Table 1: The comparative evaluation results of medical terminology assignment in terms of $S@K$ and $P@K$.

Apporach \ Metric	S@1	S@2	S@3	S@4	P@1	P@2	P@3	P@4
LocalMining	72.0%	84.0%	91.0%	95.0%	72.0%	72.1%	69.7%	68.3%
Local+Global	**83.0%**	**92.0%**	**98.0%**	**100.0%**	**83.0%**	**81.5%**	**80.3%**	**78.8%**

Table 2: Comparative illustration of the representative question samples with locally mined terminologies and locally+globally recommended terminologies. Answers are not displayed due to limited space.

QA pairs	Locally Mined Terminologies	Local Mining + Global Learning
Is it safe to color my hair during pregnancy ?	hair structure, dyed hair, feeling safe, patient currently pregnant, first trimester pregnancy...	hair structure, patient currently pregnant, coal tar allergy, hair color change, disorder of endocrine system...
If I get an infection caused by gum disease, can that be transferred to my fetus ?	infectious disease, gingival disease, entire fetus, inflammation, periodontal disease...	infectious disease, prematurity of fetus, gingival disease, periodontal disease low birth weight infant...

HealthTap, which involve 5,958 unique doctors. For ground truth construction, we invited three professionals with master degrees majored in medicine programme. The labelers were trained with a short tutorial and a set of demonstrating examples. A majority voting scheme among the three labelers can partially alleviate the subjectivity problem. The annotators were required to label only top five recommended terminologies for each QA pair, and they were labeled either as "positive" or "negative". 100 QA pairs were labeled as testing set.

We adopted two metrics that are able to characterize precisions from different aspects. The first is average $S@K$ over all testing QA pairs, which measures the probability of finding a relevant terminology among the top K recommended ones. To be specific, for each testing QA pair, $S@K$ is assigned to 1 if a relevant terminology is positioned in the top K and 0 otherwise. The second one is average $P@K$ that stands for the proportion of recommended terminologies that are relevant[20]. $P@K$ is defined as $P@K = \frac{|\mathcal{C} \cap \mathcal{R}|}{|\mathcal{C}|}$ where $\mathcal{C}$ is a set of the top K terminologies, and $\mathcal{R}$ is the manually labeled positive ones.

Table 1 displays the comparison. We can see that the local mining approach achieves the worst performance. This is reasonable, because irrelevant concepts may be mapped to terminologies because of their presence in the QA pairs.

Table 2 comparatively illustrates the representative QA pair samples with locally minded terminologies and locally+globally recommended ones. Intuitively, the terminologies are more comprehensive and reliable after enhancement with global learning.

5. CONCLUSIONS AND FUTURE WORK

This paper presented a medical terminology assignment scheme to bridge the vocabulary gap between health seekers and community generated knowledge. A strong unified framework of local mining and global learning is proposed to tackle this research issue, instead of the conventional isolated utilization. It proposes the concept entropy impurity approach to comparatively detect and normalize the medical concepts locally, which naturally construct a corpus-aware terminology vocabulary with the help of external knowledge. In addition, it builds a novel global learning model to enhance the local coding results. This model seamlessly integrates various heterogeneous cues.

In the future, we will investigate how to flexibly organize the unstructured medical content into user needs-aware ontology by the recommended medical terminologies.

6. ACKNOWLEDGEMENTS

This work was supported by NUS-Tsinghua Extreme Search project under the grant number: R-252-300-001-490.

7. REFERENCES

[1] R. L. Cilibrasi and P. M. B. Vitanyi. The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, 2007.

[2] C. Dozier, R. Kondadadi, K. Al-Kofahi, M. Chaudhary, and X. Guo. Fast tagging of medical terms in legal text. In *Proceedings of the International Conference on Artificial Intelligence and Law*, 2007.

[3] A. e-HIM Work Group on Computer-Assisted Coding. Delving into computer-assisted coding. *Journal of American Health Information Management Association*, 2004.

[4] S. Fox and M. Duggan. Health online 2013. Survey, Pew Research Center, 2013.

[5] Y. Huang, Q. Liu, S. Zhang, and D. Metaxas. Image retrieval via probabilistic hypergraph ranking. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[6] M.-Y. Kim and R. Goebel. Detection and normalization of medical terms using domain-specific term frequency and adaptive ranking. In *Information Technology and Applications in Biomedicine, IEEE International Conference on*, 2010.

[7] L. S. Larkey and W. B. Croft. Automatic assignment of icd9 codes to discharge summaries. *PhD Thesis, University of Massachusetts at Amherst*, 1995.

[8] M. Law. Online drug information in canada. Technical report, 2012.

[9] G. Leroy and H. Chen. Meeting medical terminology needs-the ontology-enhanced medical concept mapper. *IEEE Transactions on Information Technology in Biomedicine*, 2001.

[10] L. V. Lita, S. Yu, S. Niculescu, and J. Bi. Large scale diagnostic code classification for medical patient records. In *Proceedings of the Conference on Artificial Intelligence in Medicine*, 1995.

[11] L. Nie, T. Li, M. Akbari, J. Shen, and T.-S. Chua. Wenzher: Comprehensive vertical search for healthcare domain. In *Proceedings of the International ACM SIGIR Conference*, 2014.

[12] L. Nie, M. Wang, Y. Gao, Z.-J. Zha, and T.-S. Chua. Beyond text qa: Multimedia answer generation by harvesting web information. *IEEE Transactions on Multimedia*, 2013.

[13] L. Nie, M. Wang, Z.-J. Zha, and T.-S. Chua. Oracle in image search: A content-based approach to performance prediction. *ACM Transactions on Information System*, 2012.

[14] L. Nie, M. Wang, Z.-J. Zha, G. Li, and T.-S. Chua. Multimedia answering: Enriching text qa with media information. In *Proceedings of the International ACM SIGIR Conference*, 2011.

[15] L. Nie, S. Yan, M. Wang, R. Hong, and T.-S. Chua. Harvesting visual concepts for image search with complex queries. In *Proceedings of the International Conference on Multimedia*, 2012.

[16] L. Nie, Y.-L. Zhao, X. Wang, J. Shen, and T.-S. Chua. Learning to recommend descriptive tags for questions in social forums. *ACM Transactions on Information System*, 2014.

[17] H. Suominen, F. Ginter, S. Pyysalo, A. Airola, T. Pahikkala, S. Salanterä, and T. Salakoski. Machine learning to automate the assignment of diagnosis codes to free-text radiology reports: a method description. In *Proceedings of the ICML Workshop on Machine Learning for Health-Care Applications*, 2008.

[18] P. Velardi, M. Missikoff, and R. Basili. Identification of relevant terms to support the construction of domain ontologies. In *Proceedings of the workshop on Human Language Technology and Knowledge Management*, 2001.

[19] L. Yves A., S. Lyudmila, and F. Carol. Automating icd-9-cm encoding using medical language processing: A feasibility study. In *Proceedings of the AMIA Annual Symposium*, 2000.

[20] Y.-L. Zhao, L. Nie, X. Wang, and T.-S. Chua. Personalized recommendations of locally interesting venues to tourists via cross region community matching. *ACM Transactions on Intelligent Systems and Technology*, 2013.

[21] G. Zuccon, B. Koopman, A. Nguyen, D. Vickers, and L. Butt. Exploiting medical hierarchies for concept-based information retrieval. In *Proceedings of the Seventeenth Australasian Document Computing Symposium*, 2012.