
AAN: Developing Educational Tools for Work Force Training

Alexander R. Fabbri[†] Irene Li[†] Swapnil Hingmire[‡] Dragomir Radev[†]
[[†]]Yale University [[‡]]Tata Consultancy Services
{alexander.fabbri,irene.li,dragomir.radev}@yale.edu, swapnil.hingmire@tcs.com

Abstract

Millions of computing jobs remain unfilled, largely due to an insufficiently trained workforce which cannot keep up with the latest technological advances. We aim to address this problem by developing techniques for creating, visualizing and accessing educational resources for work force training and retraining. Our focus is on Computing, and we have chosen Natural Language Processing (NLP) as an initial field of study for which we are developing initial tools and building upon to apply to new domains. The platform on which the current system operates is **AAN** (“All About NLP”)¹. Below we will describe our previous and ongoing work as well as planned expansion to new domains like computer vision and epidemiology.

1 Overview

AAN started from an initial taxonomy for Artificial Intelligence (AI)-related topics from university-level courses and the manual collection of relevant tutorials, as opposed to related paper-based corpora [1]. We expanded our collection to include surveys, course lectures and online tutorials. Notably, all resources were manually curated for quality purposes and associated with a particular taxonomy topic and medium. We released this corpus of tutorials and associated annotations in our paper TutorialBank [2]. The corpus initially included 6,000 resources on NLP and AI-related topics but has expanded to include nearly 14,000 resources.

We have also worked on developing tools for keyword and concept extraction from lectures, tutorials and surveys as well as the research problem of automatically creating prerequisite chains among these concepts. Such a tool is essential for work force training as employees can determine what they know and what they need to focus on to be prepared for a particular job. In our paper LectureBank [3], we introduced a dataset containing English lecture files collected from university courses which are each classified according to the AAN taxonomy as well as manually-labeled prerequisite relation topics, which is publicly available. In subsequent work, we expanded this dataset to include 1,717 lecture slide files and 322 topics. We have also developed tools for determining prerequisite relationships among topics, in particular an unsupervised graph-based method which we plan to use when applying to a new domain where labeled data may be scarce or non-existent [4].

One of our current focuses is generating educational overviews of scientific topics to provide up-to-date surveys of a given topic in a prerequisite chain. We are also working on adapting our current tools developed for NLP to other domains such as Computer Vision. Initially, components for resource collection and classification as well as taxonomy creation and prerequisite learning were all performed manually. As we expand to new domains we aim to automate more and more of the pipeline until we can automatically create such components for any domain. Finally, we are constantly updating our AAN platform by adding new resources and features. Current functionalities include keyword and taxonomy search and bookmarking as well as resource recommendation from a title-abstract pair. We are working on a personalized topic visualization as well as more personalized search functionalities.

¹<http://aan.how>

References

- [1] Radev, D. R., Muthukrishnan, P., Qazvinian, V., & Abu-Jbara, A. (2013). The ACL anthology network corpus. *Language Resources and Evaluation*, 47(4), 919-944.
- [2] Fabbri, A. R., Li, I., Trairatvorakul, P., He, Y., Ting, W. T., Tung, R., Westerfield, C., & Radev, D. R. (2018). Tutorialbank: A Manually-collected Corpus for Prerequisite Chains, Survey Extraction and Resource Recommendation. *ACL 2018*.
- [3] Li, I., Fabbri, A. R., Tung, R. R., & Radev, D. R. (2019, July). What should i learn first: Introducing Lecturebank for NLP Education and Prerequisite Chain Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 6674-6681).
- [4] Li, I., Fabbri, A., Hingmire, S., Radev, D. (2020). R-VGAE: Relational-variational Graph Autoencoder for Unsupervised Prerequisite Chain Learning. *arXiv preprint arXiv:2004.10610*.