Design Choices in Related Entity Finding

Marc Bron, Manos Tsagias and Maarten de Rijke
The University of Amsterdam ILPS team at the TREC Entity Track 2010

The task: given a Source Entity (S) + Relation (R) + Target Type (T) find homepages of entities that satisfy the S, R and T constraints.

Debut

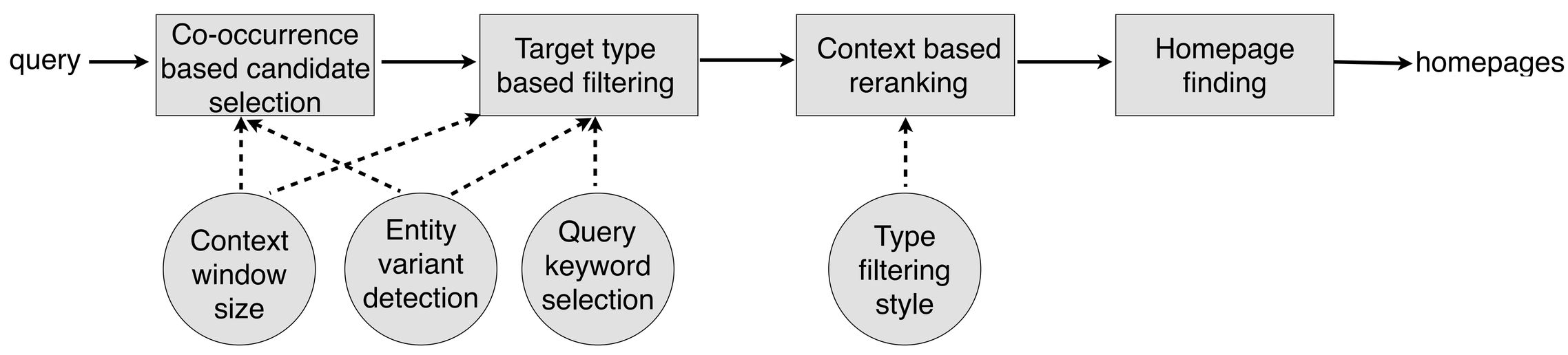
<PER>Schumacher</PER> made his <ORG>Formula One</ORG> debut with the <ORG>Jordan-Ford</ORG> team at the <ORG>1991 Belgian Grand Prix</ORG>, driving car number 32 as a replacement for the imprisoned <PER>Bertrand Gachot</PER>. <PER>Schumacher</PER>, still a contracted <PROD>Mercedes</PROD> driver, was signed by <PER>Eddie Jordan</PER> after <PROD>Mercedes</PROD> paid <ORG>Jordan</ORG> $150,000 for his debut. The week before the race, <PER>Schumacher</PER> impressed <ORG>Jordan</ORG> designer <PER>Gary Anderson</PER> and team manager <PER>Trevor Foster</PER> during a test drive at <LOC>Silverstone</LOC>. His manager <PER>Willi Weber</PER> assured <ORG>Jordan</ORG> that <PER>Schumacher</PER> knew the challenging <LOC>Spa</LOC> track well, although in fact he had only seen it as a spectator. During the race weekend, team-mate <PER>Andrea de Cesaris</PER> was meant to show <PER>Schumacher</PER> the circuit but was held up with contract negotiations.

Schumacher testing the Jordan 191

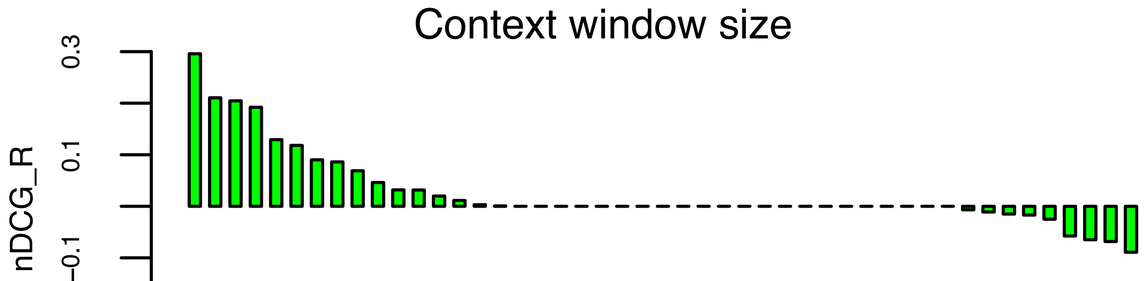
Context

Candidate entities (C)

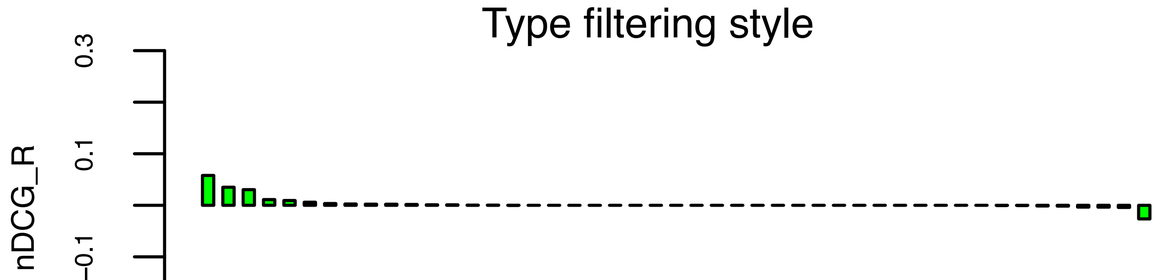
Model and Choices



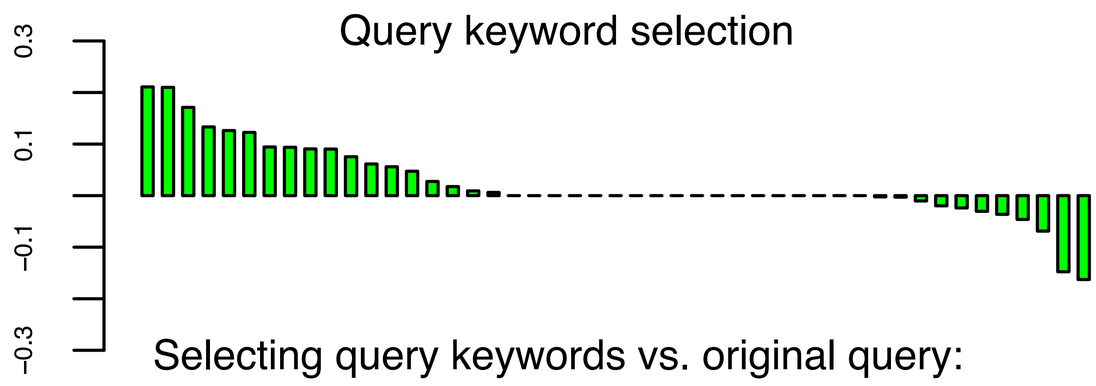
Results and Findings



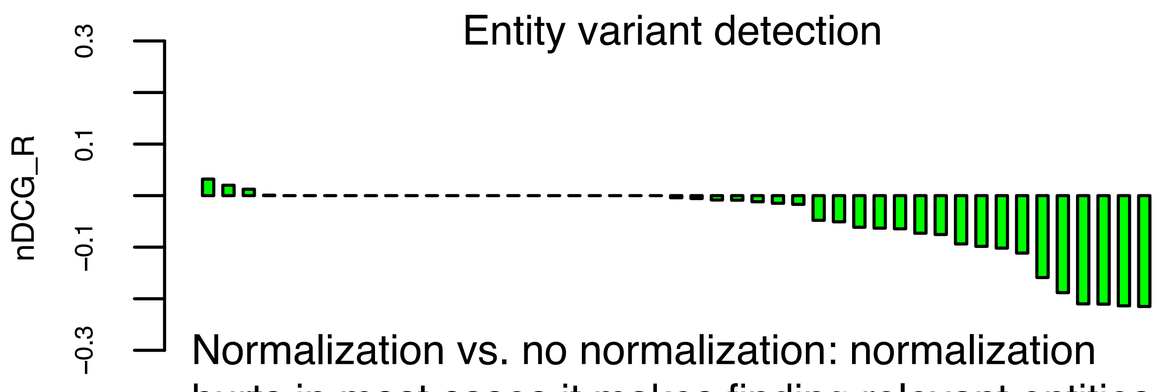
Small vs. large context: a smaller context reduces the number of candidates for frequent entities.



Aggressive vs. lenient filtering: aggressive filtering helps but only marginally.



Selecting query keywords vs. original query: selecting keywords helps in most cases. Adding plural/singular of common words is unhelpful.



Normalization vs. no normalization: normalization hurts in most cases it makes finding relevant entities harder as only one variant occurs in the ranking.

Further Information

Marc Bron - m.m.bron@uva.nl
Intelligent Systems Lab Amsterdam - http://ilps.science.uva.nl
TREC 2010

Acknowledgments

NWO
BiG Grid
the dutch e-science grid