

Problem: Extract amount, company and reason for contracts on Quebec

Dataset: 28,376 documents obtained on a FIOA request by Montreal Gazette journalist Roberto Rocha

Approach

- amount RE
- company with inc.
- reason with concernant/amount to/ etc (10 patterns)

Guidelines:

- internal money transfers do not qualify
- no company, receiver, does not qualify
- donation, qualifies

Input Data: 36 docs

Test Data: 32 docs

Results:

(with precise matching)

- On the same 36 documents were it was programmed:

System

Annotated	tp: 6	fn: 10
	fp: 0	tn: 20

* precision = $6 / 6 = 1.0$

* recall = $6 / 16 = 0.375$

– but partial reason in one and missing reason in another

– 4 fn due to lack of amount, 6 due to lack of inc.

- On 32 new documents:

System

Annotated	tp: 5	fn: 9
	fp: 0	tn: 18

* precision = $5 / 5 = 1.0$

* recall = $5 / 13 = 0.38$

– but only 2 perfect, 1 wrong company and no reason, 1 no reason and 1 reason is whole doc

– 2 fn due to lack of amount, 7 due to lack of inc.