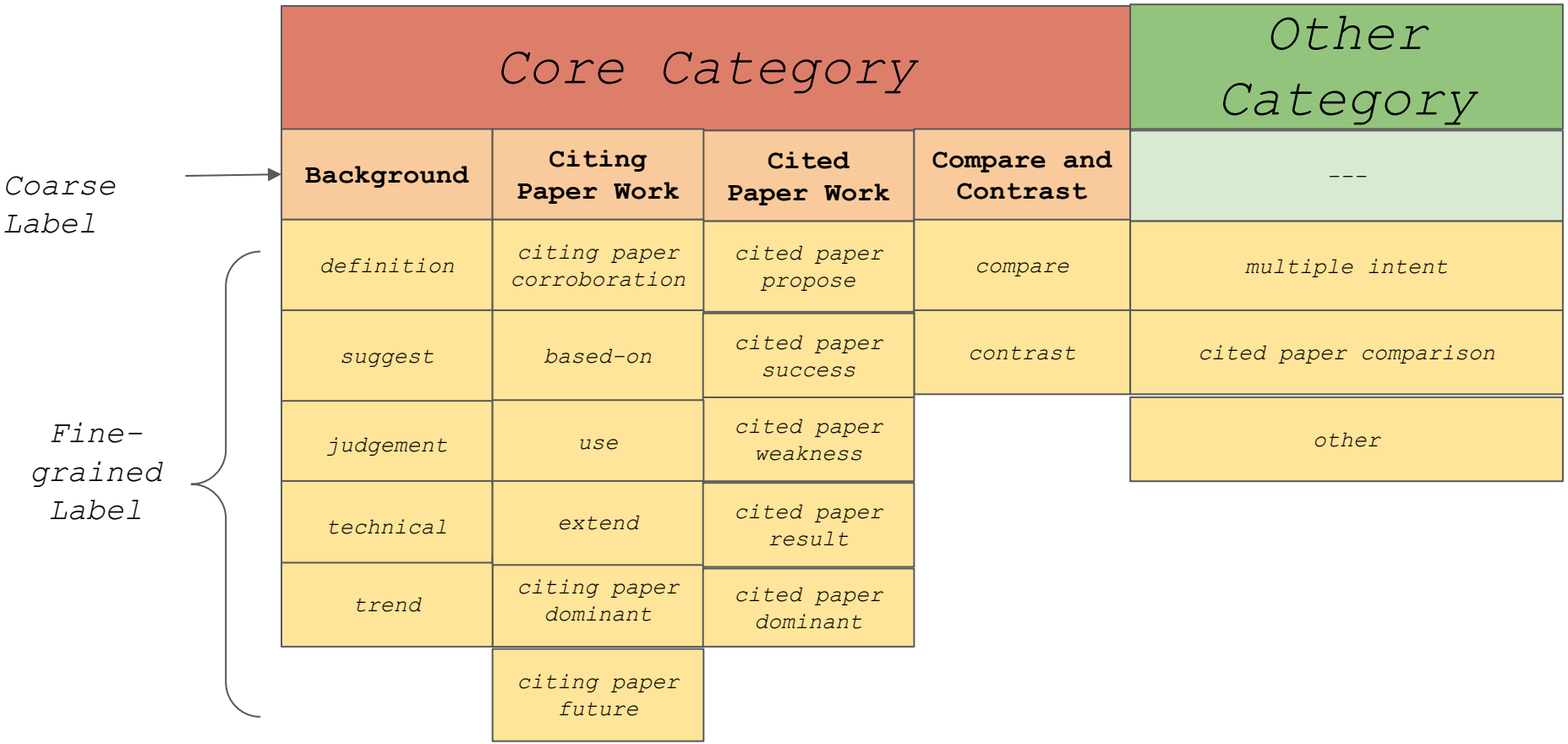
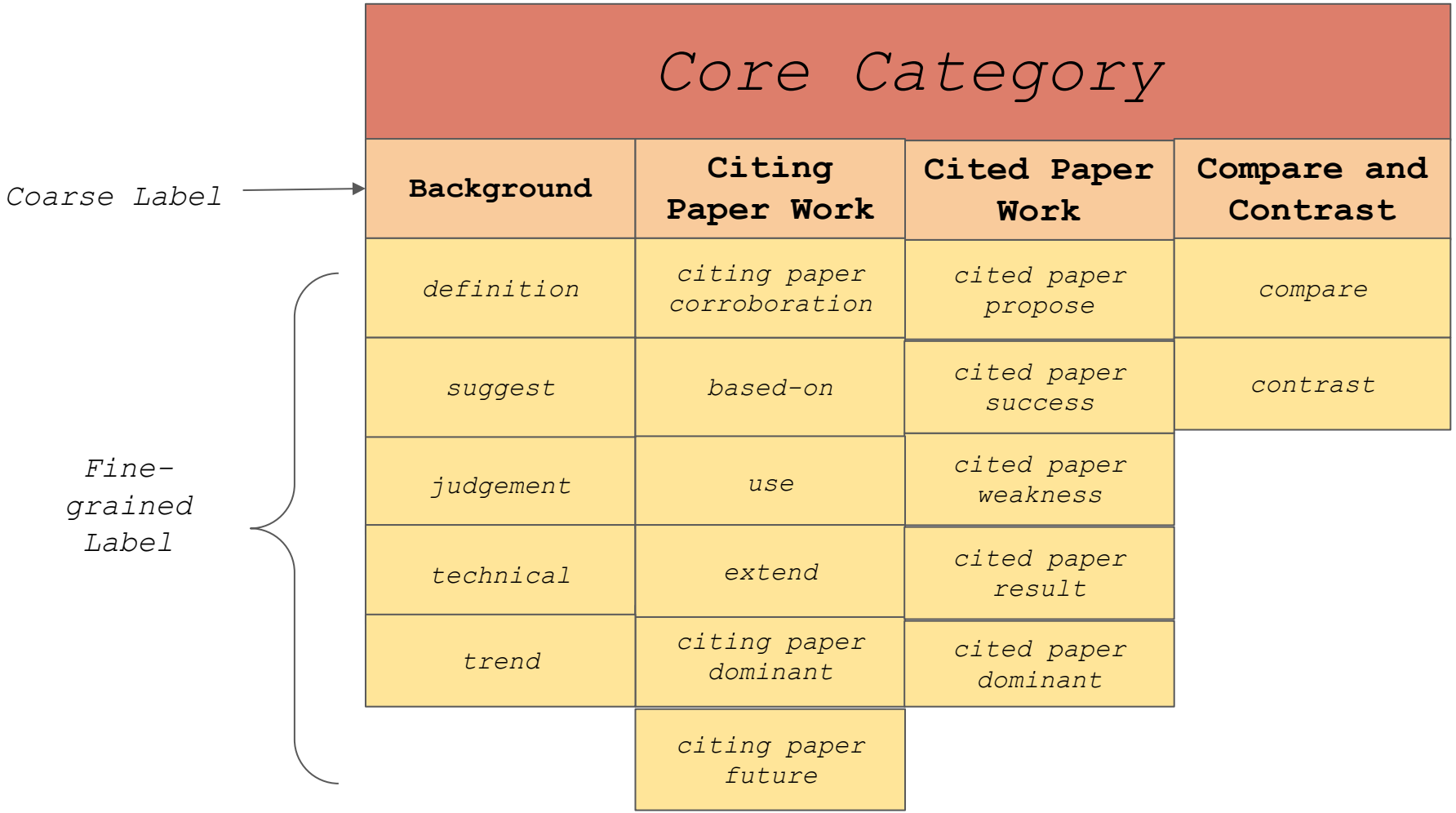
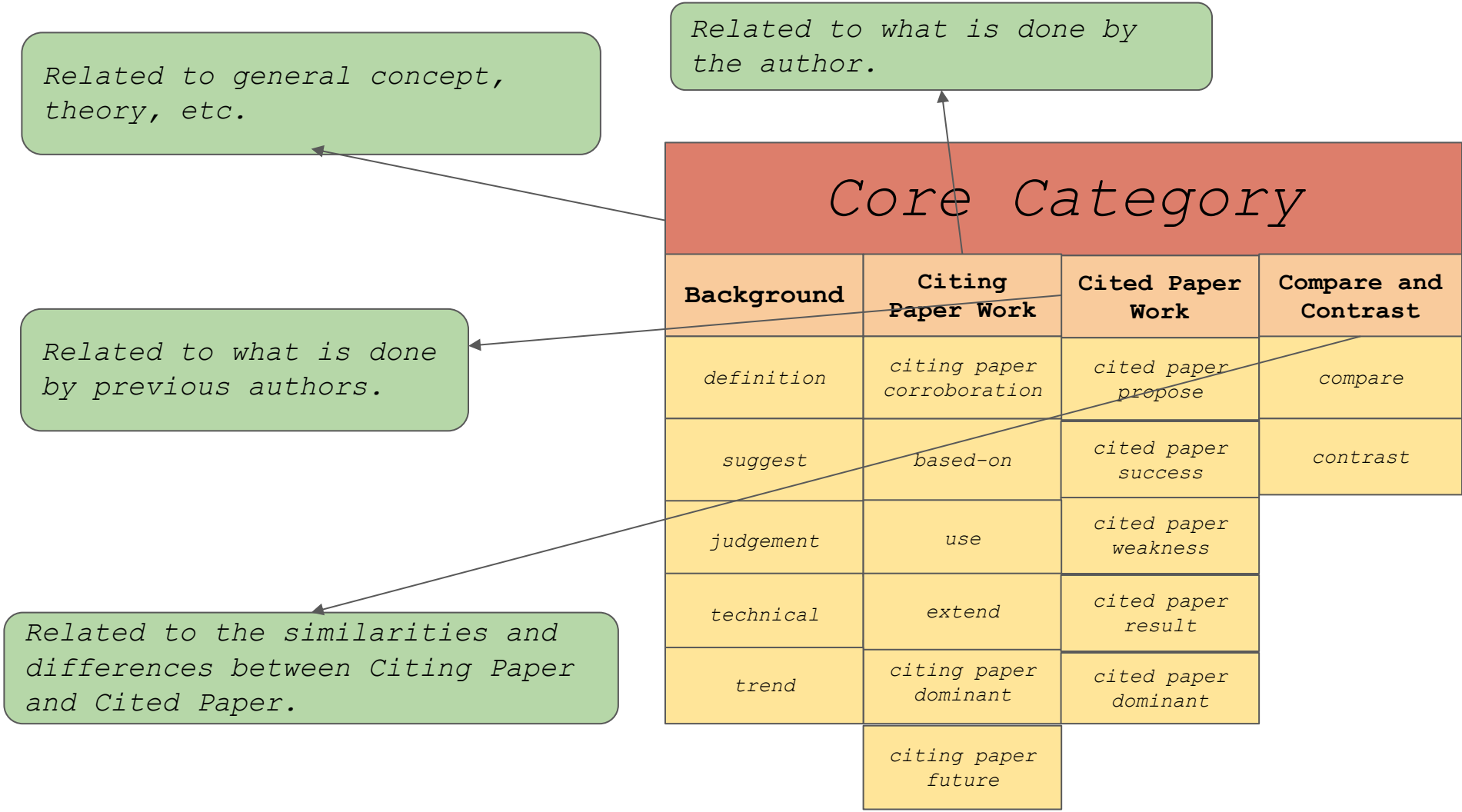
Labeling Guidance

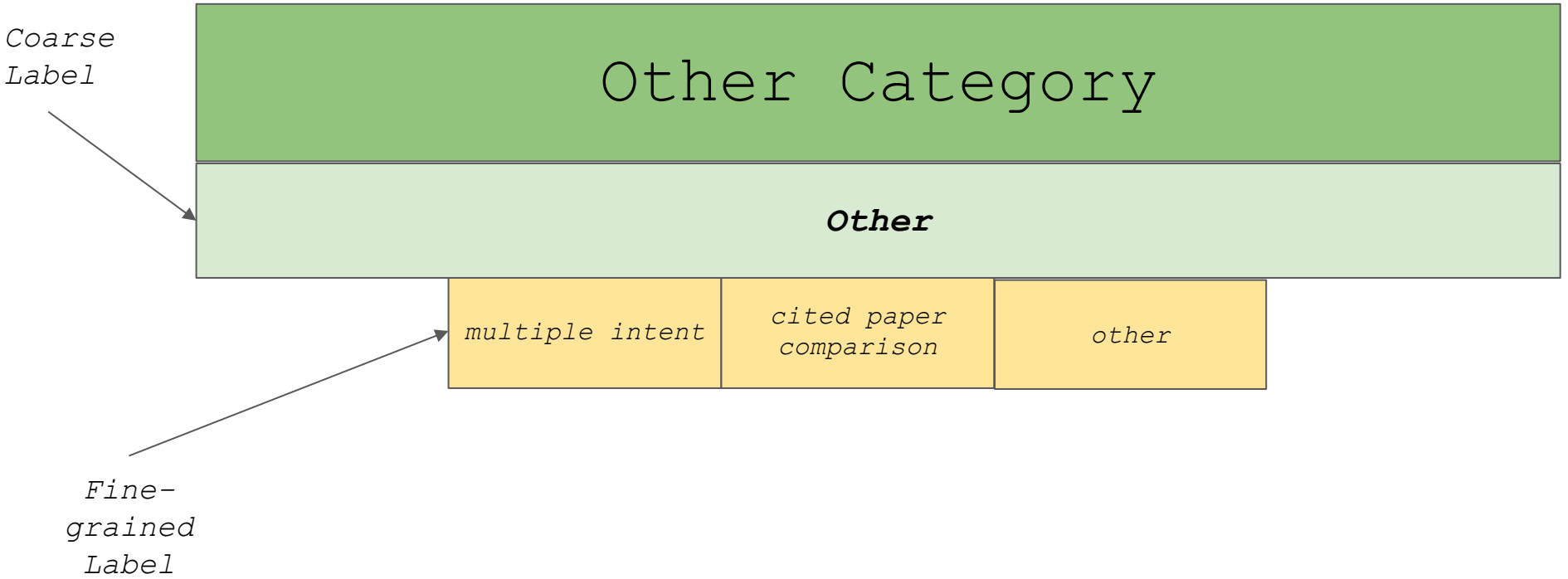
Setio Basuki
Ph.D student on Department of Computer Science and Engineering
Toyohashi University of Technology - Japan
2020

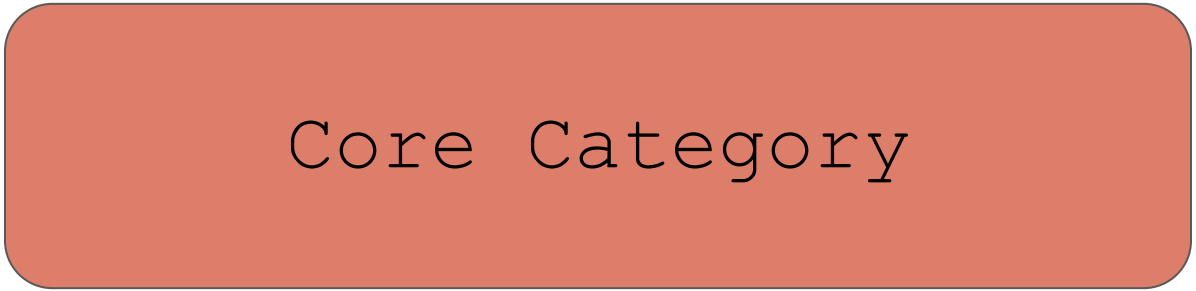

Core Category
Other Category
Coarse Label
Fine-grained Label
Background
Citing Paper Work
Cited Paper Work
Compare and Contrast
---
definition
citing paper corroboration
cited paper propose
compare
multiple intent
suggest
based-on
cited paper success
contrast
cited paper comparison
judgement
use
cited paper weakness
other
technical
extend
cited paper result
trend
citing paper dominant
cited paper dominant
citing paper future


Core Category
Coarse Label
Background
Citing Paper Work
Cited Paper Work
Compare and Contrast
definition
citing paper corroboration
cited paper propose
compare
suggest
based-on
cited paper success
contrast
judgement
use
cited paper weakness
technical
extend
cited paper result
trend
citing paper dominant
cited paper dominant
citing paper future
Fine-grained Label


Related to general concept, theory, etc.
Related to what is done by the author.
Core Category
Related to what is done by previous authors.
Related to the similarities and differences between Citing Paper and Cited Paper.
Background
Citing Paper Work
Cited Paper Work
Compare and Contrast
definition
suggest
judgement
technical
trend
citing paper corroboration
based-on
use
extend
citing paper dominant
citing paper future
cited paper propose
cited paper success
cited paper weakness
cited paper result
cited paper dominant
compare
contrast

Other category is used to accommodate the citing sentences that cannot be categorized into Core categories.




Core Category

Background

(1) definition

The label is used to explain the definition of general theory/principle/concept/topic/problem, etc.

Example:

Gigaword <citation> is a large automatically labeled dataset with 2,452 news articles and is used in TempEval-3.

(2) suggest

Giving the reader suggestion to refer, see more detail, or explore other cited papers.

Example:

A rigorous and detailed presentation of the theory, as well as the proofs of all these properties can be found in <citation>.

(3) judgement

Providing justification to theory, principle, basic concept, topic, problem, etc.

Example:

Word overlap is an important feature in many f2-type tasks <citation> especially when the sentences may contain named entities numeric or other data for which no embedding is available.

(4) technical

Explaining how a theory, principle, basic concept, etc can be implement (HOW indicators).

Example:

Clustering performance is measured on the Indo-European task according to the Rand Index, F-score, Normalized Edit Score <citation> and Normalized Variation of Information (Meila, 2003).

Background

(5) trend (significance)

This label is used when the citing sentence explains the significance of the research or trends of theory/principle/concept/topic/ problem, etc.

Example:

The recognition of Sanskrit as a highly phonetic language as also one with an extensively codified grammar <citation>, is widespread.

Citing Paper Work

(1) citing paper propose (acknowledge)

This label is used to describe what research is done by citing papers that acknowledge cited papers' work.

Example:

In this paper, we first characterize sufficient conditions for Borel-measurability of the value functions for the terminal time problem (characterized so far only for the first hitting time problem <citation>).

(2) based-on / inspire

Citing paper is built on inspiration, consideration, following, or based-on cited papers on certain aspects.

Example:

In particular, we consider the work inspired by evolution-based genetic algorithms, and the IQBE paradigm of Smith and Smith <citation> and <citation>.

(3) use

Citing paper explain that cited work is used, implemented, employed, or adopted in term of concept, dataset, method, etc. without modification, improvement, or extention.

Example:

We use 3500 sentences from CoNLL <citation> as the NER data and section 20-23 of the WSJ <citation> as the POS/chunk data (8936 sentences).

(4) extend

Citing paper extend, improve, add, or modify the work on cited paper.

Example:

We extended the abstract frameworks proposed in <citation> for describing native and SAT based ASP algorithms to capture such sophisticated features as backjumping and learning.

Citing Paper Work

(5) citing paper dominant

This scheme is used to represent the positive aspect of the proposed method or result from citing paper. The main indicators are the superiority of citing paper compared to cited paper.

Example:

Our PredNet model outperforms the model by Brabandere et al. <citation> by 29%.

(6) citing paper future

This label accommodates citing sentences that plan future research of the author.

Example:

As future work, we also plan to investigate the use of more sophisticated neural net generators, similar to DCGAN's <citation> and to extend the approach to a conditional generator applicable to structured output problems.

Cited Paper Work

(1) cited paper propose

Describing the proposed research by the cited work.

Example:

Hamm et al. <citation> proposed the use of knowledge transfer between a collection of models trained on individual devices into a single model guaranteeing differential privacy.

(2) cited paper success

Highlighting the success of cited paper.

Example:

Abadi et al. <citation> previously obtained 97% accuracy with a (8, 10−5) bound on MNIST, starting from an inferior baseline model without privacy.

(3) cited paper weakness

Highlighting the weakness of the cited papers (research gap).

Example:

However, its lack of randomization gives rise to caveats <citation>, and attackers can infer properties of the dataset <citation>.

(4) cited paper result

This label is used when citing sentences to describe the result of the cited paper (neutral, not success nor weakness).

Example:

Shokri & Shmatikov <citation> reported about 92% accuracy on SVHN with ε > 2 per model parameter and a model with over 300,000 parameters.

Cited Paper Work

(5) cited paper dominant

This scheme is used to state the superiority of cited paper compared to citing paper.

Example:

Oquab et al <citation> outperform our method (ACD-A) in mAP for 2 reasons: a) They use fully-labelled data and don't encounter domain differences: they finetune CNN on trainval set of PASCAL VOC 2012 action classification

Compare and Contrast

(1) similar

Describing what is the same between citing paper and cited paper.

Example:

Chang et al. <citation> also propose a “semisupervised” learning approach quite similar to our own model.

(2) contrast

Describing the differences or contradiction between citing paper and cited paper.

Example:

Our model is simple in design and only uses game states from the AI side, rather than using opponent's’ information <citation>.

Other Category

Other

(1) multiple intent

Describing the citing sentences that have two or more citation marks for different purposes.

Example:

SUTime <citation> designs fixed rules using a cascade finite automata <citation> on regular expressions over tokens <citation>.

(2) cited paper comparison

Comparison is done between cited papers. In this case, it doesn't really matter where the similarities or differences between them.

Example:

One employs the implication graph <citation> and the other employs resolution <citation>.

(3) other

This label is designed for citing sentences that do not meet all of the label categories described above (both core-categories and other-categories labels).

Example:

C++ in ILOG Solver <citation> or Java in Gecode/J <citation>) and even term rewriting <citation>.

Useful Keywords, Phrases, and sentence patterns

Background

(1) definition

Keywords and Phrases:

… is …
… are …
… can be seen as …
… can be defined as …
… is called …
… known as …
… is considered as …
… can be understood as …
etc.

(2) suggest

Keywords and Phrases:

…. Given in ...
… introduced in …
… can be found in …
… more comprehensive … can be found in …
… has been taken from … if further detail is required.
… for details refer to …
… more discussions … refer to …
For an overview … see …
… recent/current update …

(3) judgement

Keywords and Phrases:

… great benefits …
… is not sufficient …
… an important role …
… is not that effective …
… useful …
… inappropriate …
… more easily …
… assume …
etc.

(4) technical

Keywords and Phrases:

… is trained on …
… could be achieved by using …
… is measured …
… be generated by …
… are added to …
… to map …

Background

5. Trend (significance)

Keywords and Phrases:

… is widespread …
… recently been a growing concern …
… has attracted considerable attention since last decade …
lot of attention
… recent/current trend …
Popular / popularity
In recent years …
gain more attention
… gain interest …
.. well-known / best-known …
etc.

Citing Paper Work

(1) citing paper propose

Keywords and Phrases:

… In our study, we analyze …
We compare SynTime …
… in this paper we will focus …
We also are interested in …
Our framework combines …
etc.

(2) based-on / inspire

Keywords and Phrases:

… based on ….
… we followed …
… we consider the work inspired by …
Our work here builds on previous work …
We refer to
Inspired by
.. as in ..
etc.

(3) use

Keywords and Phrases:

In this work, we use …
Our model is implemented with …
In this paper, we employ …
… employed in our work …
We adopt …
we use the method …
We apply … <>
We compare …
etc.

(4) extend

Keywords and Phrases:

… we extended this framework …
… this paper improves …
… we explore another potential principle …
Modify
Add
combine
Adapt (small change)
etc.

Citing Paper Work

(5) citing paper dominant

Keywords and Phrases:

… our baseline model … effective reproducing the results …

… our PredNet model outperforms …

Unlike the previous works … our model could …

etc.

(6). Future work

Keywords and Phrases:

… As future work, we also plan to …
… in the future …
… as a future work …
… in our future work …
… as for future work …
… future direction …

… [keywords] + we hope / we plan / we are going to / we expect / we will / etc …

Cited Paper Work

(1) cited paper propose

Keywords and Phrases:

<citation> … propose …
<citation> … introduced …
<citation> … use …
<citation> … described …
… introduced by/in <citation>
… given in ….
etc.

(2) cited paper success

Keywords and Phrases:

… has established … as a strong standard …
… Arnold won the second places …
… demonstrated to greatly improve …
Achievement
Good result
etc.

(3) cited paper weakness

Keywords and Phrases:

… <citation> … give low standard of …
… lack of randomization given by <citation> …
… we still expect above-chance decoding performance of … <citation> …
… low performance …
… drawback …
… limitation …

etc.

(4) cited paper result

Keywords and Phrases:

… [citation mark] reported about 92% accuracy …
… have presented ..
… recent results <citation> have shown results …
… has been shown by …
… shown in …
… as reported in …
.. has/have been obtained …
… as observed in ..
...presented in ...
… illustrated in ….

Cited Paper Work

(5) cited paper dominant

Keywords and Phrases:

<citation> outperforms our …

<citation> show better performance than our method ...

Compare and Contrast

1. compare

Keywords and Phrases:

… is closely related to our work.
… our work is similar to …
… the closest to the approach we develop here.
… for the same subset …
closely related to
etc.

2. contrast

Keywords and Phrases:

… unlike the methods of … our techniques …
… Here we take a different approach …
We choose … do not use …
In contrast to …
etc.

Other

1. Multiple intent

Keywords and Phrases:

…

2. Cited paper comparison

Keywords and Phrases:

…

Note for Labeling and Annotation (1):

1. After read a Citing Sentence, please consider the Coarse Label in the Core Category first.
    a. Concepts or Theory: **BACKGROUND**
    b. Author's Work: **CITING PAPER WORK**
    c. Other's Work: **CITED PAPER WORK**
    d. Comparison between Author and Other Works: **COMPARE AND CONTRAST**
    e. Other: **OTHER**
2. After that, please decide the Fine-grained Label according to its intent.

Labeling and Annotation Practice(3)

1. If you find the difficulties to label the Citing Sentence **“Please mark the coarse label only”** and you can continue to label other citing sentences and back to the difficult citing sentences **later**.

Thank You