ldots\} = \{$ "party", "political party", $\ldots\}$. Because $p_0 \Rightarrow t_0$ and $s_0 \rightarrow t_0$, it can be implied that $s_0 \rightarrow p_0$ following Remark 3.3.2.2. Due to semantic equivalence, $\mathcal{S} \rightarrow \mathcal{T}$ and $\mathcal{P} \Rightarrow \mathcal{T}$ are valid. Hence, $\mathcal{S} \rightarrow \mathcal{P}$.

Chapter 4

Evaluation

4.1 Pairwise Entity Linking

As explained in Subsection 3.1.2, pairing a concept instance with each of its neighbor aims to provide the EL tool TagMe with more context, improve the results of EL tools and therefore, produce more relevant mentions and groundings. From this step, the *Agreed EL* results across the pairings are obtained for each concept instances. To evaluate this approach, 100 unique concept instances (Appendix A) are randomly selected whose text contains combinations of the following characteristics:

1. Varying linguistic complexity (short words, phrases)
2. Acronyms or terminologies (e.g *"chao, presence of such ad"*)
3. Plural, singular, negation forms (e.g. *"natives", "non-native"*)
4. General concepts (e.g. *"land","society"*)
5. Numerals (e.g. *"700b plan", "web 2 0 democratizing and decentralizing effect"*)
6. Typos (e.g *"indigenou people", "mar mission"*)

Furthermore, the samples' *Single EL* results and *Agreed EL* results must be somewhat different for the purpose of evaluation (Appendix B). From these 100 concepts samples, 193 mentions are obtained. Out of these 193 mentions, 114 mentions M are grounded differently by using Algorithm 3.1:*Single EL* and Algorithm 3.2:*Agreed EL* ($G_{SEL} \neq G_{AEL}$). For instance, a concept text *"national renewable energy standard"* has different G_{SEL} (Table 4.1) and G_{AEL} (Table 4.2).

From these 114 samples M, the occurance of the correct grounding detected from agreed results G_{AEL} are counted. The grounding G_{AEL} is considered correct if it makes sense and relevant to the entire topic of the concept text and the surrounding concepts. For example, the G_{AEL} (Nation[1]) of a mention *"national"* (Table 4.2) is considered correct in comparison to its G_{SEL} (The National (Abu Dhabi)[2]) counterpart (Table 4.1).

Generally speaking, this approach results in entities with higher confidence scores which helps to disambiguate and gives the more context specific entities. Around 52.6% (60 out of 114 M), *Agreed EL* leads to correction.

mention	grounding G_{SEL}	score
"national"	The National (Abu Dhabi)	0.061
"renewable energy"	Renewable energy	0.392
"renewable energy standard"	Renewable portfolio standard	0.231

Table 4.1: G_{SEL} for concept instance "national renewable energy standard"

mention	grounding G_{AEL}	score
"national"	Nation	0.115
"renewable energy"	Renewable energy	0.420
"renewable energy standard"	Renewable portfolio standard	0.211

Table 4.2: G_{AEL} for concept instance "national renewable energy standard"

However, this approach does not always yield the most desirable results. Around 12.2% (14 out of 114 M), G_{AEL} are less accurate than G_{SEL}. This is because the true meaning of the entity within the confine of the concept can be diluted further by the concept surrounding it. For example, a mention *"non native"* is detected from a concept text *"non native dropout rate"* which is grounded by using:

- *Single EL*: *Introduced species*[3] with confidence score $= 0.083$
- *Agreed EL*: *Invasive species*[4] with confidence score $= 0.062$

In addition, mentions detected from the pairings may be different from ones detected from the Single EL due to the concatenation of the texts. For example, a mentions from a concept text *"language"* by using:

- *Single EL*: mention *"language"* is grounded to *Language*[5] with confidence score $= 0.009$

[1] `https://en.wikipedia.org/wiki/Nation`
[2] `https://en.wikipedia.org/wiki/The_National_(Abu_Dhabi)`
[3] `https://en.wikipedia.org/wiki/Introduced_species`
[4] `https://en.wikipedia.org/wiki/Invasive_species`
[5] `https://en.wikipedia.org/wiki/Language`

- *Agreed EL*: mention *"language"* becomes *"language translation"* and is grounded to *Translation*[6] with confidence score $= 0.335$,

The concept of *"language"* experience dilution due to pairings to *"translation"* which leads to the change of mention from *"language"* to *"language translation"*. In this case, The results *Agreed EL* does not interpret the concept *"language"* correctly, although they are has significantly higher score than the one of *Single EL*.

In the rare cases, the dilution is sometimes worse as the original mention from *Single EL* is not even detected. For example, the mention *"vote"* from *"accurateÂ vote"* is not recognized with *Agreed EL*. On the positive note, such dilution may be beneficial to the concept clustering since some words should not be recognized as an entity.

Despite the shortcomings, the overall results of *Agreed EL* are satisfactory and more context accurate based on the graph. This is not of huge implication because the entities are used as a part of various concept representation for clustering and improved approximation of concepts grouping instead of being involved directly as input in similarity detection. Hence, mentions and groundings from the agreed entities are used in favour of ones from the single entities for the next steps.

4.2 K-Means Clustering

As explained in Section 3.2, the concepts instances are incrementally clustered from k = 5 to 1500 (both extremes are included) with 5 units intervals. This amounts to 300 clustering results for each configuration. As ground truth (which concepts a cluster is supposed to contain) is also unavailable, the quality of each clustering needs to be measured intrinsically for each k clustering results. The general convention in this case seem to suggest that the use of silhouette score is suitable one. Hence, silhouette score is measured for each clustering of the 12 configurations (Figure 4.1 for Doc2Vec and Figure 4.2 for BERT).

As seen in Figure 4.1 and Figure 4.2 the scores fluctuate across the number of clusters. As such, it is not enough to simply pick the highest scores. Therefore, I find three peak values of each curve to be considered for the best k - clustering candidates. These peak values are obtained using a python implementation from scipy[7]. The following are the peak [k-cluster, silhouette score] pairs:

- Doc2Vec:

  (*a*) $[185, 0.0607], [305, 0.0602], [445, 0.0592]$

[6] https://en.wikipedia.org/wiki/Translation

[7] https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html

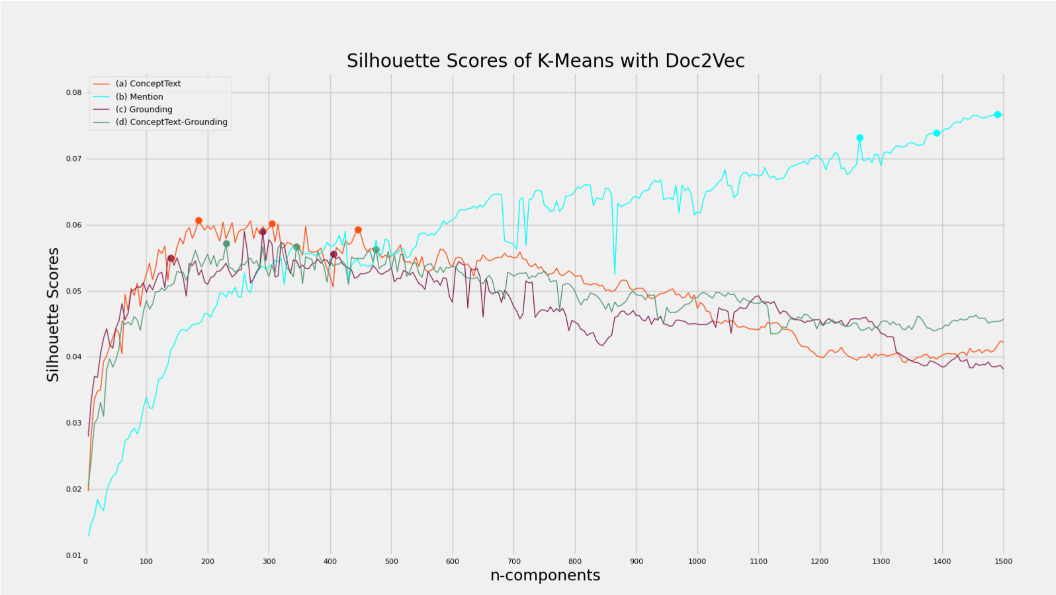


Figure 4.1: Silhouette scores for K-Means clustering using Doc2Vec from $k = 5$ to 1500

(*b*) $[1490, 0.0767]$, $[1390, 0.0738]$, $[1265, 0.0732]$

(*c*) $[290, 0.0590]$, $[405, 0.0556]$, $[140, 0.0550]$

(*d*) $[230, 0.0571]$, $[345, 0.0567]$, $[475, 0.0563]$

- BERT:

  (*a*) $[365, 0.0461]$, $[720, 0.0451]$, $[255, 0.0448]$

  (*b*) $[295, 0.0498]$, $[545, 0.0489]$, $[395, 0.0480]$

  (*c*) $[220, 0.0464]$, $[375, 0.0460]$, $[495, 0.0423]$

  (*d*) $[310, 0.0558]$, $[470, 0.0555]$, $[580, 0.0548]$

for *(a)* original concept text, *(b)* concatenation of its mentions, *(c)* concatenation of its grounding and *(d)* concatenation of its concept text and groundings.

From these candidates, the highest quality clustering results of peak scores for every configuration are inspected manually. For this, the 100 unique concept instances (Appendix A) are again used this time as anchor concept instances. These anchors are used to evaluate the peak $k-$cluster across the 8 configuration in regard to their cluster membership. A good $k-$cluster is a cluster whose concept membership makes the most sense in terms of semantic relatedness. The evaluation starts with the lowest k (185) to the highest (1490) for every peak [k-cluster, silhouette score] pairs and configurations.

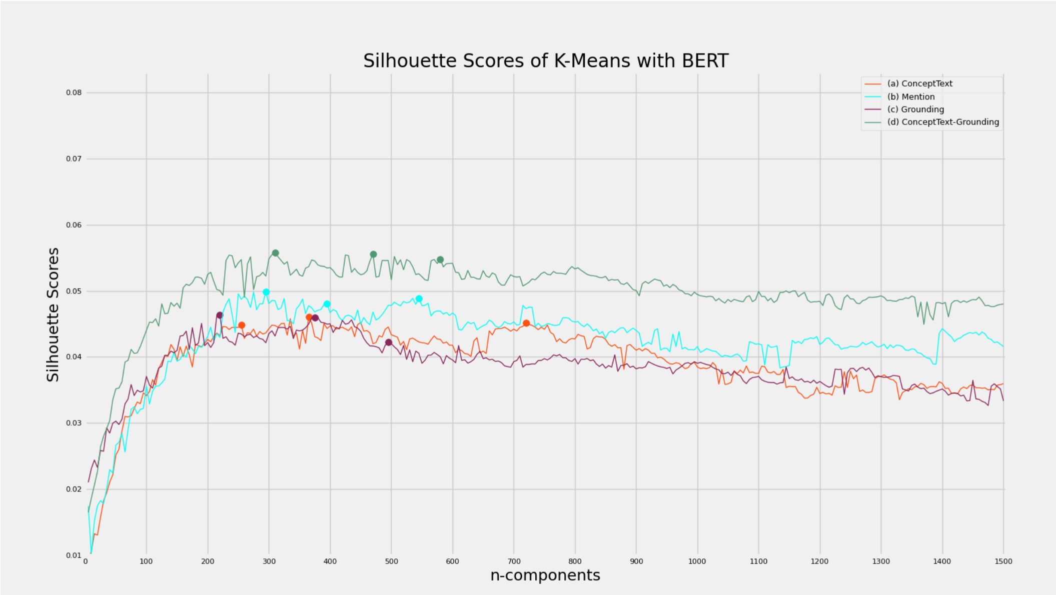


Figure 4.2: Silhouette scores for K-Means clustering using BERT from $k = 5$ to 1500

Beginning with $k = 185$, we can already omit the usage of *(b) mention* and *(c) grounding* input for feature space and consequently, the next processing steps. The clusters produced by this configuration using either of the models are for the most part badly grouped and its membership does not make sense. In other words, the concepts in most clusters are barely related and some are not of the similar topic at all. Even at the highest sillouhette score, clusters using *(b)* (Table 4.3) or *(c)* seem to have an incoherent concept membership. This might be due to the fact that solely using *(b)* and *(c)* strips additional information from the already lacking concept text. In addition, some concept texts do not have *mention* at all.

185 (Doc2Vec)	185 (BERT)	365 (BERT)
700b plan	700b plan	700b plan
future generation	make change	make water cheaper
progress	grow better	water
confidence spending	antus establishment	end
build confidence	spending limit	side effect
...	...	...

Table 4.3: Cluster of *(b)* with *"700b plan"* as anchor. (There is no other common member among the three clustering configuration results beyond the one(s) listed here)

Using mention in conjuction with other concept representation seem to be pointless as well since it will only repeat the words or entity that is already contained in concept text. Hence, it is more meaningful to base the clustering on *concept texts (a)* and their combination with *grounding (d)* and omit the usage of *mention* and *grounding* altogether. Combining *concept text* and *grounding* as the concept representation *(d)* may lead to a better cluster. This is especially true if *groundings* are correct for a given concept text. Interestingly, such effectiveness due to this combination is only observed when the Doc2Vec model is used (Table 4.4). For clusters with BERT, *concept text* as concept representation yield best cluster membership overall (Table 4.5).

(a) concept text	*(d) concept text-grounding*
"700b plan"	"700b plan"
"picken plan"	"policymaker"
"progress"	"picken plan"
"picken plan wind turbine"	"domestic company"
"strategic planning"	"monetary policy"
"planning ahead"	"fiscal policy"
"planning"	"debate on holocaust"
"partition plan"	"global policy"
"failed start up"	"domestic producer"
"step"	"policy"
	"immigration control policy"
	"english only policy"
	...

Table 4.4: *(a)* and *(d)* based cluster using Doc2Vec and "700b plan" as anchor. (There is no other common member between (a) and (d) beyond the ones) listed here)

(a) concept text	*(d) concept text-grounding*
"700b plan"	"700b plan"
"700 mile fence"	"domestic company"
"700b bailout"	"fiscal policy"
	"picken plan"
	...

Table 4.5: *(a)* and *(d)* based cluster using BERT and "700b plan" as anchor. (There is no other common member between (a) and (d) than "700b plan")

In general, BERT model produce a better cluster than Doc2Vec given the same k. The members of clusters produced by K-Means with BERT model do not only contain more similar words but also have more relevant topic. Hence, it is easier to infer a meaningful topic from a cluster produced by BERT than Doc2Vec. For the next step, the $k = 365$ cluster of *concept text* using BERT is used.

4.3 Textual Entailment

AllenNLP is used to determine the TE relation between the concept instances for every cluster with the initial entailment factor $\varepsilon_0 = 1.1$ as explained in Section 3.3. Due to its relatively good membership as evaluated in Section 4.2, the chosen clusters are one resulted from the following configuration:

- $k = 365$ clustering technique
- BERT as embedding model
- *concept text* as concept representation

The evaluation is carried out with the help of two surveys targeted to human experts. First survey aims to give an initial guide which of the resulting concept mergings and implications due to entailment with ε_0 make sense and which characteristic they have. The second is conducted with different value of ε or a threshold of $\mathbb{E}$.

To reiterate, the surveys are used to assess the following cases:

(1) Bidirectional textual entailment (BTE): *are the group of concept instances which entail one another semantically equal?* (Subsection 3.3.1)

(2) Unidirectional textual entailment (UTE): *does the premise imply a new relation, given an existing argumentative relation?* (Subsection 3.3.2). This has two subcases:

   (a) Implied relation due to entailed source
   (b) Implied relation due to entailed target

(3) Combination of UTE and BTE: *does implied relation due to UTE involving concept instances that are semantically equivalent make sense?*

4.3.1 Survey on concept linking with ε_0

The first survey contains 2 section: first section to evaluate case (2) and (3) and the second one to evaluate case (1). For the first section, 15 implied relation based on Remarks 3.3.2.1, 10 based on Remarks 3.3.3.1, 15 based on Remarks 3.3.2.2 and 10 based on Remarks 3.3.3.2 are randomly chosen. The respondents are asked if the implied relation makes sense given the existing argumentative relation without knowing the entailment. The questions are formulated like so:

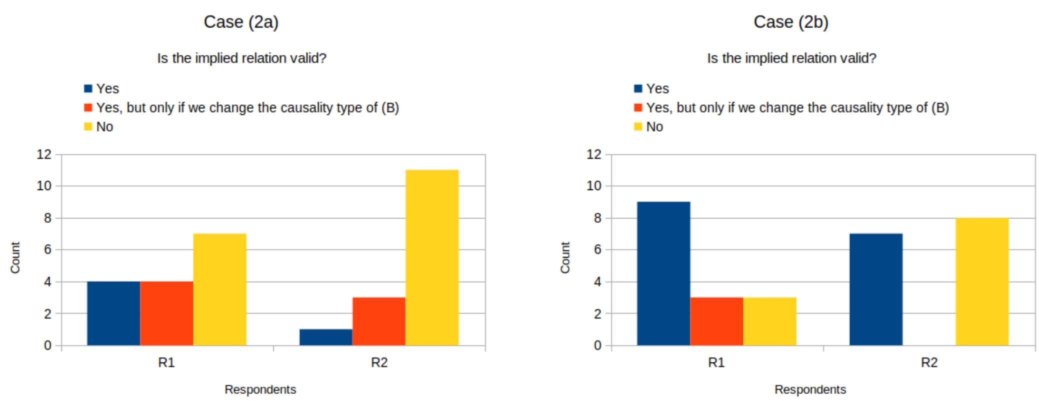
> *(A) "hassle of regulating conflict of interest" affects negatively "government". (B) "hassle of regulating conflict of interest" affects negatively "federal government". Does (A) implies (B)?*

The second section ask the respondents to try to regroup (if necessary) the concept instances that are considered as semantically equivalents.

Overall, the first survey responses indicate that $\varepsilon_0 = 1.1$ is too low. The answers from the respodents differ greatly in regard to many of the merging and linking candidates. For starter, some concept instances that are regarded as a semantic equivalent do not actually have the same meaning according to the survey respondents (Table 4.6) and thus, should be grouped into different *composite concepts* if possible. On many cases, repondents have difficulty to decide and agree which concept instance should belong into the same group.

Merging candidates	Survey responses
"technology" ⇔ "creator" ⇔ "innovation" ⇔ "creationism" ⇔ "technology advance" ⇔ "modernisation" ⇔ "evolution"	"technology" ⇔ "technology advance"
	"creator" ⇔ "creationism"
	"modernisation" ⇔ "innovation"
	"evolution"
"weapon" ⇔ "arm" ⇔ "army"	"weapon"
	"arm"
	"army"
cut taxe ⇔ increase taxe ⇔ higher taxe ⇔ tax rate reduction ⇔ clas warfare to pas tax burden	increase taxe ⇔ higher taxe
	cut taxe ⇔ tax rate reduction
	clas warfare to pas tax burden

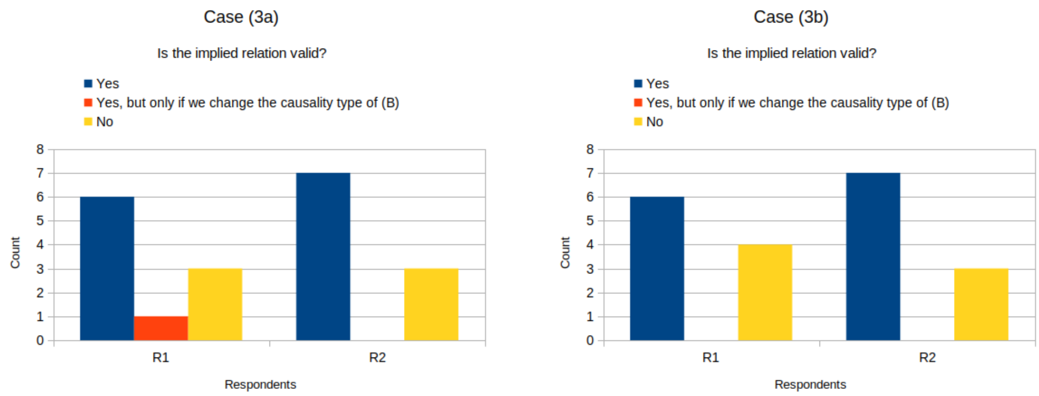
Table 4.6: BTE using ε_0 yields merging candidates which actually are not semantically similar and thus, should be split according to the respondents



(a) Entailed to the source **(b)** Entailed to the target

Figure 4.3: The count of each answer in the 1st survey for case (2) for each respondents

The respondents' answers on UTE cases also give the same indication. Respondents have difficulty to make sense the implied relation on many occasions with only the average 16.7% implied relations are valid for case (2a). Furthermore, 50% implied relations are valid for case (2b).



(a) Entailed to the source

(b) Entailed to the target

Figure 4.4: The count of each answer in the 1st survey for case (3) for each respondents

It seems that implied relations for the case (3) seems to fare well with 65% for case (3a) and case (3b). However, upon close inspection, repondents do not seem to agree wheter the implied relation makes sense on many occasions including those for case (2). The breakdown of each response is attached on Appendix C.

On the positive light, the respondents are in total agreement with the merging and linking candidates which are associated with a near 100% $\mathbb{E}$ including the sample cases as seen in Table 3.8 and Table 3.9. This indicates that a higher value of ε or $\mathbb{E}$ threshold needs to be applied before any conclusion is to be made.

4.3.2 Survey on concept linking with $\mathbb{E} > 0.9$

Based on the first survey, the TE approach (Section 3.3) is repeated using differentε and consequently $\mathbb{E}$ threshold. Upon further inspection, the lowest $\mathbb{E}$ that matches the grouping for case (1) and results in repondents' agreement for case (2) from the survey turns out to be 0.9. With it, TE approach for concept linking are performed and new set of *composites* and implied relations are generated. Besides *composites concepts* and implied relation which involve $\mathbb{E} > 90\%$, almost all sample *composites* and implied relation that are previouly used in survey do not exist in the new set.

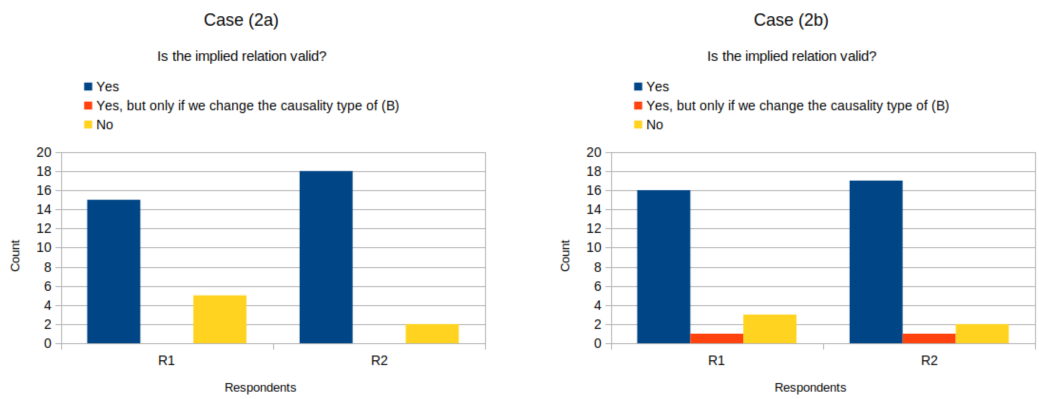
As expected, the *composite concepts* are more granular than those with $\varepsilon = 1.1$. Their components are actually paraphrases and in many cases are slight typographical errors and lexical variations (Table 4.7).

Seeing the encouraging results of $\mathbb{E}$ threshold, the second survey is conducted. This time, the survey aims only to assess the implied relation due to UTE (both case (2) and (3)). With the total of 70 questions, 20 implied relation for case (2a),

Merging candidates with $\varepsilon = 1.1$	Merging candidates with $\mathbb{E} > 0.9$
"kid" ⇔ "young person" ⇔ "nclb child" ⇔ "one child" ⇔ "youth" ⇔ "child"	"kid" ⇔ "child"; "young person" ⇔ "youth"; "one child"; "nclb child"
"health industry" ⇔ "healthcare worker" ⇔ "health consciou" ⇔ "healthcare system" ⇔ "health care supplier" ⇔ "healthcare industry" ⇔ "health care cost" ⇔ "health care subscriber" ⇔ "health crisi" ⇔ "health risk" ⇔ "health information" ⇔ "spread health" ⇔ "health care" ⇔ "healthcare" ⇔ "health care system"	"health industry"; "healthcare worker"; "health consciou"; "healthcare system" ⇔ "health care system"; "health care supplier"; "healthcare industry"; "health care cost"; "health care subscriber"; "health crisi"; "health risk"; "health information"; "spread health"; "health care" ⇔ "healthcare"

Table 4.7: Differences of ones considered semantic equivalences between BTE with the initial $\varepsilon = 1.1$ and the $\mathbb{E}$ threshold > 0.9.

15 for (3a), another 20 for case (2b) and 15 for (3b) are randomly chosen. The same as the first section of the first survey, The respondents are asked the same way as the section 1 in the 1st survey.



(a) Entailed to the source

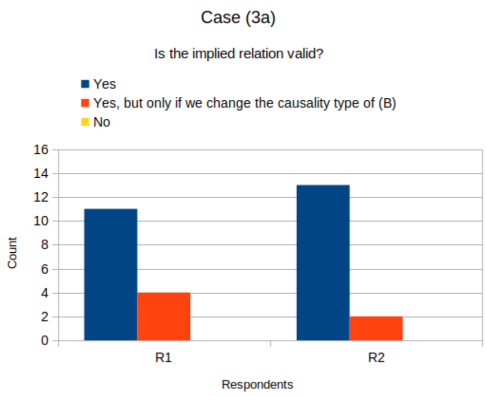
(b) Entailed to the target

Figure 4.5: The count of each answer in the 2nd survey for case (2) for each respondents

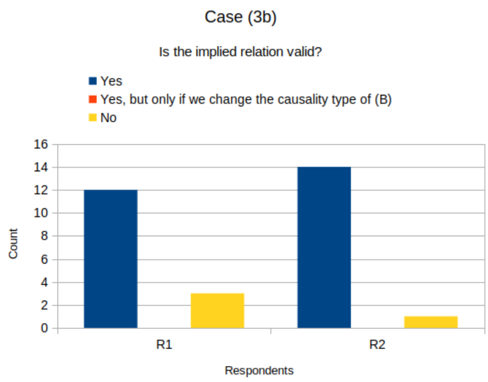
The assessment result is very positive for both cases (case (2) and (3) with their respective subcases) with favourable percentage that agrees on the implied relation. On average, 82.5% implied relations for both case (2a) and (2b) are valid according to the respondents (Figure 4.5).

For the case (3a), the average 80% of implied relations are considered true on the average while 20% of them need the opposite implied relation (implied relation should be the opposite of the existing argumentative relation) as seen in Figure 4.6.

Similarly, 86.6% implied relations are valid for case (3b) with 26% implied relations are invalid. The breakdown of this survey can be seen in Appendix D.



(a) Entailed to the source



(b) Entailed to the target

Figure 4.6: The count of each answer in the 2nd survey for case (3) for each respondents

The surveys show that with appropriate $\mathbb{E}$ threshold, TE is helpful to link concept instances for various cases. However, TE may not be sufficient for certain arguments and their corresponding concept instances as it lacks the argumentative bias that the AKG represents as a whole. Consider the following cases in regard to the concepts *"penalty"*, *"fine and penalty"* and *"death penalty"*:

- Because *"penalty"* $\Rightarrow$ *"fine and penalty"*, *"penalty"* $\rightarrow$ *"company"* implies "fine and penalty" $\rightarrow$ *"company"*. This follows the Remark 3.3.2.1 and is considered true by the respondents.

- In contrast, respondents cannot decide whether *"penalty"* $\rightarrow$ *"society"* implies *"death penalty"* $\rightarrow$ *"society"* eventhough *"penalty"* $\Rightarrow$ *"death penalty"* which, same as the previous case, follows the Remark 3.3.2.1.

Overall, TE can be applied to reveal new implied relation related to argumentative relation as well as to merge semantic equivalents given that the E threshold is high enough. On the rare occasion, some linking candidates still however need a close assessment.

Chapter 5

Conclusion

This thesis demonstrates the usefulness of textual entailment in linking argumentative concept instances in conjunction with entity linking and K - Means clustering. With various concept representations obtained from entity linking and configurations of K - Means, the most appropriate k number cluster of semantically related concept instances can be found. With this approach, textual entailment can be applied with less computation time while still retaining a high chance of success as opposed to the application of textual entailment directly to the complete set of concept instances.

By taking advantage of bidirectional and unidirectional textual entailment between two concept instances, concept linking is carried out to merge semantically equivalent concept instances and to create new implied argumentative relations using the existing argumentative relations. With a near 100% entailment score as a threshold, the thesis at hand shows that concept linking based on TE can yield very good results.

Some cases however need to be handled separately as TE cannot take the relation or the argument bias represented by the AKG into account. Further research on the nature of argumentation and its relation with TE is required to handle such cases.

On the aspect of concept representation, there is a possible method to get a more accurate representation of a concept by exploiting the Wikipedia category. This is however largely dependent on the performance of TagMe or Wikipedia-based EL tool (wikifier). In rare cases during the writing of this, it has been observed that TagMe cannot give an accurate grounding. Although a certain threshold can be applied to ensure that only mention and grounding with high confidence score are accepted, this may not work well. This is especially true when the mentions recorded in Wikipedia only point to one single Wikipedia article or entity. In this case, TagMe will always give inaccurate grounding regardless of how much context is given to the mention as there is no other possible grounding. Therefore, exploiting the category in this state will be pointless in this state and that is why it is omitted from this thesis. To make this approach work, a better wikifier or more sophisticated EL tool needs to be used. Furthermore, the knowledge base of such EL has to accommodate the categorization of their entities.

Another future improvement that can be applied to benefit entity linking, clustering, and textual entailment is typographical error elimination. Before the further process, such typographical error needs to be recognized and corrected. This can be done by finding some candidates with the shortest Levenshtein distance (or other similarity metrics). Moreover, the adjacent argumentative concepts can also be used to choose the most probable correction among the aforementioned candidates.

Appendix A

Concept Samples

No.	Concept
1	700b plan
2	accurate vote
3	achievement of killing osama
4	agitate usa west relation
5	become a forced norm
6	beneficial mind altering effect
7	biofuel production
8	border fence
9	brightest
10	burden immigrant
11	cap and trade system
12	carbon trading
13	chance for a stable government would significantly decrease after usa withdrawal from iraq
14	chao
15	cheaper travel and greater acces
16	child s mental health
17	circumventing certain more ordinary legal
18	collapsing of the skull of the partially born fetu
19	division extremist want
20	donation center to use footage in commercial
21	drug use can be beneficial to user
22	educate public about gun
23	ego to unhealthy level
24	election

25	embryonic stem cell research greater in potential
26	fulfilling life
27	full dollarization
28	future conflict
29	gameplay among youth ha increased
30	giving indium nuclear aid
31	gm crop mix with native plant
32	gun fatality
33	help other state be more stable
34	hunger
35	ill advised global
36	illegal
37	important political alternative
38	improve poor person life style
39	incentive for illegal immigrant to remain inside a country
40	increase chance of mar mission
41	increasing migrant right
42	independent scotland
43	indigenou people
44	indium with relatively little energy
45	indulging in viewing perfect man woman
46	inflation pressure
47	innocent person should not be persecuted
48	insecurity
49	institutionalized in destructive way
50	insufficient broadband market choice
51	interaction with other culture
52	iraqi leader
53	iraqi troop to defect to the insurgency
54	israeli woman
55	japan
56	job los
57	judicial system
58	keeping company honest
59	keystone
60	kosovo independence
61	land
62	large polygamou family
63	latino

64	leaking lubricating oil from wind power
65	lessening the strength of hurricane
66	lifting gaza
67	language
68	man
69	mandating military service
70	mar mission
71	market acces
72	mature rapidly
73	medical advancement
74	member of the team
75	method than circumcision
76	microfinance
77	national renewable energy standard
78	national hysterium
79	need to protect against opposing the group
80	new technique
81	non native dropout rate
82	obama african decent
83	ozone layer damaging
84	presence of such ad
85	public welfare
86	purpose
87	resident
88	retiree
89	roma person
90	stunt
91	stabilizing the economy
92	teen
93	terrorist cause
94	think it okay
95	time
96	victim
97	void in people heart
98	wa
99	worsen antus gun opinion
100	web 2 0 democratizing and decentralizing effect

Appendix B

Single and Agreed EL results

Single (SEL) and Agreed Entity Linking (AEL) Samples

mention with different SEL and AEL groundings
mention which is transformed
less accurate AEL grounding
more accurate AEL grounding

		SEL		AEL	
concept	**mention**	**grounding**	**silhouette score**	**grounding**	**silhouette score**
700b plan	plan	Plan	0.0018	Economic policy	0.2555
accurate vote	accurate	ACCURATE	0.2570	Accuracy and pre-cision	0.0007
accurate vote	vote	Voting	0.2580	n/a	n/a
achievement of killing osama	achievement	Goal	0.0015	Goal	0.0492
achievement of killing osama	osama	Osama (film)	0.0100	Osama bin Laden	0.0640
achievement of killing osama	killing	Death of Osama bin Laden	0.0037	Murder	0.0814
agitate usa west relation	agitate	Agitator (device)	0.0015	Agitator (device)	0.0015
agitate usa west relation	usa	United States	0.1326	United States	0.1249
agitate usa west relation	west	West Ger-many	0.0368	Western world	0.1085
agitate usa west relation	relation	Charles Sanders Peirce	0.0039	Property (philoso-phy)	0.1063
become a forced norm	norm	Norm (mathe-matics)	0.0297	Norm (so-cial)	0.2425
beneficial mind altering effect	beneficial	Benefi-cial in-sects	0.0005	Probiotic	0.0948
beneficial mind altering effect	mind	Mind	0.0057	Mind	0.0806
biofuel production	biofuel	Biofuel	0.5977	Biofuel	0.5294
biofuel production	production	Produc-tion (e-conomics)	0.2036	n/a	n/a
border fence	border fence	Border barrier	0.0346	Mexico–United States barrier	0.1374
brightest	brightest*	n/a	n/a	The Best and the Brightest	0.2125
burden immigrant	burden	Tax inci-dence	0.1906	Legal burden of proof	0.1424
burden immigrant	immigrant	Immigra-tion	0.2246	Immigra-tion	0.1701
cap and trade system	cap and trade	Emissions trading	0.5035	Emissions trading	0.5669
cap and trade system	system	System	0.2067	System	0.1968

carbon trading	carbon trad ing	Carbon emission trading	0.1842	Emissions trading	0.3787
chance for a stable gov- ernment would signifi- cantly decrease after usa withdrawal from iraq	chance	Indeter- minism	0.0619	Indeter- minism	0.0722
chance for a stable gov- ernment would signifi- cantly decrease after usa withdrawal from iraq	stable	Stable	0.0230	Sorting algorithm	0.0312
chance for a stable gov- ernment would signifi- cantly decrease after usa withdrawal from iraq	government	Govern- ment	0.1041	Govern- ment	0.1580
chance for a stable gov- ernment would signifi- cantly decrease after usa withdrawal from iraq	usa	United States	0.1470	United States	0.1485
chance for a stable gov- ernment would signifi- cantly decrease after usa withdrawal from iraq	iraq	Iraq	0.3698	Iraq	0.4254
chao	chao	Chao (Sonic)	0.0328	Discor- dianism	0.1808
cheaper travel and greate r acces	travel	Travel	0.0031	Tourism	0.3160
child s mental health	child	Child	0.2426	Child abuse	0.3308
child s mental health	mental heal th	Mental health	0.2985	Mental health	0.3356
circumventing certain mor e ordinary legal	ordinary	Ordinary (officer)	0.1831	Ordinary (officer)	0.1124
circumventing certain mor e ordinary legal	legal	Legal personal- ity	0.1738	Law	0.2157
collapsing of the skull o f the partially born fetu	collapsing	Collapse of the World Trade Center	0.0010	Collapse of the World Trade Center	0.0010
collapsing of the skull o f the partially born fetu	skull	Skull	0.0529	Human skull	0.1402
collapsing of the skull o f the partially born fetu	fetu	Fetu	0.0429	Fetu	0.0429
division extremist want	division	Division (mili- tary)	0.0062	A Divi- sion (New York City Subway)	0.0969
division extremist want	want	Want	0.0007	Want	0.0007
division extremist want	extremist	Extremism	0.0322	Islamic extremism	0.0322
donation center to use footage in commercial	donation	Donation	0.0738	Donation	0.0645
donation center to use footage in commercial	center	Centrism	0.0283	Centrism	0.0283

donation center to use footage in commercial	footage	Footage	0.0525	Footage	0.0791
donation center to use footage in commercial	commercial	Advertis-ing	0.1391	Televi-sion ad-vertise-ment	0.1451
drug use can be benefi-cial to user	drug use	Recre-ational drug use	0.2406	Recre-ational drug use	0.2348
drug use can be benefi-cial to user	can	Beverage can	0.0649	Can (band)	0.1024
drug use can be benefi-cial to user	beneficial	Probiotic	0.0589	HSBC	0.1024
drug use can be benefi-cial to user	user	Drug user	0.0919	Drug user	0.1312
educate public about gun	educate	Education	0.1724	Education	0.0986
educate public about gun	public	Public univer-sity	0.1284	Public	0.1492
educate public about gun	gun	Gun	0.0562	Gun	0.1208
ego to unhealthy level	ego	Self-con-cept	0.1356	Id, ego and su-per-ego	0.1831
ego to unhealthy level	unhealthy	Health	0.1972	Health	0.1850
ego to unhealthy level	level	GCE Ad-vanced Level (U-nited Kingdom)	0.0925	GCE Ad-vanced Level (U-nited Kingdom)	0.1049
election	election	By-elec-tion	0.0212	Election	0.1846
embryonic stem cell re-search greater in poten-tial	potential	Action potential	0.2382	Quantum computing	0.2271
embryonic stem cell re-search greater in poten-tial	embryonic s tem cell	Embryonic stem cell	0.5999	Embryonic stem cell	0.5632
fulfilling life	life	Life (magazine)	0.0029	Life	0.1954
full dollarization	dollariza-tion	Currency substitu-tion	0.3608	Currency substitu-tion	0.6483
future conflict	future	Future	0.2544	Future	0.1534
future conflict	conflict	Conflict (narra-tive)	0.2563	War	0.2260
gameplay among youth ha increased	gameplay	Gameplay	0.1080	Gameplay	0.0172
gameplay among youth ha increased	ha	Hectare	0.0787	Hectare	0.0539
gameplay among youth ha increased	youth	Youth	0.1192	n/a	n/a
giving indium nuclear aid	indium	Indium	0.3454	Indium	0.2719
giving indium nuclear aid	nuclear	Nuclear weapon	0.2273	Nuclear weapon	0.1249

giving indium nuclear aid	aid	Artificial intelligence	0.1358	Humanitarian aid	0.0406
gm crop mix with native plant	gm crop	Genetically modified crops	0.3136	Genetically modified crops	0.4189
gm crop mix with native plant	mix	Audio mixing (recorded music)	0.0051	Mongrel	0.0803
gm crop mix with native plant	native plant	Native plant	0.1900	Native plant	0.1886
gun fatality	gun	Gun	0.1968	Firearm	0.1761
gun fatality	fatality	Death	0.2073	Death	0.1673
help other state be more stable	help	The Help	0.0567	Help! (song)	0.0402
help other state be more stable	state	Alabama	0.0595	State (polity)	0.1135
help other state be more stable	stable	Sorting algorithm	0.0106	Numerical stability	0.0106
hunger	hunger	Hunger	0.0330	Malnutrition	0.2140
ill advised global	ill	Illinois	0.0294	Disease	0.1070
ill advised global	global	Globalization	0.0319	Globalization	0.1074
illegal	illegal	Law	0.0031	Crime	0.2598
important political alternative	political	Political (song)	0.0037	Politics	0.2965
important political alternative	alternative	Alternative culture	0.0076	Alternative culture	0.1939
improve poor person life style	poor person	Poor person	0.0096	Poor person	0.0096
improve poor person life style	life style	Life (magazine)	0.0182	Lifestyle (sociology)	0.0912
incentive for illegal immigrant to remain inside a country	incentive	Incentive	0.0828	Incentive	0.0656
incentive for illegal immigrant to remain inside a country	illegal immigrant	Illegal immigration to the United States	0.2977	Illegal immigration	0.3555
incentive for illegal immigrant to remain inside a country	inside	Inside (Ronnie Milsap album)	0.0015	Inside (Ronnie Milsap album)	0.0015
incentive for illegal immigrant to remain inside a country	country	Nation state	0.1530	Nation state	0.1199
increase chance of mar mission	chance	Probability	0.0339	Luck	0.0909

increase chance of mar mission	mar	Gospel of Mark	0.1435	Morocco	0.0661
increase chance of mar mission	mission	Christian mission	0.1769	Christian mission	0.0702
increasing migrant right	migrant	Immigration	0.2538	Immigration	0.1318
increasing migrant right	right	Right-wing politics	0.2339	Relative direction	0.0015
independent scotland	independent scotland	Scottish independence	0.4440	Scottish independence	0.2443
independent scotland	scotland	Kingdom of Scotland	0.6062	Scotland	0.4128
indigenou people	people	People (magazine)	0.0026	Person	0.3873
indium with relatively little energy	indium	Indium	0.3075	Indium	0.3162
indium with relatively little energy	little	Little owl	0.0013	Little owl	0.0309
indium with relatively little energy	energy	United States Department of Energy	0.1260	Energy	0.1869
indulging in viewing perfect man woman	viewing	Viewing (funeral)	0.0007	Heathrow Airport	0.0007
indulging in viewing perfect man woman	perfect man	The Perfect Man	0.1161	Perfect Man (Shinhwa album)	0.1161
indulging in viewing perfect man woman	man woman	A Man and a Woman	0.0213	A Man and a Woman (song)	0.0816
inflation pressure	inflation	Inflation (cosmology)	0.2736	Inflation	0.3775
inflation pressure	pressure	Atmospheric pressure	0.1855	Atmospheric pressure	0.0463
innocent person should not be persecuted	innocent	Innocence	0.1397	n/a	n/a
innocent person should not be persecuted	person	Hypostasis (philosophy and religion)	0.1212	Person	0.1608
innocent person should not be persecuted	persecuted	Persecution	0.2461	Persecution	0.1535
insecurity	insecurity	InSecurity	0.0143	Emotional security	0.2136
institutionalized in destructive way	institutionalized	Institutionalized (song)	0.0109	Involuntary commitment	0.1154

institutionalized in de-structive way	way	By the Way	0.0007	Tao	0.1125
insufficient broadband market choice	broadband	Broadband	0.2334	Broadband	0.1547
insufficient broadband market choice	market	Market (eco-nomics)	0.2765	Market (eco-nomics)	0.1920
insufficient broadband market choice	choice	Utility	0.2561	n/a	n/a
interaction with other culture	interaction	Interac-tion	0.2286	Social relation	0.2301
interaction with other culture	culture	Culture	0.2343	Culture	0.2300
iraqi leader	iraqi	Iraq	0.2145	Iraq War	0.1253
iraqi leader	leader	Supreme leader	0.1982	Leader-ship	0.1927
iraqi troop to defect to the insurgency	iraqi	Iraqis	0.2299	Ba'athist Iraq	0.2032
iraqi troop to defect to the insurgency	troop	Troop	0.1624	Troop	0.1123
iraqi troop to defect to the insurgency	defect	Defection	0.1483	Defection	0.0594
iraqi troop to defect to the insurgency	insurgency	Insur-gency	0.2734	Iraqi in-surgency (2003–11)	0.1830
israeli woman	israeli	Israelis	0.1817	Israel	0.1045
israeli woman	woman	Woman	0.1626	n/a	n/a
japan	japan	Japan	0.2651	Empire of Japan	0.4365
job los	job	Job (bib-lical figure)	0.0045	Employ-ment	0.2194
judicial system	judicial system	Judiciary	0.0299	Judicial system of China	0.3195
keeping company honest	keeping company	Keeping Company	0.1119	Keeping Company	0.1119
keeping company honest	honest	Honest (Future album)	0.0044	Dishon-esty	0.0825
keystone	keystone	Keystone (archi-tecture)	0.0693	Hercules (constel-lation)	0.2886
kosovo independence	kosovo	Kosovo War	0.7526	Kosovo	0.7500
kosovo independence	kosovo in-dependence	2008 Kosovo declara-tion of indepen-dence	0.5014	2008 Kosovo declara-tion of indepen-dence	0.3814
land	land	Land (e-conomics)	0.0015	Land law	0.2738
large polygamou family	family	Family (biology)	0.0291	Family	0.1376

latino	latino	Race and ethnicity in the United States Census	0.5000	Latino (demonym)	0.5945
leaking lubricating oil from wind power	leaking	Internet leak	0.0042	Internet leak	0.0042
leaking lubricating oil from wind power	lubricating oil	Lubricant	0.2169	Motor oil	0.1778
leaking lubricating oil from wind power	wind power	Wind power	0.3150	Wind power	0.3537
lessening the strength of hurricane	strength	Ultimate tensile strength	0.1306	Virtue	0.1321
lessening the strength of hurricane	hurricane	Tropical cyclone	0.1944	Tropical cyclone	0.1733
lifting gaza	lifting	DDT (professional wrestling)	0.0020	Momentum	0.0887
lifting gaza	gaza blockade	Blockade of the Gaza Strip	0.3500	Blockade of the Gaza Strip	0.4381
language	language*	Language	0.0091	Translation	0.3346
man	man	MAN SE	0.0032	Human	0.1617
mandating military service	military service	Military service	0.0205	Conscription	0.1704
mar mission	mar	Gospel of Mark	0.2775	Morocco	0.0685
mar mission	mission	Christian mission	0.2796	Christian mission	0.0685
market acces	market	Market (place)	0.0081	Market (economics)	0.2951
mature rapidly	mature	Sexual maturity	0.0013	Adult	0.2645
medical advancement	medical	Medicine	0.0047	Health care	0.2148
member of the team	member	Network affiliate	0.1284	Board of directors	0.0925
member of the team	team	The A-Team	0.1287	Team	0.2338
method than circumcision	method	Socratic method	0.1955	Scientific method	0.0922
method than circumcision	circumcision	Circumcision	0.4379	Circumcision	0.4839
microfinance	microfinance	Microfinance	0.2469	Microcredit	0.4834
national renewable energy standard	national	The National (Abu Dhabi)	0.0606	Nation	0.0716

national renewable energy standard	renewable energy	Renewable energy	0.3900	Renewable energy	0.4500
national renewable energy standard	renewable energy standard	Renewable portfolio standard	0.2316	Renewable portfolio standard	0.2917
national hysterium	national	The National (band)	0.0014	Nationalism	0.1824
need to protect against opposing the group	group	Group (mathematics)	0.0043	Social group	0.1563
new technique	technique	The Technique	0.0018	Scientific technique	0.2563
non native dropout rate	non native	Introduced species	0.0834	Invasive species	0.0703
non native dropout rate	dropout	Dropout (communications)	0.0147	High school dropouts	0.0879
non native dropout rate	rate	Heart rate	0.0780	Information theory	0.0900
obama african decent	obama	Barack Obama	0.3019	Barack Obama	0.1630
obama african decent	african	African Americans	0.2888	Black people	0.1980
ozone layer damaging	ozone layer	Ozone layer	0.2552	Ozone depletion	0.4124
presence of such ad	presence	Divine presence	0.2000	Divine presence	0.0700
presence of such ad	ad	Common Era	0.2100	Advertising	0.1100
public welfare	public welfare	Welfare	0.0300	Welfare economics	0.2500
purpose	purpose	Purpose (Justin Bieber album)	0.0006	Intention	0.1736
resident	resident	Resident (title)	0.0050	Residency (domicile)	0.1472
retiree	retiree	Pensioner	0.0282	Retirement	0.2199
roma person	roma	Romani language	0.3162	Romani people	0.2988
roma person	person	Person	0.1762	Person	0.1969
stunt	stunt	Stunt	0.2300	Stunt	0.0800
stunt	development	Filmmaking	0.2100	Economic development	0.1600
stabilizing the economy	stabilizing	Lyapunov stability	0.2217	Lyapunov stability	0.0010
stabilizing the economy	economy	Economic system	0.2316	n/a	n/a
teen	teen	Teen film	0.0074	Adolescence	0.2381

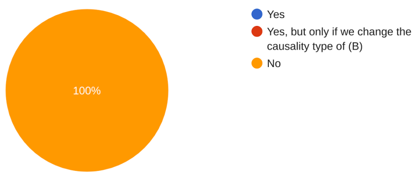
terrorist cause	terrorist	Terrorism	0.2187	Terrorism	0.2146
terrorist cause	cause	Causation (law)	0.1762	Social issue	0.1915
think it okay	think	Think (Aretha Franklin song)	0.0020	Thought	0.1137
think it okay	okay	OK	0.0111	OK	0.0111
time	time	Time (magazine)	0.0051	Time	0.2031
victim	victim	Victim (1961 film)	0.0038	Victimol-ogy	0.1597
void in people heart	void	Vacuum	0.0232	Void mar-riage	0.0772
void in people heart	people	People (magazine)	0.0864	People!	0.1473
void in people heart	heart	Heart (band)	0.0957	Heart (band)	0.1275
wa	wa	Western Australia	0.0225	Washing-ton (s-tate)	0.0581
worsen antus gun opinion	gun	Artillery	0.0624	Gun	0.1339
worsen antus gun opinion	opinion	Freedom of speech	0.0560	Freedom of speech	0.1719
web 2 0 democratizing and decentralizing effect	web	World Wide Web	0.1677	Internet	0.1844
web 2 0 democratizing and decentralizing effect	web 2	Web 2.0	0.1731	Web 2.0	0.1660
web 2 0 democratizing and decentralizing effect	democratiz-ing	Democra-tization	0.1256	Democra-tization	0.1244
web 2 0 democratizing and decentralizing effect	decentral-izing	Decen-traliza-tion	0.1753	Decen-traliza-tion	0.1979

Appendix C

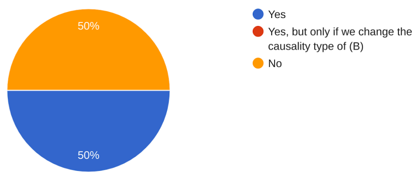
First UTE Survey

C.1 Case (2a)

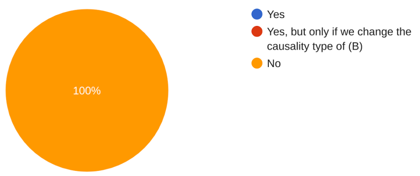
(A) "sex offender" affects negatively "right to reproduce". (B) "castrating sex offender" affects negatively "right to reproduce". Does (A) implies (B)?
2 responses



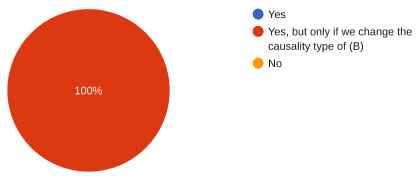
(A) "false hope in iraq" affects positively "limited succes of surge". (B) "limited drilling project" affects positively "limited succes of surge". Does (A) implies (B)?
2 responses



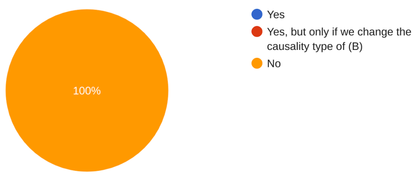
(A) "individual right to bear arm" affects positively "domestic tyranny". (B) "right to deny treatment" affects positiv...domestic tyranny". Does (A) implies (B)?
2 responses



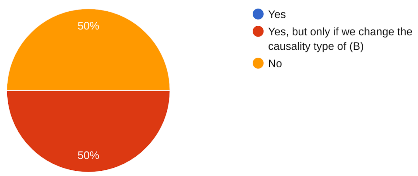
(A) "estate tax banned" affects positively "charitable giving". (B) "repealing estate tax" affects positively "charitable giving". Does (A) implies (B)?
2 responses



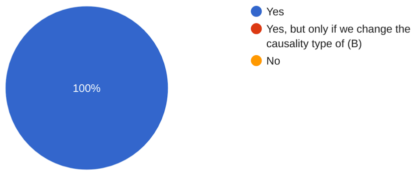
(A) "interpersonal interaction" affects positively "everyone trying to form relationship with other face to face". (B)...ther face to face". Does (A) implies (B)?
2 responses



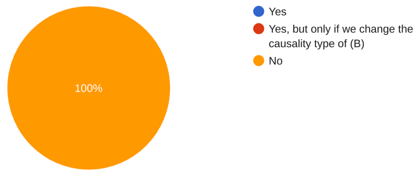
(A) "principal of one person one vote" affects positively "voting public". (B) "usa veto of palestinian un vote" affects ...ely "voting public". Does (A) implies (B)?
2 responses



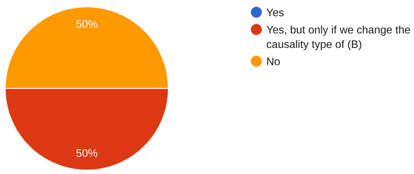
(A) "healthier menu" affects positively "dieter". (B) "producer of healthy food" affects positively "dieter". Does (A) implies (B)?
2 responses



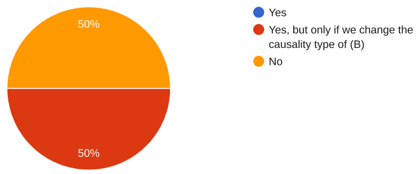
(A) "difference in way of life between" affects positively "happiness". (B) "cure many of society ill" affects positively "happiness". Does (A) implies (B)?
2 responses



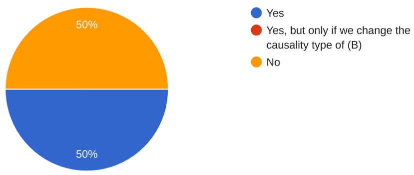
(A) "despair in child" affects positively "development". (B) "disappoint child" affects positively "development". Does (A) implies (B)?
2 responses



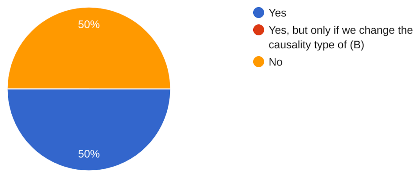
(A) "weaker greenhouse effect" affects negatively "emission do not decrease". (B) "transport emission" affec... do not decrease". Does (A) implies (B)?
2 responses



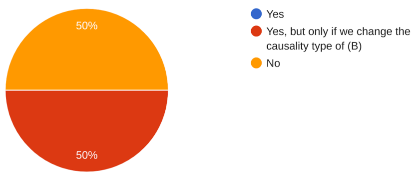
(A) "money spent on treating offender" affects negatively "prison cost taxation". (B) "castrating sex offender" a...ison cost taxation". Does (A) implies (B)?
2 responses



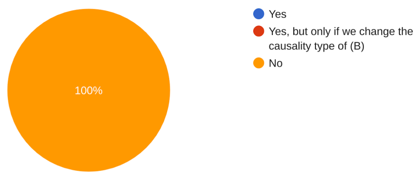
(A) "circumventing certain more ordinary legal" affects negatively "institution". (B) "greater cost of insurance" affects ne...ively "institution". Does (A) implies (B)?
2 responses



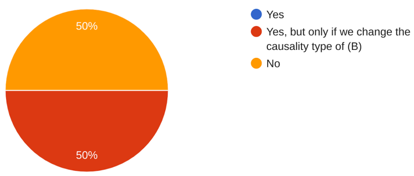
(A) "country s bank system" affects positively "stability". (B) "future risk taking by bank" affects positively "stability". Does (A) implies (B)?
2 responses



(A) "terrorist attack" affects negatively "public welnes". (B) "married" affects negatively "public welnes". Does (A) implies (B)?
2 responses

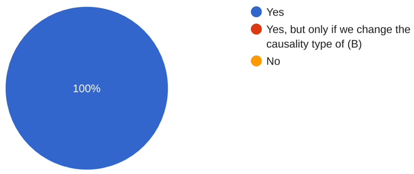


(A) "stability" affects negatively "group that propagate terrorism". (B) "islamic extremist insurgency" affects negatively...pagate terrorism". Does (A) implies (B)?
2 responses

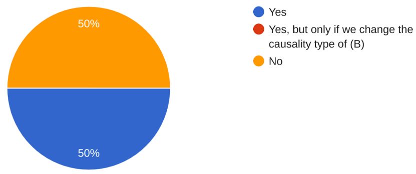


C.2 Case (2b)

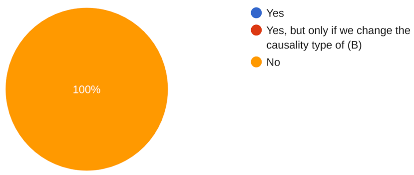
(A) "needle exchange" affects negatively "infection". (B) "needle exchange" affects negatively "risk of pulmonary embolism". Does (A) implies (B)?
2 responses



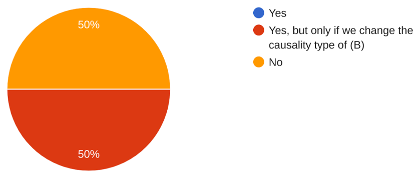
(A) "dropout rate" affects negatively "teacher". (B) "dropout rate" affects negatively "teacher to cheat". Does (A) implies (B)?
2 responses



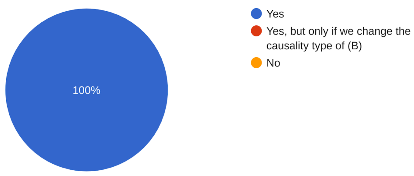
(A) "investing appropriately in water utility" affects positively "price". (B) "investing appropriately in water utility" ...tively "stock price". Does (A) implies (B)?
2 responses



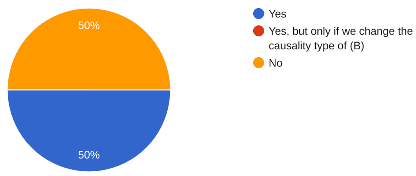
(A) "sitting in front of the tv" affects negatively "activity". (B) "sitting in front of the tv" affects negatively "use of champagne". Does (A) implies (B)?
2 responses



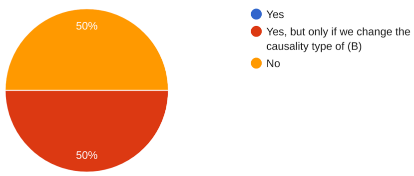
(A) "emergency" affects positively "circumventing certain more ordinary legal". (B) "emergency" affects positively ...k will be sloppier". Does (A) implies (B)?
2 responses



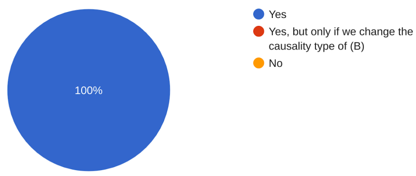
(A) "college regulation" affects positively "easier regulation". (B) "college regulation" affects positively "better gun control law". Does (A) implies (B)?
2 responses



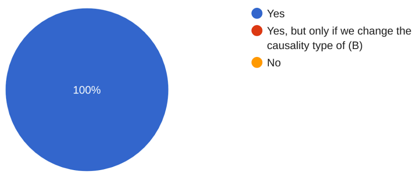
(A) "medical marijuana" affects positively "harder one". (B) "medical marijuana" affects positively "harder life for all of usa". Does (A) implies (B)?
2 responses



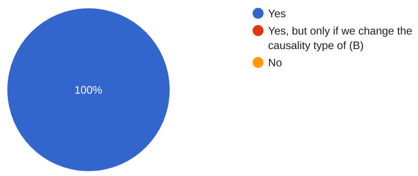
(A) "candidate to fundraise more" affects negatively "candidate". (B) "candidate to fundraise more" affects ne... party candidate". Does (A) implies (B)?
2 responses



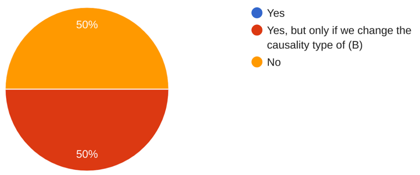
(A) "independent kosovo the most viable" affects positively "eu". (B) "independent kosovo the most viable" aff...ively "eu budget". Does (A) implies (B)?
2 responses



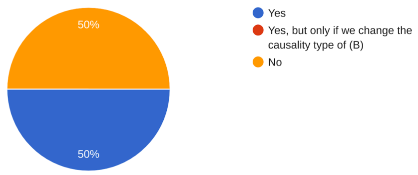
(A) "water life in the yangze river" affects positively "ecology". (B) "water life in the yangze river" affects positively "crop". Does (A) implies (B)?
2 responses



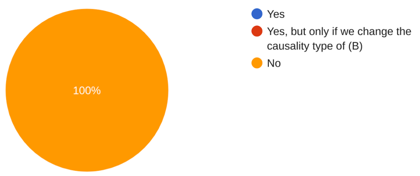
(A) "poor rating from student" affects negatively "teacher". (B) "poor rating from student" affects negatively "teache...st below passing". Does (A) implies (B)?
2 responses



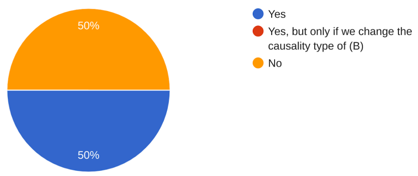
(A) "failing teaching method" affects negatively "teacher". (B) "failing teaching method" affects negatively "teacher to o...st below passing". Does (A) implies (B)?
2 responses



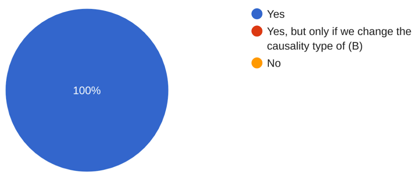
(A) "child performance" affects positively "mature rapidly". (B) "child performance" affects positively "massive job los". Does (A) implies (B)?
2 responses



(A) "terrorist attack" affects negatively "married". (B) "terrorist attack" affects negatively "public welnes". Does (A) implies (B)?
2 responses

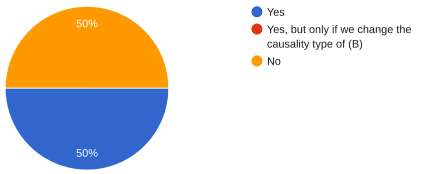


(A) "corrupt afghan government" affects negatively "middle east". (B) "corrupt afghan government" affects negatively "... eastern woman". Does (A) implies (B)?
2 responses

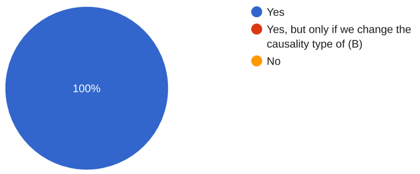


C.3 Case (3a)

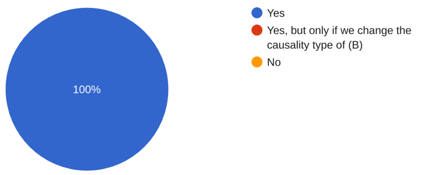
(A) "activity" affects positively "environmental health". (B) "screening" affects positively "environmental health". Does (A) implies (B)?
2 responses



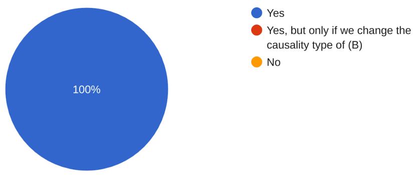
(A) "terrorism" affects negatively "person targeted by terrorism". (B) "terrorist attack" affects negatively "person targeted by terrorism". Does (A) implies (B)?
2 responses



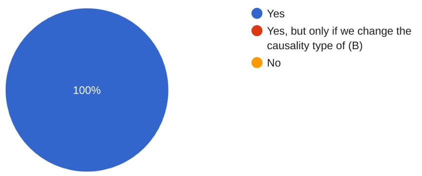
(A) "lower price" affects positively "working clas person". (B) "make water cheaper" affects positively "working clas person". Does (A) implies (B)?
2 responses



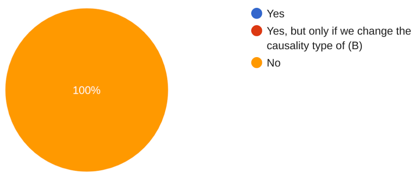
(A) "greenhouse ga emission" affects negatively "air pollution". (B) "human greenhouse ga emission" affects negatively "air pollution". Does (A) implies (B)?
2 responses



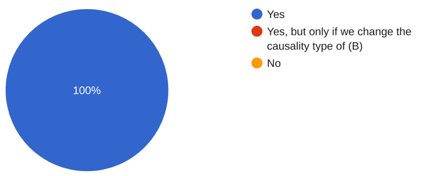
(A) "physical injury" affects negatively "teacher". (B) "permanent injury" affects negatively "teacher". Does (A) implies (B)?
2 responses



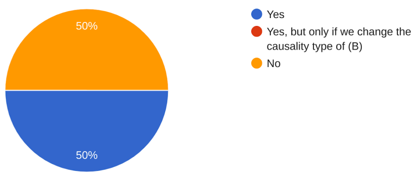
(A) "tear" affects negatively "break pedal wear". (B) "violence" affects negatively "break pedal wear". Does (A) implies (B)?
2 responses



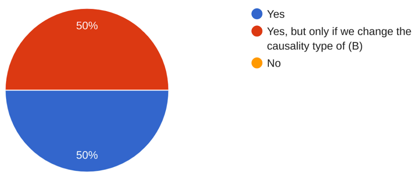
(A) "lower price" affects positively "working clas person". (B) "lower borrowing cost" affects positively "working clas person". Does (A) implies (B)?
2 responses



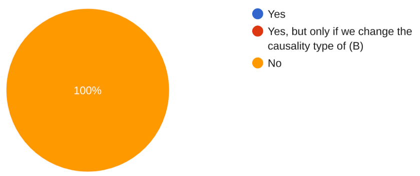
(A) "rallying cause of the insurgency" affects negatively "withdrawal from iraq". (B) "hasty withdrawal from iraq" af...hdrawal from iraq". Does (A) implies (B)?
2 responses



(A) "global affair" affects negatively "stability". (B) "global health" affects negatively "stability". Does (A) implies (B)?
2 responses

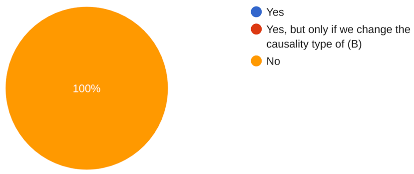


(A) "road safety" affects positively "international relation". (B) "world security" affects positively "international relation". Does (A) implies (B)?
2 responses

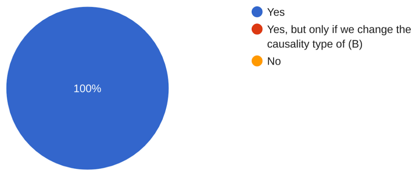


C.4 Case (3b)

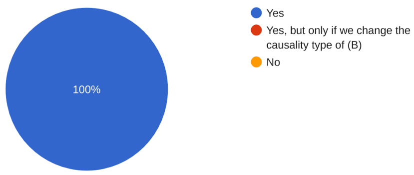
(A) "school would make more money" affects positively "market liquidity". (B) "school would make more money" affects positively "price". Does (A) implies (B)?
2 responses



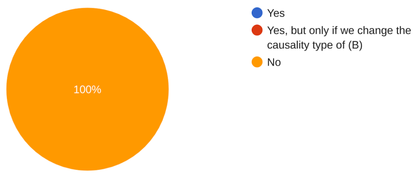
(A) "market acces" affects positively "market liquidity". (B) "market acces" affects positively "shop". Does (A) implies (B)?
2 responses



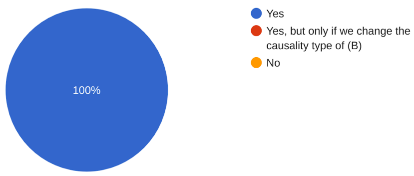
(A) "web 2 0 technology" affects positively "financial firm". (B) "web 2 0 technology" affects positively "budget". Does (A) implies (B)?
2 responses



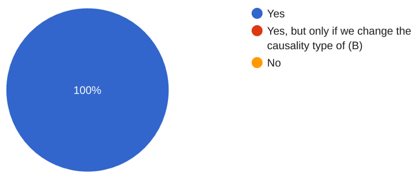
(A) "knowledge of human viability on alien planet" affects positively "effectively apply local knowledge". (B) ...ce in governance". Does (A) implies (B)?
2 responses



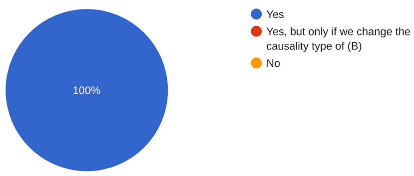
(A) "equality" affects positively "party". (B) "equality" affects positively "choice within a party". Does (A) implies (B)?
2 responses



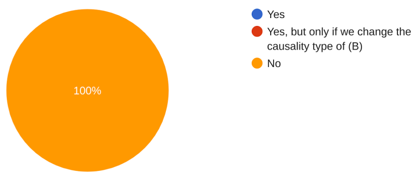
(A) "sound bite politic" affects negatively "member of the political party". (B) "sound bite politic" affects negatively "democrat". Does (A) implies (B)?
2 responses



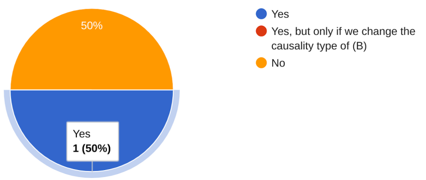
(A) "child to study science" affects positively "effectively apply local knowledge". (B) "child to study science"...exposure to work". Does (A) implies (B)?
2 responses



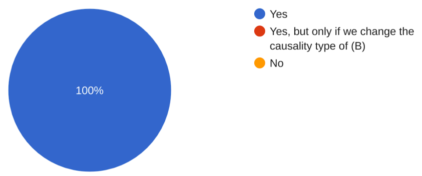
(A) "child to study science" affects positively "value of experience". (B) "child to study science" affects positively "wif...er understanding". Does (A) implies (B)?
2 responses



(A) "politician understand real world" affects positively "party". (B) "politician understand real world" affects positively...e party platform". Does (A) implies (B)?
2 responses



(A) "flag controversial issue for the public" affects positively "party". (B) "flag controversial issue for the public" affect...he party platform". Does (A) implies (B)?
2 responses

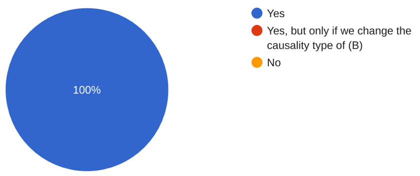


Appendix D

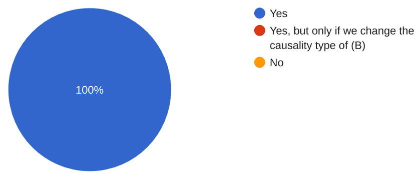
Second UTE Survey

D.1 Case (2a)

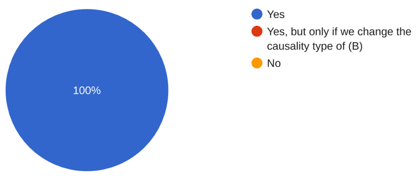
(A) "protection" affects positively "global demand for french champagne". (B) "safety net" affects positively "global d...french champagne". Does (A) implies (B)?
2 responses



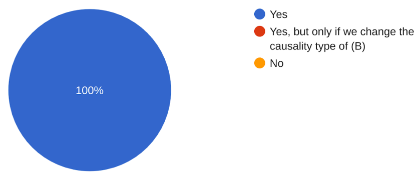
(A) "fresh face" affects positively "government". (B) "new candidate" affects positively "government". Does (A) implies (B)?
2 responses



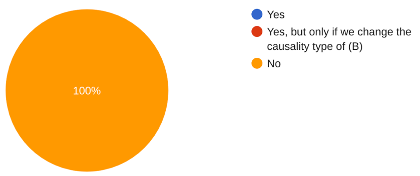
(A) "significant energy" affects positively "energy". (B) "massive quantity of energy" affects positively "energy". Does (A) implies (B)?
2 responses



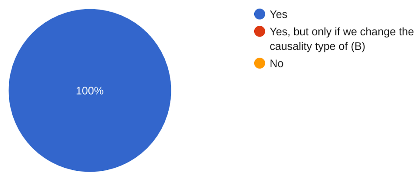
(A) "nuclear energy" affects positively "meet growing energy demand". (B) "nuclear power" affects positively "mee...g energy demand". Does (A) implies (B)?
2 responses



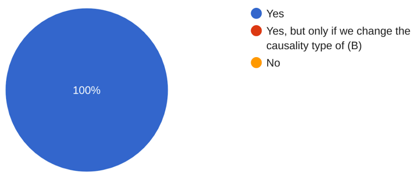
(A) "environment" affects positively "energy". (B) "massive quantity of energy" affects positively "energy". Does (A) implies (B)?
2 responses



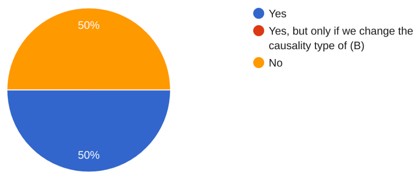
(A) "hydrogen car" affects positively "new industry". (B) "society using hydrogen car" affects positively "new industry". Does (A) implies (B)?
2 responses



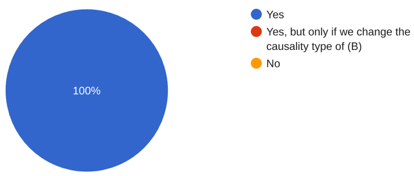
(A) "open source software" affects negatively "customer support". (B) "opennes of open source software" affec...stomer support". Does (A) implies (B)?
2 responses



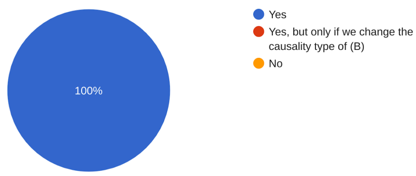
(A) "migration" affects positively "social cultural". (B) "immigrant to learn language" affects positively "social cultural". Does (A) implies (B)?
2 responses



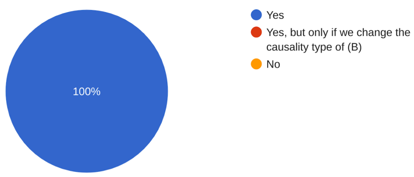
(A) "usa deficit" affects negatively "government". (B) "usa budget deficit" affects negatively "government". Does (A) implies (B)?
2 responses



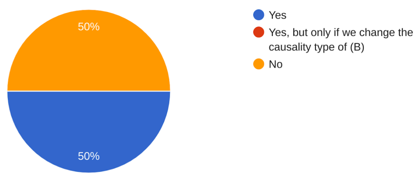
(A) "party affiliation" affects negatively "lessening voter turnout". (B) "member of the political party" affects negatively...ing voter turnout". Does (A) implies (B)?
2 responses



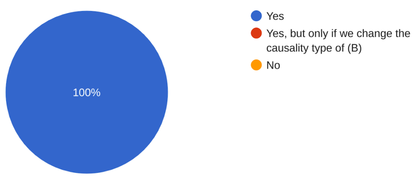
(A) "fossil fuel subsidy" affects negatively "climate change". (B) "subsidizing oil" affects negatively "climate change". Does (A) implies (B)?
2 responses



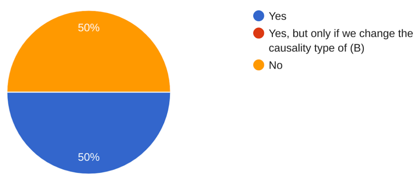
(A) "dangerou" affects negatively "japanese citizen". (B) "reporter in a very dangerou position" affects negatively "japanese citizen". Does (A) implies (B)?
2 responses



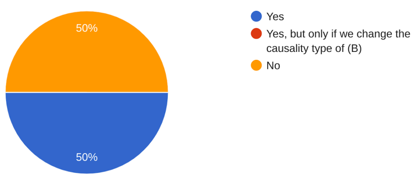
(A) "dollarization" affects negatively "risk of inflation". (B) "full dollarization" affects negatively "risk of inflation". Does (A) implies (B)?
2 responses



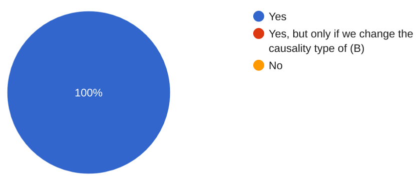
(A) "state spending" affects negatively "society". (B) "state cost" affects negatively "society". Does (A) implies (B)?
2 responses



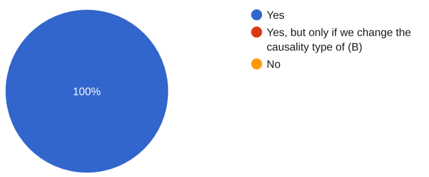
(A) "confidence" affects positively "society". (B) "shareholder confidence" affects positively "society". Does (A) implies (B)?
2 responses



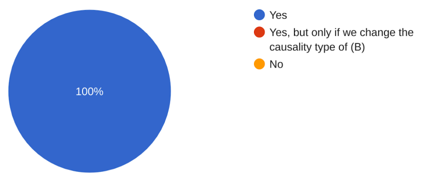
(A) "polygamy" affects positively "spread of venereal disease". (B) "large polygamou" affects positively "spread of venereal disease". Does (A) implies (B)?
2 responses



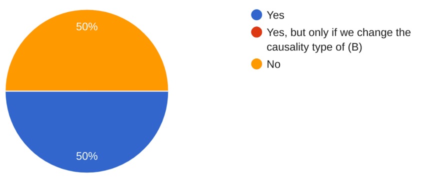
(A) "kosovo independence" affects positively "existing autonomy". (B) "support for kosovo independence" affec...sting autonomy". Does (A) implies (B)?
2 responses



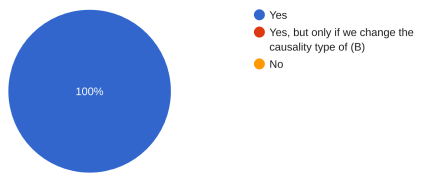
(A) "one child policy" affects positively "improve china for young generation". (B) "china one child policy" affects posi... young generation". Does (A) implies (B)?
2 responses



(A) "penalty" affects positively "society". (B) "death penalty" affects positively "society". Does (A) implies (B)?
2 responses

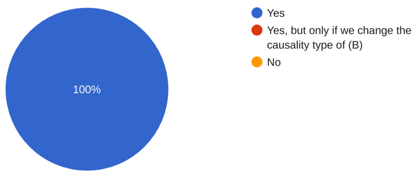


(A) "penalty" affects positively "company". (B) "fine and penalty" affects positively "company". Does (A) implies (B)?
2 responses

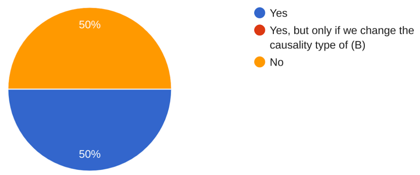


D.2 Case (2b)

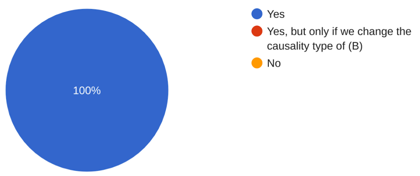
(A) "hassle of regulating conflict of interest" affects negatively "government". (B) "hassle of regulating conflict of int...ederal government". Does (A) implies (B)?
2 responses



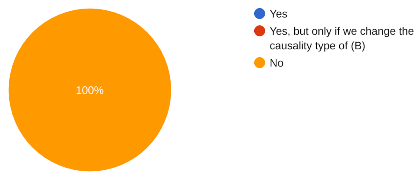
(A) "dictatorial move" affects negatively "government". (B) "dictatorial move" affects negatively "government debt". Does (A) implies (B)?
2 responses



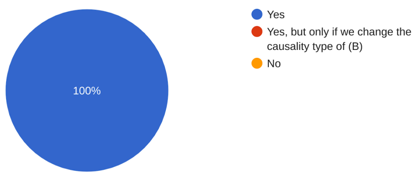
(A) "extremist" affects negatively "safety". (B) "extremist" affects negatively "ensure safety". Does (A) implies (B)?
2 responses



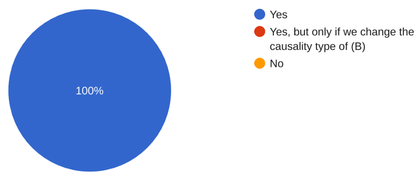
(A) "tourism demand" affects negatively "development". (B) "tourism demand" affects negatively "waterfront development". Does (A) implies (B)?
2 responses



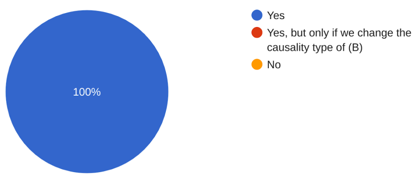
(A) "protection" affects positively "global demand for french champagne". (B) "protection" affects positively "safety net". Does (A) implies (B)?
2 responses



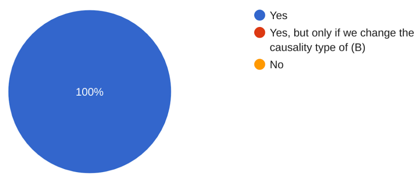
(A) "youth mind" affects positively "student". (B) "youth mind" affects positively "older mba student". Does (A) implies (B)?
2 responses



(A) "teacher independence" affects positively "student". (B) "teacher independence" affects positively "average student". Does (A) implies (B)?
2 responses

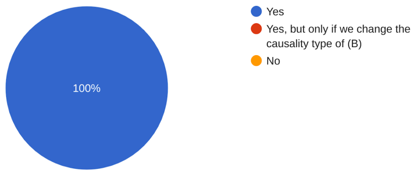


(A) "renewable alternative" affects positively "energy". (B) "renewable alternative" affects positively "world energy". Does (A) implies (B)?
2 responses



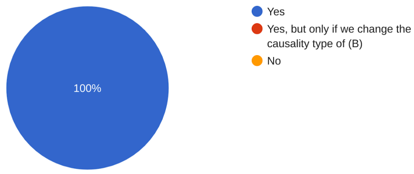
(A) "global warming" affects positively "energy". (B) "global warming" affects positively "world energy". Does (A) implies (B)?

2 responses



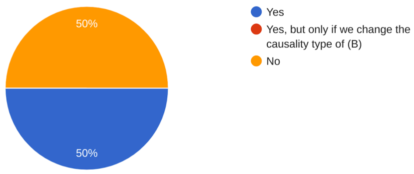
(A) "china demand pressure on world energy supply" affects positively "energy". (B) "china demand pressure on ...antity of energy". Does (A) implies (B)?

2 responses



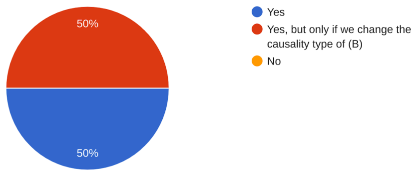
(A) "doctor" affects positively "public option". (B) "doctor" affects positively "public option monopoly". Does (A) implies (B)?

2 responses



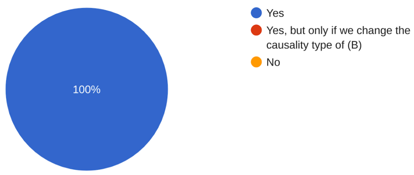
(A) "campaigning only to swing state" affects negatively "state". (B) "campaigning only to swing state" affect...tate government". Does (A) implies (B)?

2 responses



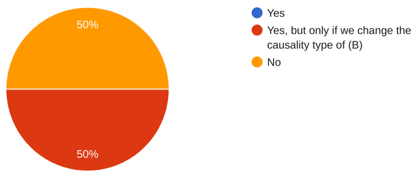
(A) "state with les renewable energy" affects negatively "state". (B) "state with les renewable energy" affects negatively...tate government". Does (A) implies (B)?

2 responses



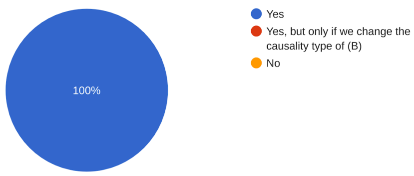
(A) "separate offense for date rape" affects positively "feel confident of succes in woman". (B) "separate offense ...ting expectation". Does (A) implies (B)?

2 responses



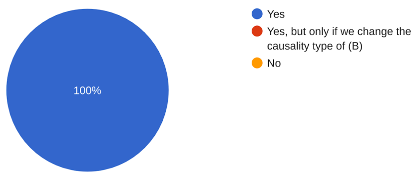
(A) "learn essential information" affects positively "standardized test". (B) "learn essential information" affects positively "get std test". Does (A) implies (B)?

2 responses

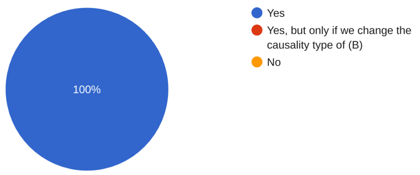


(A) "jealousy" affects negatively "polygamou wife". (B) "jealousy" affects negatively "wife of polygamist". Does (A) implies (B)?

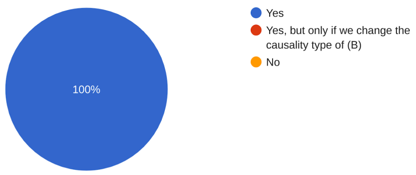
2 responses



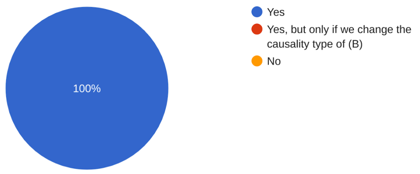
(A) "open primary" affects positively "nomination inconsistent with party view". (B) "open primary" affects positi...arty member view". Does (A) implies (B)?
2 responses



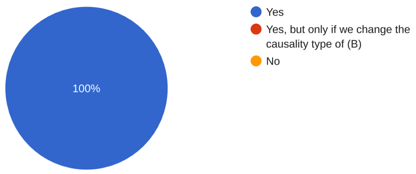
(A) "official from adjusting price" affects negatively "state". (B) "official from adjusting price" affects negatively "state government". Does (A) implies (B)?
2 responses



(A) "muslim resentment" affects negatively "kosovo". (B) "muslim resentment" affects negatively "independent kosovo the most viable". Does (A) implies (B)?
2 responses

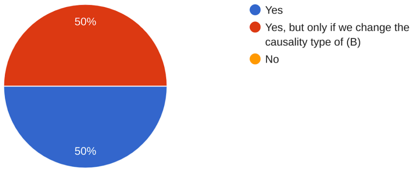


(A) "protect inmate" affects positively "prison". (B) "protect inmate" affects positively "prisoner". Does (A) implies (B)?
2 responses

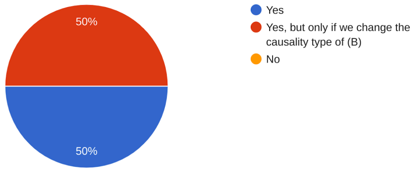


D.3 Case (2b)

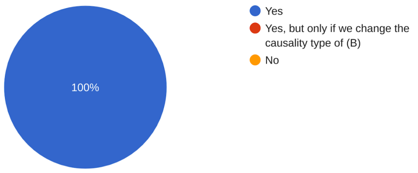
(A) "competition" affects positively "auto competition". (B) "unregulated competition" affects positively "auto competition". Does (A) implies (B)?
2 responses



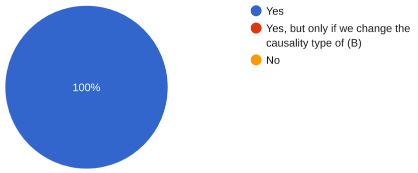
(A) "competitive" affects positively "society". (B) "foreign competition" affects positively "society". Does (A) implies (B)?
2 responses



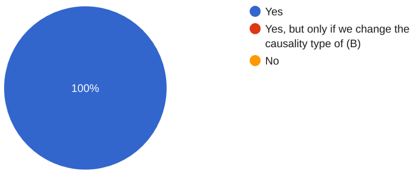
(A) "competition" affects positively "business". (B) "foreign competition" affects positively "business". Does (A) implies (B)?
2 responses



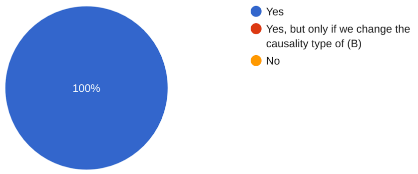
(A) "illegal immigration" affects negatively "united state". (B) "illegal immigrant in arizona" affects negatively "united state". Does (A) implies (B)?
2 responses



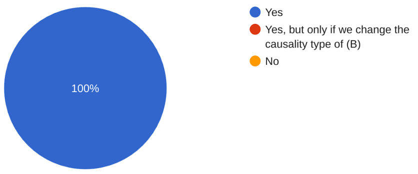
(A) "advancement" affects positively "medical advancement". (B) "eu expansion" affects positively "medical advancement". Does (A) implies (B)?
2 responses



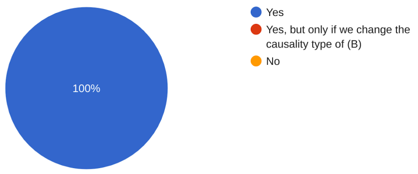
(A) "usa deficit" affects negatively "citizen". (B) "usa budget deficit" affects negatively "citizen". Does (A) implies (B)?
2 responses



(A) "violent video game" affects negatively "individual". (B) "like violent video game" affects negatively "individual". Does (A) implies (B)?
2 responses

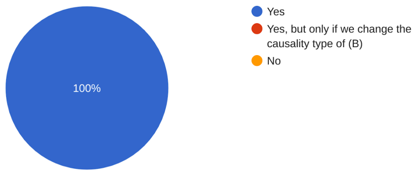


(A) "human s health" affects negatively "humanity". (B) "woman s health" affects negatively "humanity". Does (A) implies (B)?
2 responses



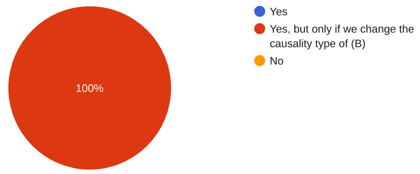
(A) "corruption" affects negatively "democratic proces". (B) "corrupt afghan government" affects negatively "democratic proces". Does (A) implies (B)?

2 responses



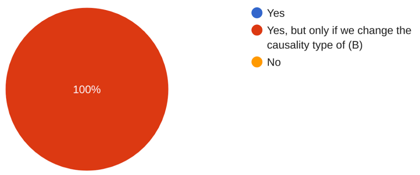
(A) "dependence on foreign oil" affects negatively "usa". (B) "usa break dependence on foreign oil" affects negatively "usa". Does (A) implies (B)?

2 responses



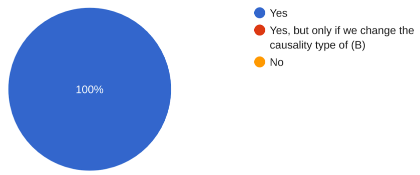
(A) "dependence on foreign oil" affects negatively "domestic producer". (B) "usa break dependence on foreign oil" af...mestic producer". Does (A) implies (B)?

2 responses



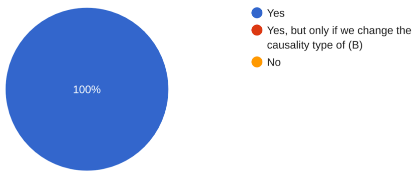
(A) "voter participation" affects negatively "citizen". (B) "voter turnout" affects negatively "citizen". Does (A) implies (B)?

2 responses



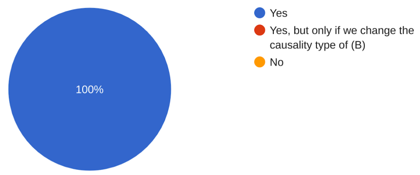
(A) "confidence" affects positively "economic". (B) "shareholder confidence" affects positively "economic". Does (A) implies (B)?

2 responses



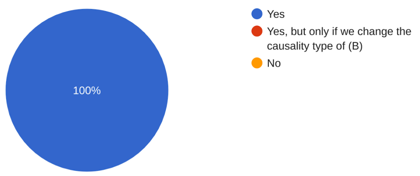
(A) "economy" affects positively "consumer". (B) "economic growth" affects positively "consumer". Does (A) implies (B)?

2 responses



(A) "kosovo independence" affects positively "separatist movement". (B) "independent kosovo the most viable" af...ratist movement". Does (A) implies (B)?

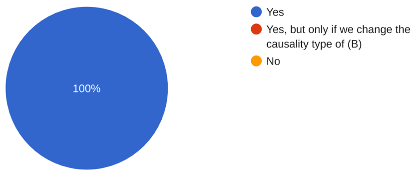
2 responses



D.4 Case (2b)

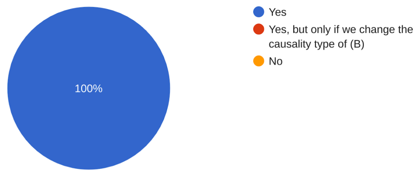
(A) "sex abuse" affects negatively "everyone". (B) "sex abuse" affects negatively "everyone in society". Does (A) implies (B)?

2 responses



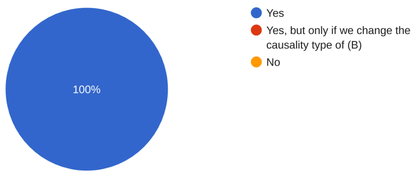
(A) "civil society" affects positively "chinese person". (B) "civil society" affects positively "chinese population". Does (A) implies (B)?

2 responses



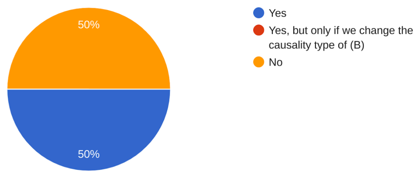
(A) "harm" affects negatively "student". (B) "harm" affects negatively "needy student". Does (A) implies (B)?

2 responses



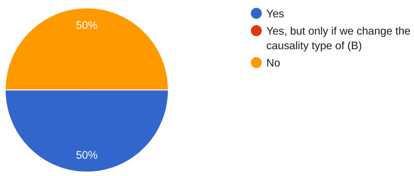
(A) "advancement" affects negatively "public insurance". (B) "advancement" affects negatively "greater prevention an...public insurance". Does (A) implies (B)?

2 responses



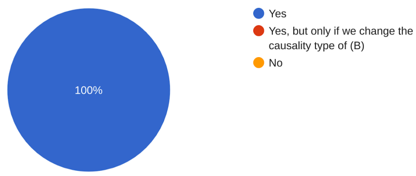
(A) "insurer" affects positively "private insurer". (B) "insurer" affects positively "competing with private insurer". Does (A) implies (B)?

2 responses



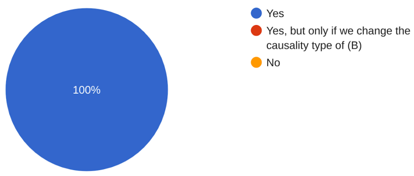
(A) "private insurance is pushed out" affects negatively "private insurance". (B) "private insurance is pushed out" affects...surance industry". Does (A) implies (B)?

2 responses



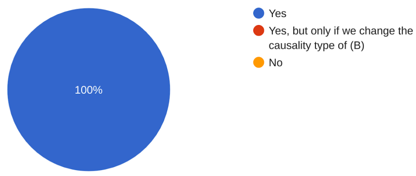
(A) "give doctor hospital million of new patient" affects positively "insurance company". (B) "give doctor hospital milli..."health insurance". Does (A) implies (B)?

2 responses

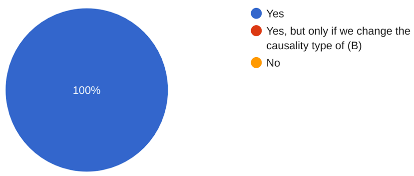


(A) "state can regulate the sale" affects positively "citizen". (B) "state can regulate the sale" affects positively "eu the citizen". Does (A) implies (B)?

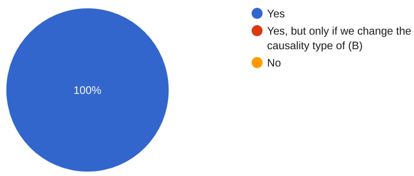
2 responses



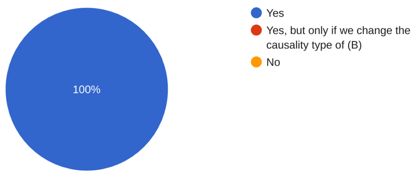
(A) "long term threat" affects negatively "citizen". (B) "long term threat" affects negatively "citizen right". Does (A) implies (B)?
2 responses



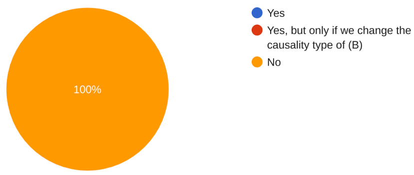
(A) "greater mainstream party representation" affects positively "democracy". (B) "greater mainstream party representa...ocratic process". Does (A) implies (B)?
2 responses



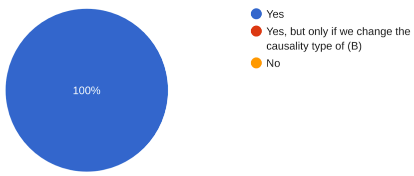
(A) "identity card" affects negatively "illegal immigration". (B) "identity card" affects negatively "illegal immigrant in arizona". Does (A) implies (B)?
2 responses



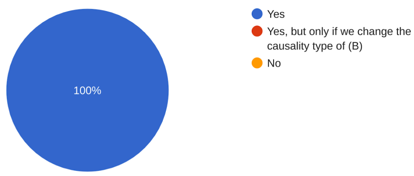
(A) "school environment" affects positively "better". (B) "school environment" affects positively "better sex". Does (A) implies (B)?
2 responses



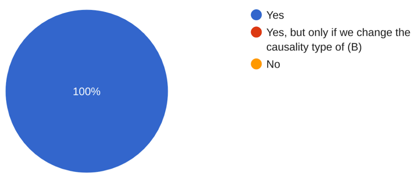
(A) "foreign oil dependency" affects negatively "economy". (B) "foreign oil dependency" affects positively "economic progress". Does (A) implies (B)?
2 responses



(A) "infection" affects negatively "needle user". (B) "infection" affects negatively "needle exchange user". Does (A) implies (B)?
2 responses



(A) "crime" affects negatively "united state". (B) "crime" affects negatively "image of united state". Does (A) implies (B)?
2 responses



Bibliography

[1] Meni Adler, Jonathan Berant, and Ido Dagan. Entailment-based text exploration with application to the health-care domain. In *Proceedings of the ACL 2012 System Demonstrations*, ACL '12, pages 79–84, USA, July 2012. Association for Computational Linguistics. 2.3, 3

[2] Yamen Ajjour, Henning Wachsmuth, Johannes Kiesel, Martin Potthast, Matthias Hagen, and Benno Stein. Data Acquisition for Argument Search: The args.me Corpus. In Christoph Benzmüller and Heiner Stuckenschmidt, editors, *KI 2019: Advances in Artificial Intelligence*, volume 11793, pages 48–59. Springer International Publishing, Cham, 2019. ISBN 978-3-030-30178-1 978-3-030-30179-8. doi: 10.1007/978-3-030-30179-8_4. URL http://link.springer.com/10.1007/978-3-030-30179-8_4. 2.4

[3] Khalid Al-Khatib, Yufang Hou, Henning Wachsmuth, Charles Jochim, Francesca Bonin, and Benno Stein. End-to-End Argumentation Knowledge Graph Construction. *AAAI*, 34(05):7367–7374, April 2020. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v34i05.6231. URL https://aaai.org/ojs/index.php/AAAI/article/view/6231. (document), 1, 1, 2.5, 3

[4] I. Androutsopoulos and P. Malakasiotis. A Survey of Paraphrasing and Textual Entailment Methods. *jair*, 38:135–187, May 2010. ISSN 1076-9757. doi: 10.1613/jair.2985. URL https://jair.org/index.php/jair/article/view/10651. 2.3

[5] Rico Angell, Nicholas Monath, Sunil Mohan, Nishant Yadav, and Andrew McCallum. Clustering-based Inference for Zero-Shot Biomedical Entity Linking. *arXiv:2010.11253 [cs]*, October 2020. URL http://arxiv.org/abs/2010.11253. 1

[6] Arvind Arasu, Christopher Ré, and Dan Suciu. Large-Scale Deduplication with Constraints Using Dedupalog. In *2009 IEEE 25th International Conference on Data Engineering*, pages 952–963, Shanghai, China, March 2009.

IEEE. ISBN 978-1-4244-3422-0. doi: 10.1109/ICDE.2009.43. URL http://ieeexplore.ieee.org/document/4812468/. 2.1, 2.3, 3

[7] Jonathan Berant, Ido Dagan, and Jacob Goldberger. Global learning of focused entailment graphs. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 1220–1229, USA, July 2010. Association for Computational Linguistics. 2.3, 3, 3.3.1, 3.3.1.1

[8] Roi Blanco, Giuseppe Ottaviano, and Edgar Meij. Fast and Space-Efficient Entity Linking for Queries. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 179–188, Shanghai China, February 2015. ACM. ISBN 978-1-4503-3317-7. doi: 10.1145/2684822.2685317. URL https://dl.acm.org/doi/10.1145/2684822.2685317. 2.1

[9] Lucas Carstens and Francesca Toni. Towards relation based Argumentation Mining. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 29–34, Denver, CO, 2015. Association for Computational Linguistics. doi: 10.3115/v1/W15-0504. URL http://aclweb.org/anthology/W15-0504. 2.4

[10] D. Cavar, J. Herring, and A. Meyer. Law Analysis using Deep NLP and Knowledge Graphs, 2018. URL /paper/Law-Analysis-using-Deep-NLP-and-Knowledge-Graphs-Cavar-Herring/c549f824620d041dbe490869bc1ebe78df9331e3. 1

[11] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Bipolarity in Argumentation Graphs: Towards a Better Understanding. *International Journal of Approximate Reasoning*, 54(7):876–899, 2013. ISSN 0888-613X. doi: 10.1016/j.ijar.2013.03.001. URL http://dx.doi.org/10.1016/j.ijar.2013.03.001. 1, 2.5

[12] Lihan Chen, Jiaqing Liang, Chenhao Xie, and Yanghua Xiao. Short Text Entity Linking with Fine-grained Topics. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 457–466, Torino Italy, October 2018. ACM. ISBN 978-1-4503-6014-2. doi: 10.1145/3269206.3271809. URL https://dl.acm.org/doi/10.1145/3269206.3271809. 2.1

[13] Robert Craven and Francesca Toni. Argument graphs and assumption-based argumentation. *Artificial Intelligence*, 233:1–59, April 2016. ISSN 00043702. doi: 10.1016/j.artint.2015.12.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0004370215001800. 1

[14] Silviu Cucerzan. Large-Scale Named Entity Disambiguation Based on Wikipedia Data. In *EMNLP-CoNLL*, page 9, 2007. 2.1

[15] Ido Dagan, Oren Glickman, and Bernardo Magnini. The PASCAL Recognising Textual Entailment Challenge. In Joaquin Quiñonero-Candela, Ido Dagan, Bernardo Magnini, and Florence d'Alché-Buc, editors, *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment*, volume 3944, pages 177–190. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-33427-9 978-3-540-33428-6. doi: 10.1007/11736790_9. URL `http://link.springer.com/10.1007/11736790_9`. 2.3

[16] Martin Davies, Ashley Barnett, and Tim Van Gelder. *Using Computer-Aided Argument Mapping to Teach Reasoning - Tim van Gelder Publications*, volume 8 of *Windsor Studies in Argumentation*, pages 131–176. WSIA, 2019. ISBN 978-0-920233-87-0. URL `https://sites.google.com/site/timvangelder/publications-1/using-computer-aided-argument-mapping-to-teach-reasoning`. 1, 2.5

[17] Earl DeLapp. REDeLapp/Panama-Papers-Network-Analysis, October 2020. URL `https://github.com/REDeLapp/Panama-Papers-Network-Analysis`. 1

[18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*, May 2019. URL `http://arxiv.org/abs/1810.04805`. 2.2

[19] Xin Luna Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Kevin Murphy, Shaohua Sun, and Wei Zhang. From data fusion to knowledge fusion. *Proc. VLDB Endow.*, 7(10):881–892, June 2014. ISSN 2150-8097. doi: 10.14778/2732951.2732962. URL `https://dl.acm.org/doi/10.14778/2732951.2732962`. 1

[20] Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, and Wei Zhang. Knowledge-Based Trust: Estimating the Trustworthiness of Web Sources. *arXiv:1502.03519 [cs]*, February 2015. URL `http://arxiv.org/abs/1502.03519`. 1

[21] Lisa Ehrlinger and Wolfram Wöß. Towards a Definition of Knowledge Graphs. In *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems*, volume 1695, Leipzig, Germany, 2016. URL `http://ceur-ws.org/Vol-1695/paper4.pdf`. 1

[22] Ahmed K. Elmagarmid, Panagiotis G. Ipeirotis, and Vassilios S. Verykios. Duplicate Record Detection: A Survey. *IEEE Trans. Knowl. Data Eng.*, 19(1):1–16, January 2007. ISSN 1041-4347. doi: 10.1109/TKDE.2007.250581. URL `http://ieeexplore.ieee.org/document/4016511/`. 2.1

[23] Xiuyi Fan and Francesca Toni. Conflict resolution with argumentation dialogues. In *10th International Conference on Autonomous Agents and Multiagent Systems 2011, AAMAS 2011*, volume 2, pages 1095–1096, January 2011. 1

[24] Vanessa Wei Feng and Graeme Hirst. Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 987–996, USA, June 2011. Association for Computational Linguistics. ISBN 978-1-932432-87-9. 2.4

[25] Paolo Ferragina and Ugo Scaiella. TAGME: On-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management - CIKM '10*, page 1625, Toronto, ON, Canada, 2010. ACM Press. ISBN 978-1-4503-0099-5. doi: 10.1145/1871437.1871689. URL `http://portal.acm.org/citation.cfm?doid=1871437.1871689`. 2.1

[26] Rauscher Francois, Matta Nada, and Atifi Hassan. How to Extract Knowledge from Professional E-Mails. In *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 687–692, Bangkok, Thailand, November 2015. IEEE. ISBN 978-1-4673-9721-6. doi: 10.1109/SITIS.2015.113. URL `http://ieeexplore.ieee.org/document/7400638/`. 2.1

[27] Octavian-Eugen Ganea, Marina Ganea, Aurelien Lucchi, Carsten Eickhoff, and Thomas Hofmann. Probabilistic Bag-Of-Hyperlinks Model for Entity Linking. *arXiv:1509.02301 [cs]*, January 2016. URL `http://arxiv.org/abs/1509.02301`. 2.1

[28] Ning Gao, Mark Dredze, and Douglas W. Oard. Person entity linking in email with NIL detection. *Journal of the Association for Information Science and Technology*, 68(10):2412–2424, October 2017. ISSN 23301635. doi: 10.1002/asi.23888. URL `http://doi.wiley.com/10.1002/asi.23888`. 2.1

[29] Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. AllenNLP: A Deep Semantic Natural Language Processing Platform. In *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, pages 1–6, Melbourne, Australia, 2018. Association for Computational Linguistics. doi: 10.

18653/v1/W18-2501. URL http://aclweb.org/anthology/W18-2501. 2.3, 3, 3.3

[30] Debela Gemechu and Chris Reed. Decompositional Argument Mining: A General Purpose Approach for Argument Graph Construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 516–526, Florence, Italy, 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1049. URL https://www.aclweb.org/anthology/P19-1049. 2.4

[31] Oren Glickman and Ido Dagan. A probabilistic setting and lexical cooccurrence model for textual entailment. In *Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment*, EMSEE '05, pages 43–48, USA, June 2005. Association for Computational Linguistics. 2.3

[32] Yuhang Guo, Wanxiang Che, Ting Liu, and Sheng Li. A Graph-based Method for Entity Linking. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 1010–1018, Chiang Mai, Thailand, 2011. Asian Federation of Natural Language Processing. URL https://www.aclweb.org/anthology/I11-1113.pdf. 2.1

[33] Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. Entity Linking in Queries: Tasks and Evaluation. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 171–180, Northampton Massachusetts USA, September 2015. ACM. ISBN 978-1-4503-3833-2. doi: 10.1145/2808194.2809473. URL https://dl.acm.org/doi/10.1145/2808194.2809473. 2.1

[34] Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. Entity Linking in Queries: Efficiency vs. Effectiveness. In Joemon M Jose, Claudia Hauff, Ismail Sengor Altıngovde, Dawei Song, Dyaa Albakour, Stuart Watt, and John Tait, editors, *Advances in Information Retrieval*, volume 10193, pages 40–53. Springer International Publishing, Cham, 2017. ISBN 978-3-319-56607-8 978-3-319-56608-5. doi: 10.1007/978-3-319-56608-5_4. URL http://link.springer.com/10.1007/978-3-319-56608-5_4. 2.1

[35] Keith W. Hipel, Liping Fang, and D. Marc Kilgour. The Graph Model for Conflict Resolution: Reflections on Three Decades of Development. *Group Decis Negot*, 29(1):11–60, February 2020. ISSN 1572-9907. doi: 10.1007/s10726-019-09648-z. URL https://doi.org/10.1007/s10726-019-09648-z. 1

[36] T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(7):881–892, July 2002. ISSN 0162-8828. doi: 10.1109/TPAMI.2002.1017616. URL http://ieeexplore.ieee.org/document/1017616/. 2.2

[37] Khalid Al Khatib, Yufang Hou, Henning Wachsmuth, Charles Jochim, Francesca Bonin, and Benno Stein. End-to-End Argumentation Knowledge Graph Construction. In *AAAI*, 2020. doi: 10.1609/AAAI.V34I05.6231. 2.5

[38] Hanna Köpcke, Andreas Thor, and Erhard Rahm. Evaluation of entity resolution approaches on real-world match problems. *Proc. VLDB Endow.*, 3(1-2):484–493, September 2010. ISSN 2150-8097. doi: 10.14778/1920841.1920904. URL https://dl.acm.org/doi/10.14778/1920841.1920904. 2.1

[39] Jey Han Lau and Timothy Baldwin. An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 78–86, Berlin, Germany, 2016. Association for Computational Linguistics. doi: 10.18653/v1/W16-1609. URL http://aclweb.org/anthology/W16-1609. 2.2

[40] John Lawrence and Chris Reed. Argument Mining: A Survey. *Computational Linguistics*, 45(4):765–818, January 2020. ISSN 0891-2017, 1530-9312. doi: 10.1162/coli_a_00364. URL https://www.mitpressjournals.org/doi/abs/10.1162/coli_a_00364. 2.4

[41] Quoc V. Le and Tomas Mikolov. Distributed Representations of Sentences and Documents. *arXiv:1405.4053 [cs]*, May 2014. URL http://arxiv.org/abs/1405.4053. 2.2

[42] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv:1907.11692 [cs]*, July 2019. URL http://arxiv.org/abs/1907.11692. 2.3

[43] Edgar Meij, Krisztian Balog, and Daan Odijk. Entity linking and retrieval for semantic search. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, pages 683–684, New York New York USA, February 2014. ACM. ISBN 978-1-4503-2351-2. doi: 10.1145/2556195.2556201. URL https://dl.acm.org/doi/10.1145/2556195.2556201. 2.1

[44] Rada Mihalcea, Doina Ta'tar, Gabriela Serban, and Andreea Mihis. Textual Entailment as a Directional Relation. *Journal of Research and Practice in Information Technology*, 41(1):12, 2009. 2.3, 3, 3.3.2

[45] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781 [cs]*, September 2013. URL `http://arxiv.org/abs/1301.3781`. 2.2

[46] Marie-Francine Moens, Erik Boiy, Raquel Mochales Palau, and Chris Reed. Automatic detection of arguments in legal texts. In *Proceedings of the 11th International Conference on Artificial Intelligence and Law - ICAIL '07*, page 225, Stanford, California, 2007. ACM Press. ISBN 978-1-59593-680-6. doi: 10.1145/1276318.1276362. URL `http://portal.acm.org/citation.cfm?doid=1276318.1276362`. 2.4

[47] Emanuela Moreale and Maria Vargas-Vera. A Question-Answering System Using Argumentation. In Raúl Monroy, Gustavo Arroyo-Figueroa, Luis Enrique Sucar, and Humberto Sossa, editors, *MICAI 2004: Advances in Artificial Intelligence*, Lecture Notes in Computer Science, pages 400–409, Berlin, Heidelberg, 2004. Springer. ISBN 978-3-540-24694-7. doi: 10.1007/978-3-540-24694-7_41. 1

[48] R. Nagarajan. Document Clustering Using Agglomerative Hierarchical Clustering Approach (ahdc) and Proposed Tsg Keyword Extraction Method. *IJRET*, 05(11):118–124, November 2016. ISSN 23217308, 23191163. doi: 10.15623/ijret.2016.0511023. URL `https://ijret.org/volumes/2016v05/i11/IJRET20160511023.pdf`. 2.2

[49] Nathan Ong, Diane Litman, and Alexandra Brusilovsky. Ontology-Based Argument Mining and Automatic Essay Scoring. In *Proceedings of the First Workshop on Argumentation Mining*, pages 24–28, Baltimore, Maryland, 2014. Association for Computational Linguistics. doi: 10.3115/v1/W14-2104. URL `http://aclweb.org/anthology/W14-2104`. 2.4

[50] Claudia María Álvarez Ortiz. Does Philosophy Improve Critical Thinking Skills? Master's thesis, The University of Melbourne, 2007. 1

[51] Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *SW*, 8(3):489–508, December 2016. ISSN 22104968, 15700844. doi: 10.3233/SW-160218. URL `https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/SW-160218`. 1

[52] Ted Pedersen, Serguei V. S. Pakhomov, Siddharth Patwardhan, and Christopher G. Chute. Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40(3):288–299, June 2007. ISSN 1532-0464. doi: 10.1016/j.jbi.2006.06.004. URL https://www.sciencedirect.com/science/article/pii/S1532046406000645. 2.2, 3

[53] Andreas Peldszus and Manfred Stede. From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1):1–31, 2013. 2.4

[54] Isaac Persing and Vincent Ng. Modeling Argument Strength in Student Essays. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 543–552, Beijing, China, 2015. Association for Computational Linguistics. doi: 10.3115/v1/P15-1053. URL http://aclweb.org/anthology/P15-1053. 2.4

[55] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv:1908.10084 [cs]*, August 2019. URL http://arxiv.org/abs/1908.10084. 2.2

[56] Anderson Rossanez, Julio Cesar dos Reis, Ricardo da Silva Torres, and Hélène de Ribaupierre. KGen: A knowledge graph generator from biomedical scientific literature. *BMC Medical Informatics and Decision Making*, 20(4):314, December 2020. ISSN 1472-6947. doi: 10.1186/s12911-020-01341-5. URL https://doi.org/10.1186/s12911-020-01341-5. 1

[57] Hassan A. Sleiman and Rafael Corchuelo. A Survey on Region Extractors from Web Documents. *IEEE Trans. Knowl. Data Eng.*, 25(9):1960–1981, September 2013. ISSN 1041-4347. doi: 10.1109/TKDE.2012.135. URL http://ieeexplore.ieee.org/document/6231632/. 2.1

[58] Kostas Stefanidis, Vasilis Efthymiou, Melanie Herschel, and Vassilis Christophides. Entity resolution in the web of data. In *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*, pages 203–204, Seoul, Korea, 2014. ACM Press. ISBN 978-1-4503-2745-9. doi: 10.1145/2567948.2577263. URL http://dl.acm.org/citation.cfm?doid=2567948.2577263. 2.1

[59] Gizem Unal. Use knowledge graphs to understand your customers, January 2020. URL https://cambridge-intelligence.com/knowledge-graphs-to-understand-customers/. 1

[60] Wendi Usino, Anton Satria, Khalid Hamed, Arif Bramantoro, Hasniaty A, and Wahyu Amaldi. Document Similarity Detection using K-Means and Cosine Distance. *ijacsa*, 10(2), 2019. ISSN 21565570, 2158107X. doi: 10.14569/IJACSA.2019.0100222. URL `http://thesai.org/Publications/ViewPaper?Volume=10&Issue=2&Code=ijacsa&SerialNo=22`. 2.2

[61] Darnes Vilariño, David Pinto, Mireya Tovar, Saul León, and Esteban Castillo. BUAP: Lexical and semantic similarity for cross-lingual textual entailment. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, SemEval '12, pages 706–709, USA, June 2012. Association for Computational Linguistics. 2.3, 3

[62] Ngoc Phuoc An Vo and Octavian Popescu. Corpora for Learning the Mutual Relationship between Semantic Relatedness and Textual Entailment. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3379–3386, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA). URL `https://www.aclweb.org/anthology/L16-1539`. 2.3, 3

[63] Tim Weninger, Rodrigo Palacios, Valter Crescenzi, Thomas Gottron, and Paolo Merialdo. Web Content Extraction - a Meta-Analysis of its Past and Thoughts on its Future. *arXiv:1508.04066 [cs]*, August 2015. URL `http://arxiv.org/abs/1508.04066`. 2.1

[64] Mary McGee Wood, Craig Jones, John Sargeant, and Phil Reed. Light-Weight Clustering Techniques for Short Text Answers in Human Computer Collaborative (HCC) CAA, 2008. 2.2

[65] Han Xiao. Hanxiao/bert-as-service, March 2021. URL `https://github.com/hanxiao/bert-as-service`. 2.2

[66] D. Xu, Chuanwei Ruan, Evren Körpeoglu, Sushant Kumar, and Kannan Achan. Product Knowledge Graph Embedding for E-commerce. *WSDM*, 2020. doi: 10.1145/3336191.3371778. 1

[67] Vijaya Krishna Yalavarthi, Xiangyu Ke, and Arijit Khan. Select Your Questions Wisely: For Entity Resolution With Crowd Errors. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 317–326, Singapore Singapore, November 2017. ACM. ISBN 978-1-4503-4918-5. doi: 10.1145/3132847.3132876. URL `https://dl.acm.org/doi/10.1145/3132847.3132876`. 2.1

[68] Wenpeng Yin, Nazneen Fatema Rajani, Dragomir Radev, Richard Socher, and Caiming Xiong. Universal Natural Language Processing with Limited Annotations: Try Few-shot Textual Entailment as a Start. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8229–8239, Online, 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.660. URL https://www.aclweb.org/anthology/2020.emnlp-main.660. 2.3

[69] Wei Zhang, Xiaogang Wang, Deli Zhao, and Xiaoou Tang. Graph Degree Linkage: Agglomerative Clustering on a Directed Graph. *arXiv:1208.5092 [cs, stat]*, August 2012. URL http://arxiv.org/abs/1208.5092. 2.2