

Cold Start Knowledge Base Population at 2014

- Knowledge Base Population (KBP) track of TAC encourages the development of systems that can match entities mentioned in natural texts with those appearing in a knowledge base and extract novel information about entities from a document collection and add it to a new or existing knowledge base.

Cold Start Knowledge Base Population at 2014

- Knowledge Base Population (KBP) track of TAC encourages the development of systems that can match entities mentioned in natural texts with those appearing in a knowledge base and extract novel information about entities from a document collection and add it to a new or existing knowledge base.
- Some example relations:
 - children of
 - city of birth
 - shareholders
 - countries of residence

- Modeled the problem using distant supervision
- Used Freebase as an existing Knowledge base.
- **Freebase:** Freebase is a large collaborative knowledge base consisting of metadata composed mainly by its community members.
- It is an online collection of structured data harvested from many sources, including individual, user-submitted wiki contributions.

- The TAC corpus consisted of three type of documents:

- The TAC corpus consisted of three type of documents:
 - **discussion forums** 99,063 English discussion forum documents selected from the BOLT Phase 1 discussion forums source data releases. Each forum includes at least 5 posts.

- The TAC corpus consisted of three type of documents:
 - **discussion forums** 99,063 English discussion forum documents selected from the BOLT Phase 1 discussion forums source data releases. Each forum includes at least 5 posts.
 - **newswire** 1,000,257 documents selected from English Gigaword Fifth Edition.

- The TAC corpus consisted of three type of documents:
 - **discussion forums** 99,063 English discussion forum documents selected from the BOLT Phase 1 discussion forums source data releases. Each forum includes at least 5 posts.
 - **newswire** 1,000,257 documents selected from English Gigaword Fifth Edition.
 - **web** 999,999 English web documents selected from various GALE web collections.

- The TAC corpus consisted of three type of documents:
 - **discussion forums** 99,063 English discussion forum documents selected from the BOLT Phase 1 discussion forums source data releases. Each forum includes at least 5 posts.
 - **newswire** 1,000,257 documents selected from English Gigaword Fifth Edition.
 - **web** 999,999 English web documents selected from various GALE web collections.
- We have submitted the knowledge base populated using our techniques on the above corpus and waiting for results.