

Developing automatic named entity recognition tools for text mining in biological and chemical corpora

Maxim Holmatov

Supervisor:

Iliia Korvigo

Laboratory of functional analysis
of the genome, MIPT

Introduction

Named entity is a real-world object, such as persons, locations, organizations, products, etc., that can be denoted with a proper name.

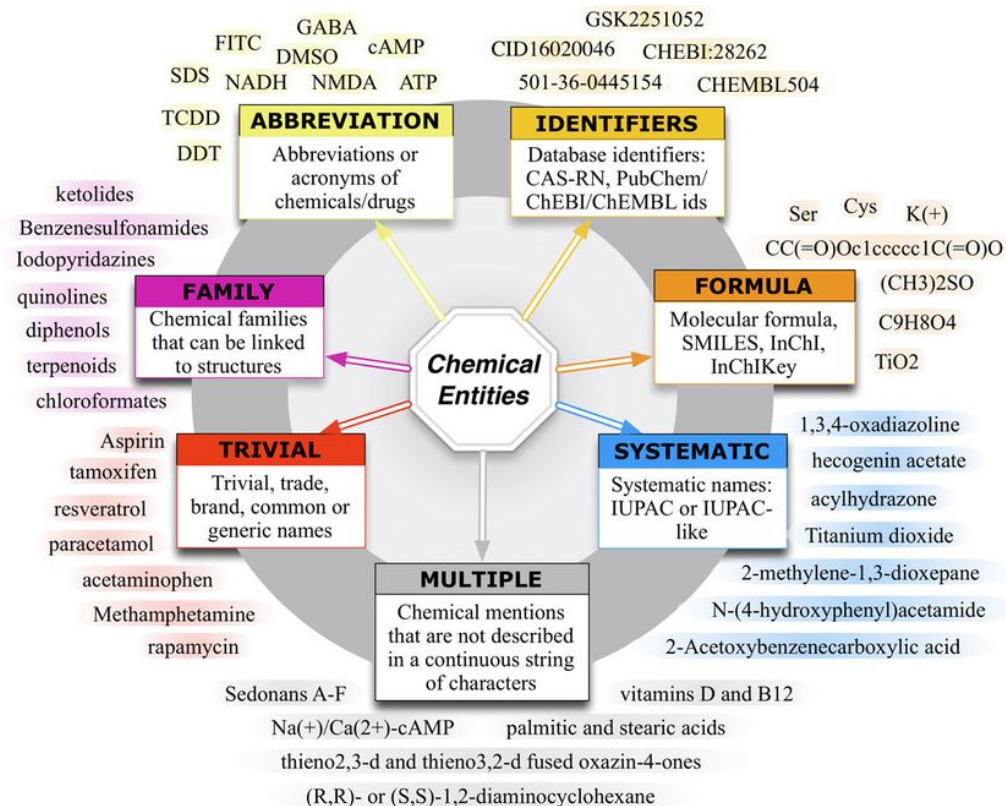
The **peri-kappa B** site mediates **human immunodeficiency virus type 2** activation in **monocytes** but not in **T cells**.

Delineation of the **CD28** signaling cascade was found to involve **protein tyrosine kinase** activity, followed by the activation of **phospholipase A2** and **5-lipoxygenase**.

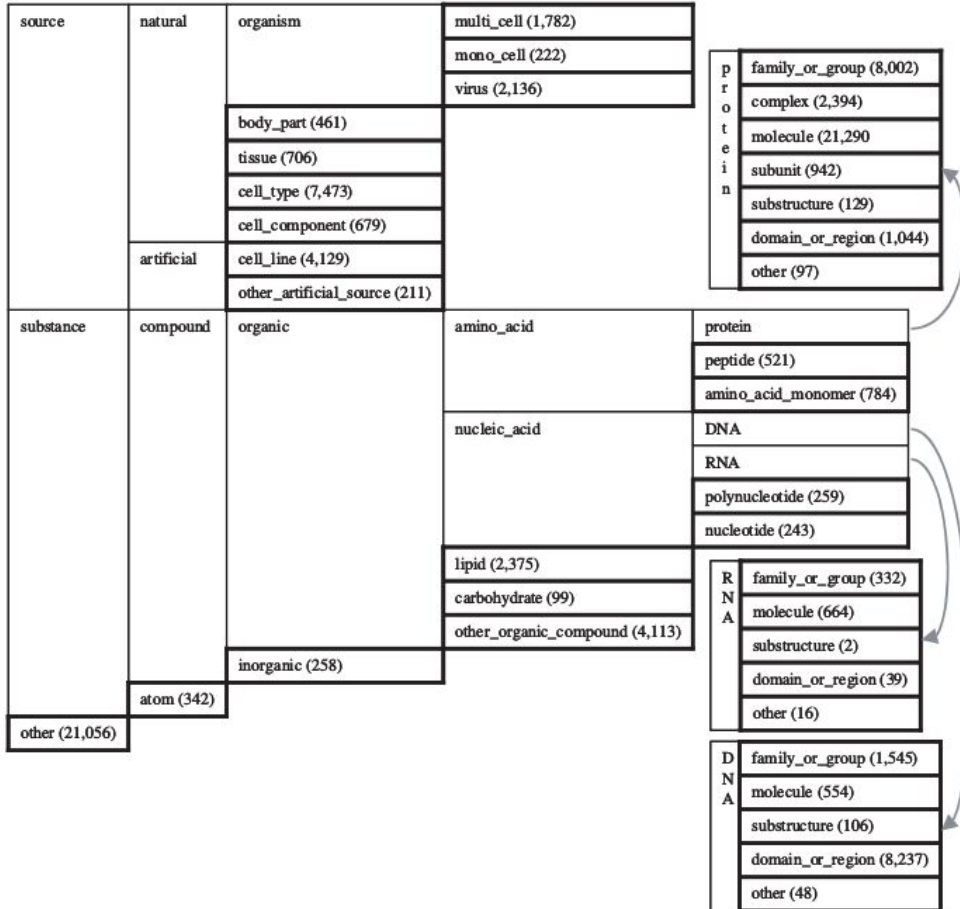
Traditional approaches:

1. Dictionary
 - too big
 - too many terms
2. Rules (e.g. regular expressions)
 - hard to maintain
 - no universal rule
3. ML (SVM, HMM, CRF, ANNs)
 - Woohoo!

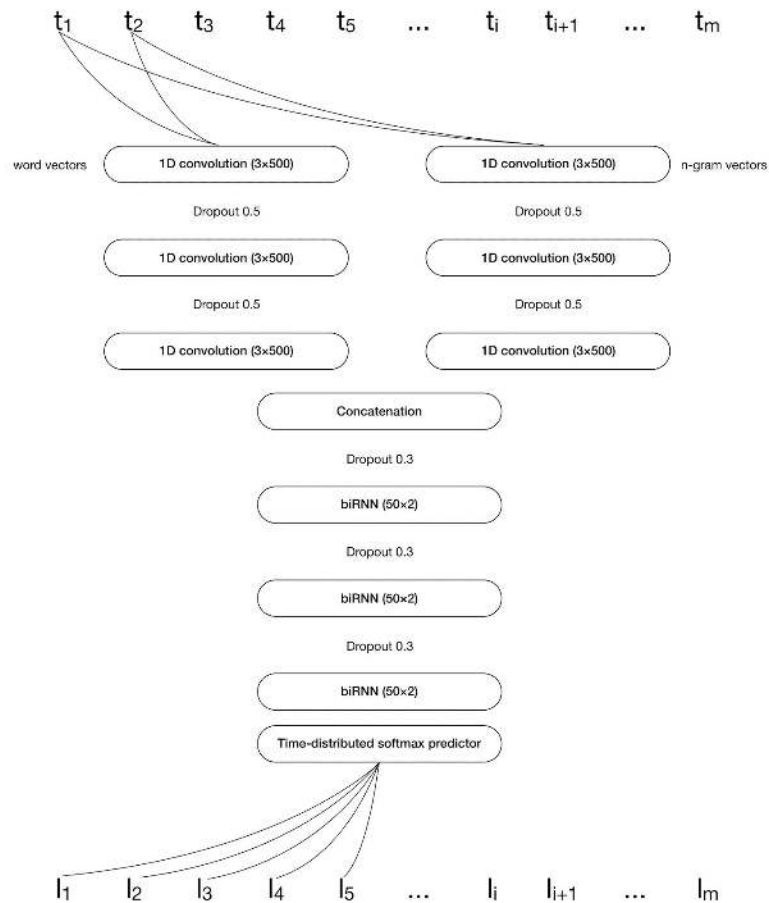
CHEMDNER



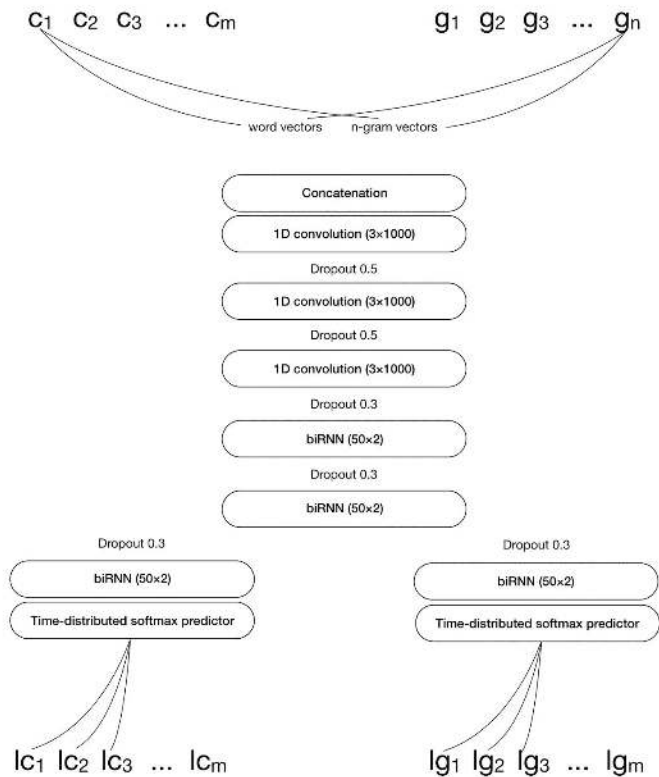
GENIA



Network Architecture for CHEMDNER



Network Architecture for GENIA



Thanks for your attention!

