Here is something for named entity recognition (NER):

1. CoNLL-2003 shared task. Shared task often provides data to play around. Reference: Introduction to the CoNLL-2003 shared task: Language-independent **named entity recognition**
2. Deep learning based NER. NER is often considered as sequence tagging problem in the deep learning setting. Reference: Neural architectures for **named entity recognition**
3. There are lots of papers out there through Google scholar, literature review is necessary to find out which are useful.
4. Exiting NER tools:
    1. CoreNLP (Stanford Tagger, supports both English and Chinese)
    2. Spacy (English only)
    3. HanLP (Chinese only)
    4. others..
5. For Chinese NER, segmentation is often applied before NER. It is not recommend to train a segmentation model alone (unless you come out an idea of joint learning, i.e., segmentation and NER together). Here are some useful tools/packages for Chinese word segmentation:
    1. jieba (very fast and supports Traditional Chinese, although not accurate enough)
    2. HanLP
    3. THULAC
    4. CoreNLP (Stanford Segmenter)
6. For rare word identification, one can search in a large vocabulary, e.g., wordnet (English only), to verity its existence.
7. Some relevant techniques:
    1. Part-of-speech (POS)
    2. Dependency parsing
    3. Tokenize
    4. …