BCMI-NLP Labeled-Alignment-Based Entailment System for NTCIR-10 RITE-2 Task

Xiao-Lin Wang, Hai Zhao, and Bao-Liang Lu
Center for Brain-like Computing and Machine Intelligence
MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligence Systems
Department of Computer Science and Engineering,
Shanghai Jiao Tong University
arthur.xl.wang@gmail.com {zhaohai,blu}@cs.sjtu.edu.cn

Overview

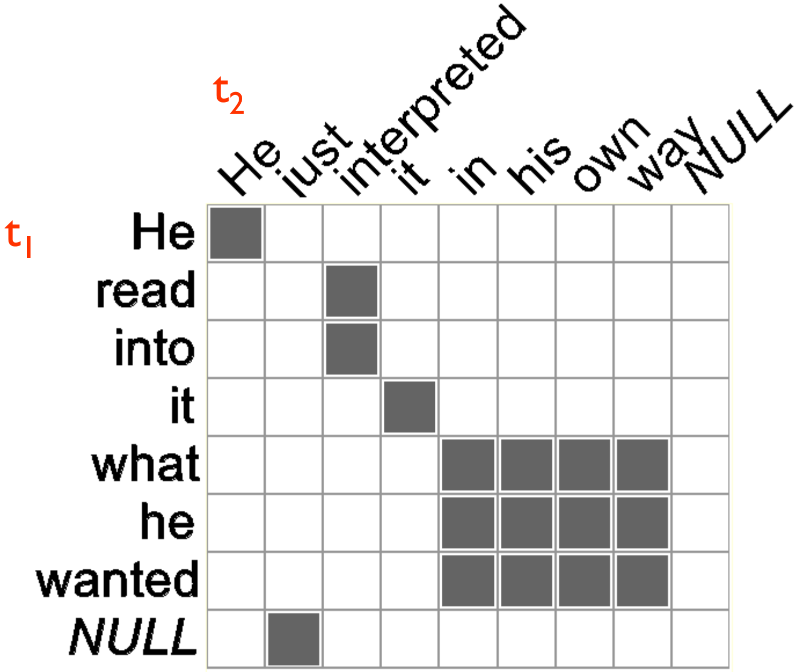
- Alignment-based RITE approach
- Labeled alignment scheme
- Labeled-alignment-based RITE approach
- Formal runs

Alignment-based RITE approach

- Intuition: sufficiently good alignment

  → close lexical and structural similarity

  → entailment relation

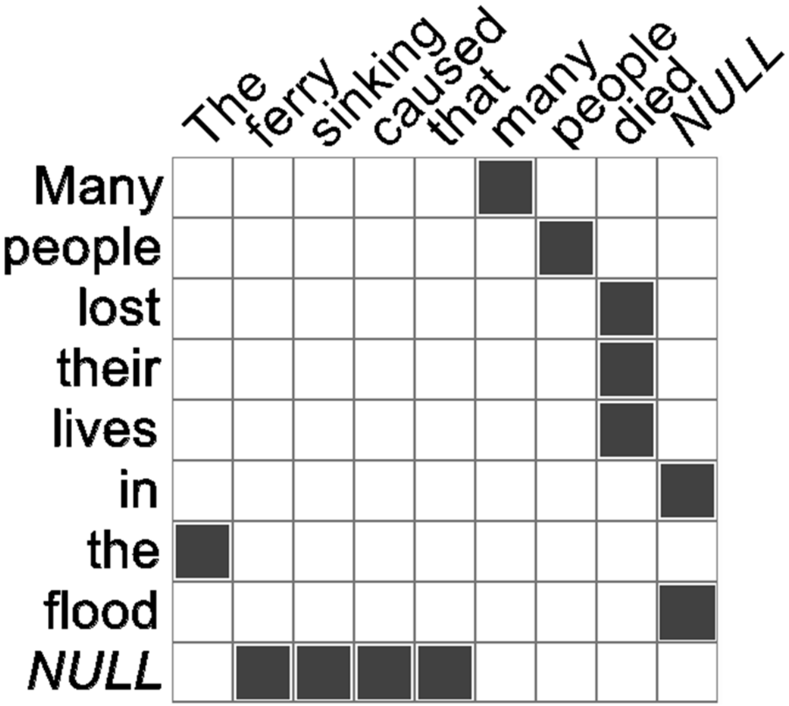


read into → interpreted
what he wanted → in his own way
NULL → just

Therefore, entailment

Alignment-based RITE approach

- Shortage:
  - originally developed for machine translation
  - not justify non-entailment pairs ☹



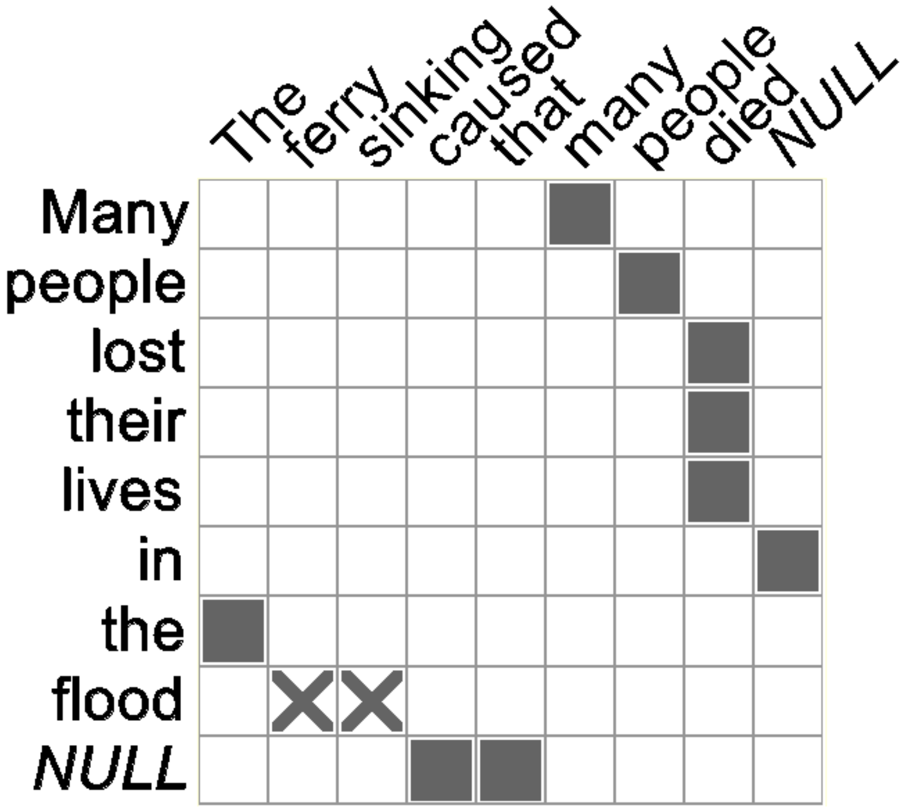
Critical non-aligned:
flood → NULL
NULL → ferry sinking

Non-critical non-aligned:
in → NULL
NULL → caused that

mixed, so hard to predict

Labeled alignment scheme

- Augment the normal alignment with negative links



Negative links:

flood ↛ ferry sinking

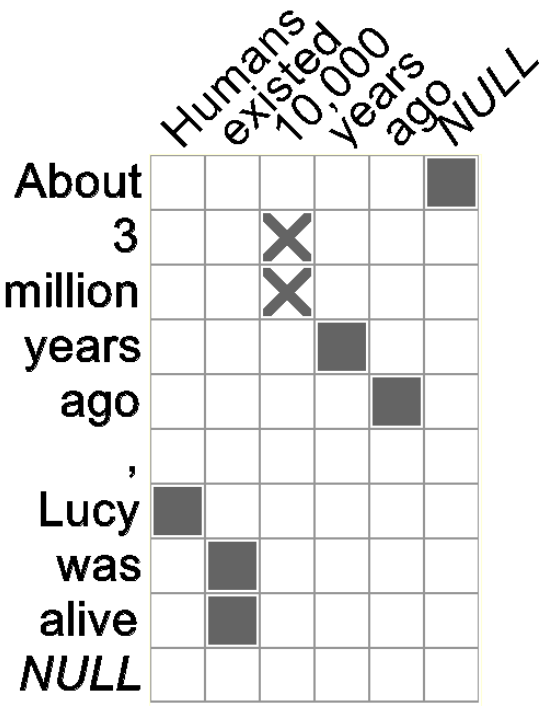
Therefore, non-entailment

Roadmap:

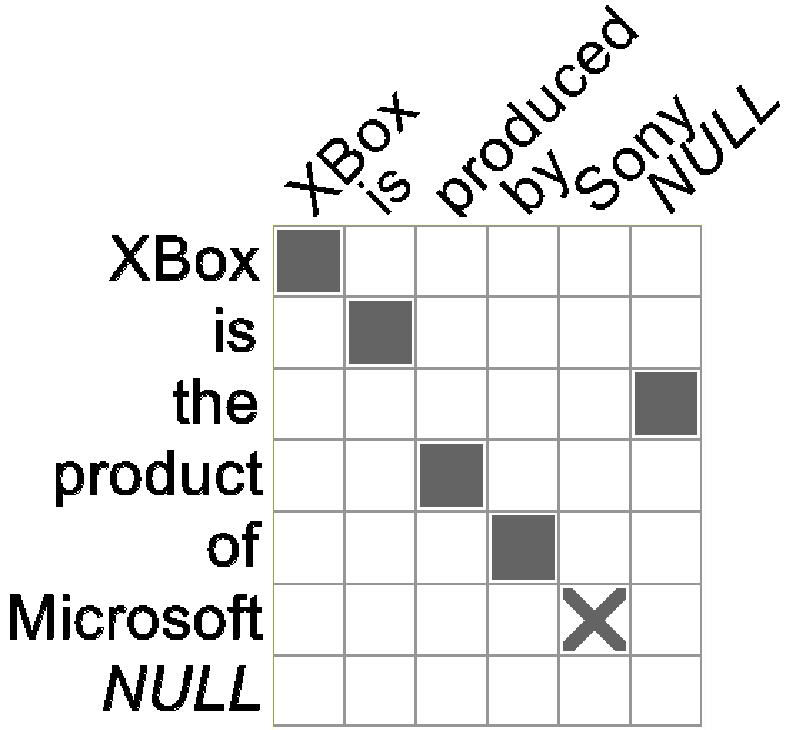
1) manually annotate
2) train an alignment model
3) train an RITE predict model

Labeled alignment scheme

- More examples



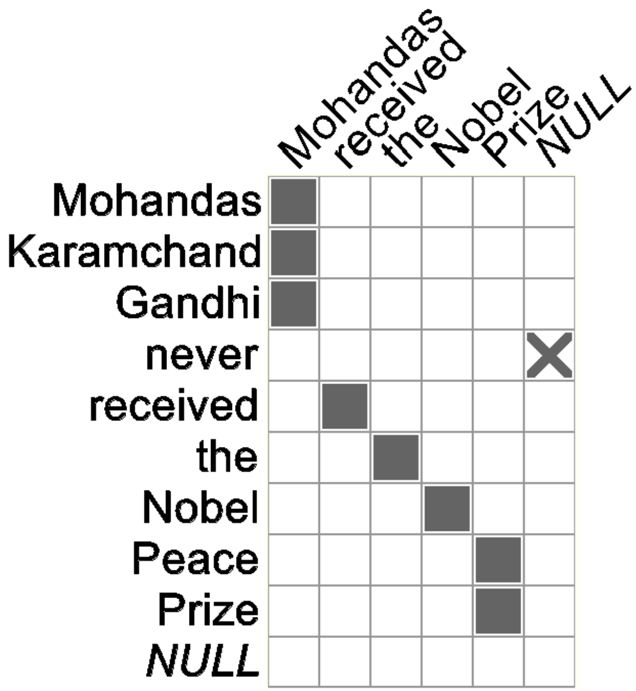
Mismatched numbers



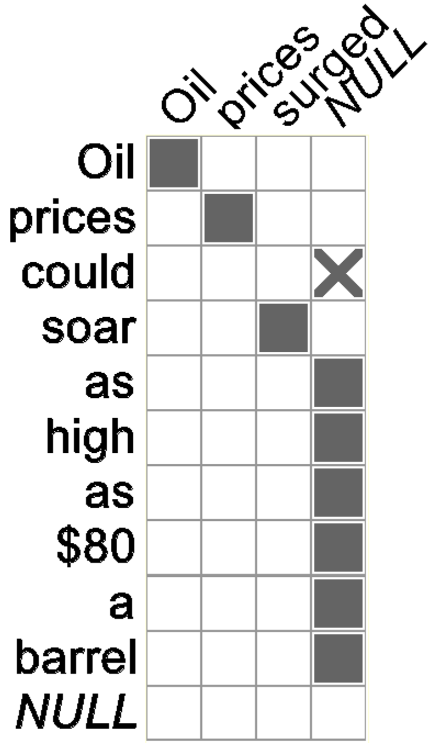
Different named entities

Labeled alignment scheme

- More examples

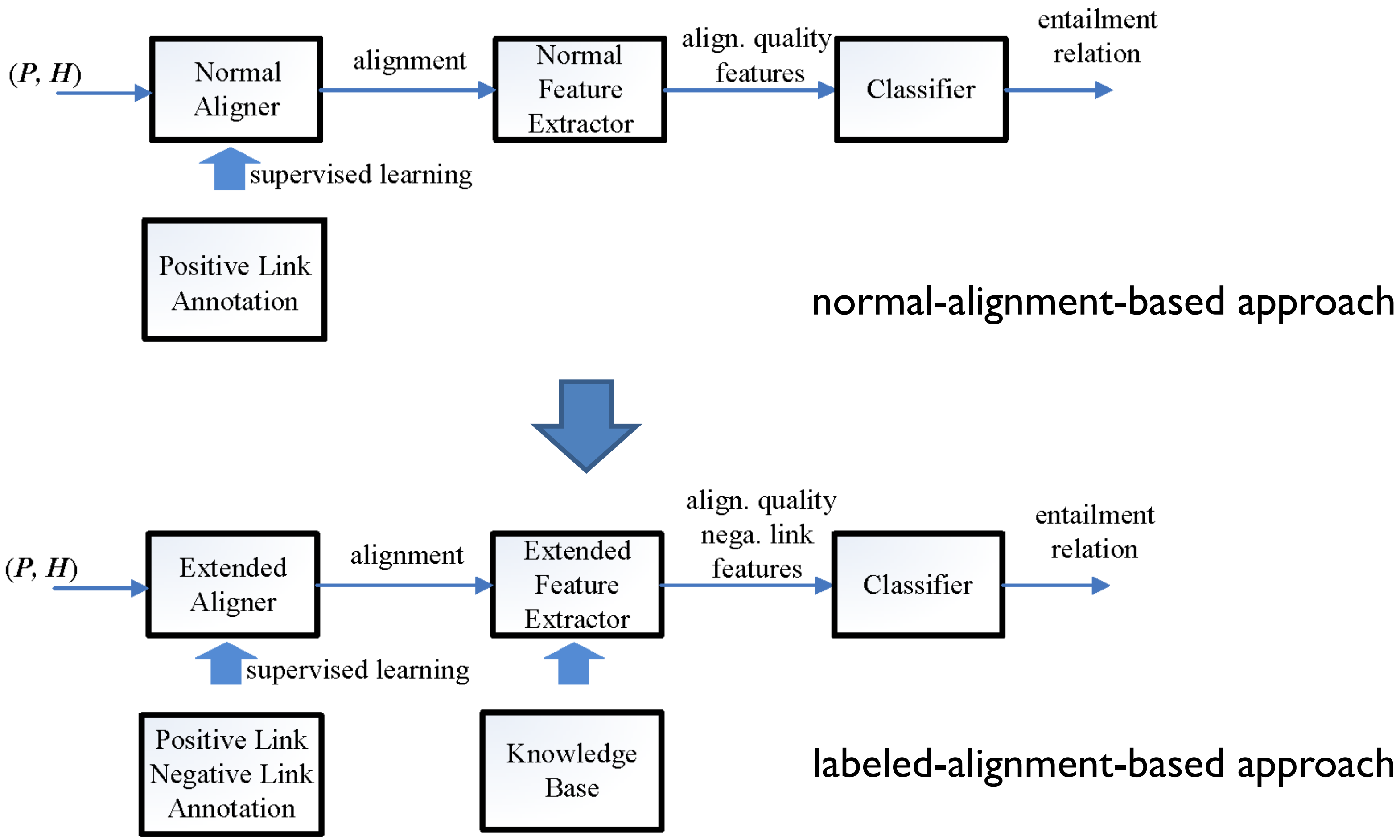


Negation



Non-happened event

Labeled-alignment-based RITE approach



Labeled-alignment-based RITE

Link Type Features

Whether e_1 and e_2 are in an antonym list
Whether e_1 and e_2 are in an synonym list
Whether e_1 and e_2 are unequal numbers
Whether e_1 and e_2 are different named entities
Relation of e_1 and e_2 in an ontology (hyponym, sibling, etc.)
Ontology-based similarities of e_1 and e_2
Count of common characters
Length of the common prefixes
Length of the common suffix
Tuple of the syntactic tags
Tuple of the ancestors in an ontology
Tuple of whether e_1 or e_2 is in a list of negative expressions
Tuple of whether e_1 or e_2 is the head of a noun phrase

Labeled-alignment-based RITE

- Classification component
  - Sample representation: single flat vector that combines the features extracted from all the links
  - Classifier: RBF-kerneled SVM (1-vs-rest for MC)
- Open question: why a cascaded system that

  - first classifies the link type

  - then classifies the RITE relation

  performs poorly ? (worse than baseline)

Labeled–alignment-based RITE

- Possible answer for the failure of a cascaded system : not robust enough under current (public) available resources

For example:

三峡旅游不再有淡旺季的问题

三峡旅游不再有枯水期的问题

- Hard to accurately decide whether two normal Chinese words are antonym or not with the knowledge base of CiLin and mini-Hownet.

Formal Run

- System Implement
  - automated supervised-learning aligner

    (N. chambers et al.,2007; B. MacCartny et al., 2008)

  - Manually annotate the training data set
    - alignment links
    - positive/negative labels

Formal Run -- aligner

Structured prediction
MIRA training process

Inputs

- training problems $\langle P_j, H_j \rangle, j = 1..n$
- corresponding gold-standard alignments E_j
- a number of learning epochs N (e.g. 50)
- a "burn-in" period $N_0 < N$ (e.g. 10)
- initial learning rate R_0 (e.g. 1) and multiplier r (e.g. 0.8)
- a vector of feature functions $\mathbf{\Phi}(E)$
- an alignment algorithm $\text{ALIGN}(P, H; \mathbf{w})$ which finds a good alignment for $\langle P, H \rangle$ using weight vector $\mathbf{w}$

Initialize

- Set $\mathbf{w} = 0$

Repeat for $i = 1$ to N

- Set $R_i = r \cdot R_{i-1}$, reducing the learning rate
- Randomly shuffle the training problems
- For $j = 1$ to n:
  - Set $\hat{E}_j = \text{ALIGN}(P_j, H_j; \mathbf{w})$
  - Set $\mathbf{w} = \mathbf{w} + R_i \cdot (\mathbf{\Phi}(E_j) - \mathbf{\Phi}(\hat{E}_j))$
- Set $\mathbf{w} = \mathbf{w} / \|\mathbf{w}\|_2$ (L2 normalization)
- Set $\mathbf{w}[i] = \mathbf{w}$, storing the weight vector for this epoch

Return an averaged weight vector:

- $\mathbf{w}_{avg} = 1/(N - N_0) \sum_{i=N_0+1}^{N} \mathbf{w}[i]$

Formal Run – aligner (decoding)

Inputs

- an alignment problem $\langle P, H \rangle$
- a number of iterations N (e.g. 100)
- initial temperature T_0 (e.g. 40) and multiplier r (e.g. 0.9)
- a bound on edit size max (e.g. 6)
- an alignment scoring function, $\textsc{score}(E)$

Initialize

- Let E be an "empty" alignment for $\langle P, H \rangle$ (containing only DEL and INS edits, no EQ or SUB edits)
- Set $\hat{E} = E$

Repeat for $i = 1$ to N

- Let $\{F_1, F_2, ...\}$ be the set of possible successors of E. To generate this set:
  - Consider every possible edit f up to size max
  - Let $C(E, f)$ be the set of edits in E which "conflict" with f (i.e., involve at least some of the same tokens as f)
  - Let $F = E \cup \{f\} \setminus C(E, f)$
- Let $s(F)$ be a map from successors of E to scores generated by $\textsc{score}$
- Set $p(F) = \exp s(F)$, and then normalize $p(F)$, transforming the score map to a probability distribution
- Set $T_i = r \cdot T_{i-1}$
- Set $p(F) = p(F)^{1/T_i}$, smoothing or sharpening $p(F)$
- Renormalize $p(F)$
- Choose a new value for E by sampling from $p(F)$
- If $\textsc{score}(E) > \textsc{score}(\hat{E})$, set $\hat{E} = E$

Return $\hat{E}$

Stimulated annealing algorithm

"Possible edit" means
"possible alignment link"

Generated from:

1) Each segmented word in t_1 → Each segmented word in t_2
2) Each syntactic node in t_1 → Each syntactic node in t_2
3) Each NE in t_1 → Each NE in t_2
4) Expression e_1 in t_1 → Expression e_2 in t_2 where,

(e_1, e_2) appears in a synonym list,
or a antonym list,
or in Cilin,
or in hownet,
or other resources

Formal Run

- System Implement
  - Segment: BaseSeg (Zhao and Kit, 2006)
  - Syntactic parser: Stanford factored parser
  - NER: BaseNER (zhao and Kit, 2008)
  - Chinese ontologies: Cilin (Mei et al.,1983; Luo, 2007), Hownet (Dong, 2003; Liu and Li, 2002; Liu and Singh, 2004)
  - Classifier: LibSVM (Lin, 2011)

Formal Run

Run	Method	Macro-F1 on BC	Macro-F1 on MC	Worse. Rank. on RITE4QA
Run01	Char-overlap	67.04	39.95	2.67*
Run02	Normal-align.	66.89	44.88	0.00*
Run03	Labeled-align.	73.84	56.82	8.00*

* not checked during formal run due to limit of time

References

- N. Chambers, D. Cer, T. Grenager, D. Hall, C. Kiddon, B. MacCartney, M.-C. de Marneffe, D. Ramage, E. Yeh, and C. D. Manning. “Learning alignments and leveraging natural logic.” In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pages 165--170. Association for Computational Linguistics, 2007.
- C. C. Chang and C. J. Lin. “Libsvm: a library for support vector machines.” ACM Transactions on Intelligent Systems and Technology (TIST), 2(3):27, 2011.
- Z. D. Dong and Q. Dong. “ Hownet-a hybrid language and knowledge resource.” In Proceedings of International Conference on Natural Language Processing and Knowledge Engineering, pages 820--824. IEEE, 2003.
- D. Klein and C. D. Manning. “Fast exact inference with a factored model for natural language parsing.”Advances in neural information processing systems, 15(2003):3--10, 2002.
- R. Levy and C. Manning.“Is it harder to parse Chinese, or the Chinese Treebank”. In Proceedings of ACL, volume 3, pages 439--446, 2003.H. Liu and P. Singh. “Conceptnet -- a practical commonsense reasoning tool-kit." BT technology journal, 22(4):211--226, 2004.

References

- Q. Liu and S. J. Li."Computation of semantical similarity for phrases based on HowNet (in Chinese)." Chinese Computational Linguistics, 7(2):59--76, 2002.Z. C. Luo. "Improvements on TongYiCi CiLin." http://blog.csdn.net/ganlantree/article/details/1845788, 2007. [accessed 10-Jan-2013].
- B. MacCartney, M. Galley, and C. D. Manning." A phrase-based alignment model for natural language inference." In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pages 802--811. Association for Computational Linguistics, 2008.
- J. J. Mei, Y. M. Zhu, and Y. Q. Gao. TongYiCi CiLin. Shanghai Dictionary Publisher, 1983.
- Y. Watanabe, Y. Miyao, J. Mizuno, T. Shibata, H. Kanayama, C.-W. Lee, C.-J. Lin, S. Shi, T. Mitamura, N. Kando, H. Shima, and K. Takeda.Overview of the Recognizing Inference in Text ({RITE-2}) at the {NTCIR}-10 Workshop. In Proceedings of NTCIR-10 Workshop Meeting, 2013.

References

- T. Xia.Research on the computation of semantical similarity for Chinese phrases (in Chinese). Computer Engineering}, 33(6):191--194, 2007.
- H. Zhao, C. N. Huang, and M. Li. "An improved Chinese word segmentation system with conditional random field." In Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing, pages 162--165. Sydney: July, 2006.
- H. Zhao and C. Y. Kit." Unsupervised segmentation helps supervised learning of character tagging for word segmentation and named entity recognition." In Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing, pages 106--111, 2008.
- Y. H. Zhu, H. Q. Hou, and Y. T. Sha."Comparison and evaluation of two algorithms for recognizing Chinese synonyms (in Chinese) ". Journal of Library Science in China, 28(4):82--85, 2002.

Thank you