|  |  |
| --- | --- |
| Title | Fonduer: Knowledge Base Construction from Richly Formatted Data |
| Author | Sen Wu, Luke Hsiao, Xiao Cheng, Braden Hancock... |
| URL | https://sing.stanford.edu/site/publications/fonduer-sigmod18.pdf |
| Content | - Knowledge base construction(KBC)  - KBC from richly formatted data aims to extract relations conveyed jointly via textual, structural, tabular and visual expressions  - “Fonduer” is a machine-learning-based KBC system for richly formatted data, which presents a new data model that accounts for 3 challenging characteristics of richly formatted data:   1. Prevalent document-level relations 2. Multimodality (多模态） 3. Data variety   Fonduer uses a new deep-learning model to automatically capture the representation needed to learn how to extract relations from richly formatted KBC  - Advantages:   1. Fonduer achieves an average improvement of 41 F1 points on the quality of the output knowledge base 2. Easy to use: after using Fonduer’s new programming model, non-domain experts are able to design KBC systems that achieve on average 23 F1 points higher quality than traditional machine-learning-based KBC approaches   - contribution:  1.design a new data model that preserves structural and semantic  information across different data modalities  2.show that existing deep-learning models tailored for text information extraction   1. Fonduer introduces a programming model in which no development cycles are spent on feature engineering |
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|