



The ChEMU evaluation campaign: Named entity recognition and event extraction of chemical reactions from patents

Karin Verspoor, Tim Baldwin, Trevor Cohn, Saber Akhondi,
Dat Quoc Nguyen, Christian Druckenbrodt, Zenan Zhai, Camilo Thorne,
Ralph Hoessel, Biaoyan Fang, Hiyori Yoshikawa

The ChEMU evaluation campaign

- Task 1: Named entity recognition
 - To identify specific types of chemical compounds
 - To assign the label of a chemical compound according to the role for which the chemical compound plays within a chemical reaction, such as Starting_material and Solvent
- Task 2: Event extraction over chemical reactions
 - This task involves event trigger detection, event typing and primary argument recognition

The ChEMU evaluation campaign

10.0 g (35.0 mmol) of **2-tert-butyl 4-ethyl 5-amino-3-methylthiophene-2,4-dicarboxylate** (Example 1A) were dissolved in 500 ml of **dichloromethane** and 11.4 g (70.1 mmol) of **N,N'-carbonyldiimidazole** (CDI) and 19.6 ml (140 mmol) of **triethylamine** were added

ID	Type	Text span
T1	Starting_material	2-tert-butyl 4-ethyl 5-amino-3-methylthiophene-2,4-dicarboxylate
T2	Solvent	dichloromethane
T3	Starting_material	N,N'-carbonyldiimidazole
T4	Reagent	triethylamine
T5	Trigger	dissolved
T6	Trigger	added

ID	Event type	Event trigger	Argument _1	Argument _2	Argument _3
E1	Reaction_step	T5	Theme:T1	Theme:T2	
E2	Reaction_step	T6	Theme:E1	Theme:T3	Theme:T4

Task 1 – NER – in Red

Task 2 – Event extraction – in Purple

The ChEMU evaluation campaign

- Motivation:
 - The chemical and pharmaceutical industries depend on the discovery of new chemical compounds
 - Most chemical compounds are described only in patent documents
 - Automatic natural language processing approaches enable information extraction from the chemical patents and support discovery and synthesis of chemical information
- Goals:
 - To develop tasks that potentially impact chemical research in both academia and industry
 - To provide the community with a new dataset of chemical entities, enriched with relation links between chemical event triggers and arguments
 - To advance the state-of-the-art in information extraction over chemical patents

The ChEMU evaluation campaign

- Why is this campaign needed?
 - There is previously only one shared task on this chemical patent domain, which is the CHEMDNER patents task at the BioCreative V workshop
 - Information extraction approaches developed for the scientific literature domain might not be directly applied to the chemical patent domain: *Patents are written in a very different way as compared to scientific literature*
- These tasks represent a new challenge for IE systems, in an area of significant pharmacological importance
- The campaign will focus attention on more complex analysis of chemical patents, provide strong baselines, and serve as a useful resource for future research