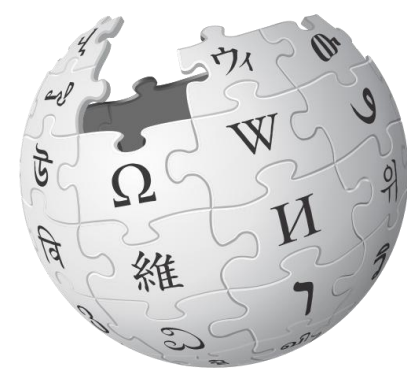# DRETa: Extracting RDF from Wikitables

Emir Muñoz, Aidan Hogan, Alessandra Mileo

National University of Ireland, Galway

Motivation

Manchester United F.C.
From Wikipedia, the free encyclopedia

Manchester United Football Club is an English professional football club, based in Old Trafford, Greater Manchester, that plays in the Premier League. Founded as Newton Heath LYR Football Club in 1878, the club changed its name to Manchester United in 1902 and moved to Old Trafford in 1910.

Manchester United have won many trophies in English football, including a record 20 League titles, a record 11 FA Cups,[3] four League Cups and a record 20 FA Community Shields. The club has also won three European Cups, one UEFA Cup Winners' Cup, one UEFA Super Cup, one Intercontinental Cup and one FIFA Club World Cup. In 1998–99, the club won a continental treble of the Premier League, the FA Cup and the UEFA Champions League, an unprecedented feat for an English club.

The 1958 Munich air disaster claimed the lives of eight players. In 1968, under the management of Matt Busby, Manchester United was the first English football club to win the European Cup. Alex Ferguson won 28 major honours, and 38 in total, from November 1986 to May 2013,[4][5] when he announced his retirement after 26 years at the club.[6] Fellow Scot David Moyes was appointed as his replacement on 9 May 2013.[7]

Manchester United is the third-richest football club in the world for 2011–12 in terms of revenue, with an annual revenue of €395.9 million, and the second most valuable club in 2013,

Manchester United

Full name	Manchester United Football Club
Nickname(s)	The Red Devils[1]
Founded	1878, 135 years ago, as Newton Heath LYR F.C.
Ground	Old Trafford

No.	Position	Player	No.	Position	Player
1	GK	David de Gea	19	FW	Danny Welbeck
2	DF	Rafael	20	FW	Robin van Persie
3	DF	Patrice Evra (vice-captain)	22	DF	Fábio
4	DF	Phil Jones	23	MF	Tom Cleverley
5	DF	Rio Ferdinand	24	MF	Darren Fletcher

Query:

```
SELECT ?player
WHERE {
    ?player dbp:currentclub dbr:Manchester_United_F.C .
}
```

player
http://dbpedia.org/resource/David_de_Gea
http://dbpedia.org/resource/Rafael_Pereira_da_Silva_(footballer_born_1990)
http://dbpedia.org/resource/Patrice_Evra
....
⚠ http://dbpedia.org/resource/Fabio_Pereira_da_Silva
http://dbpedia.org/resource/Tom_Cleverley
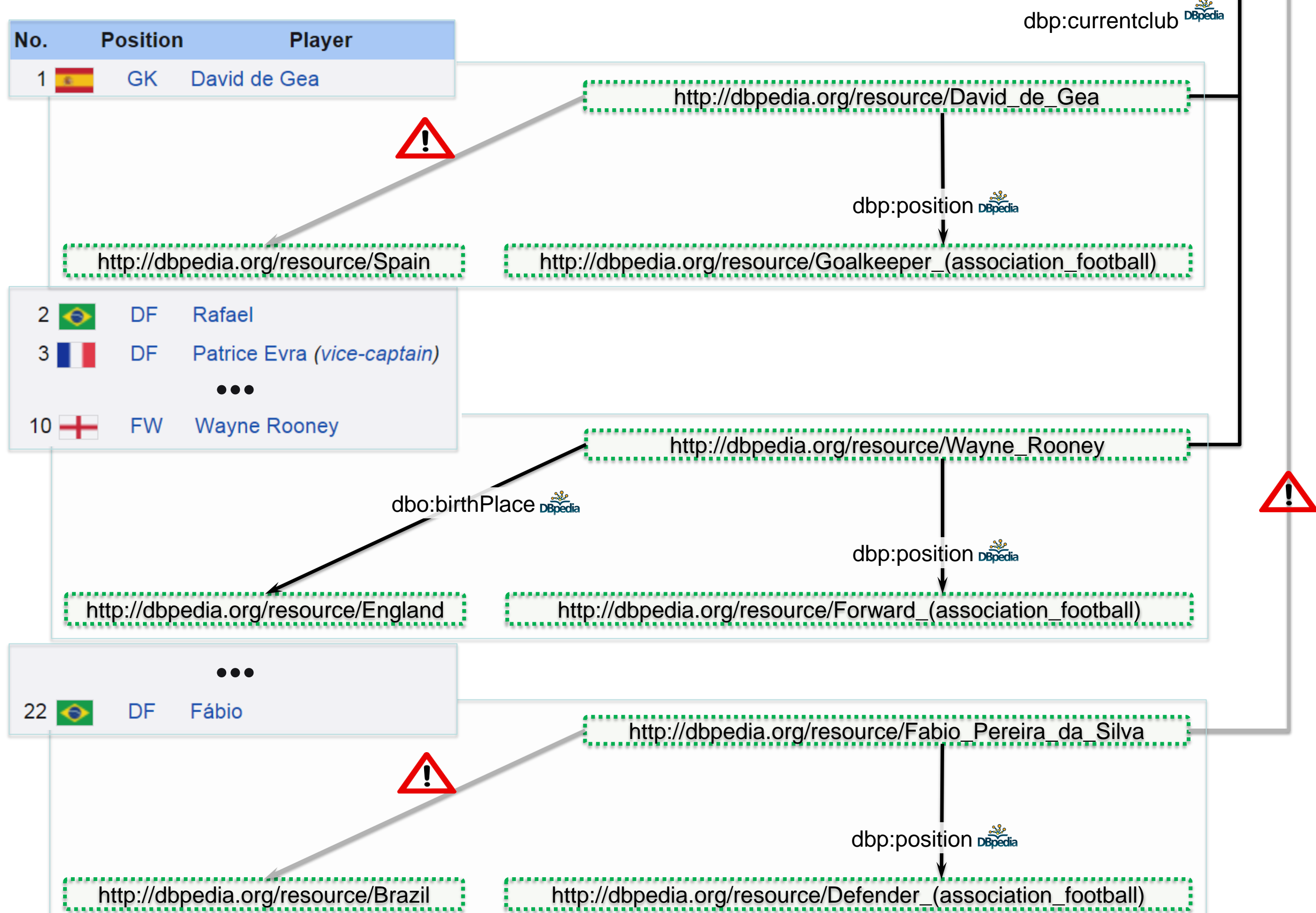http://dbpedia.org/resource/Darren_Fletcher

... Incomplete Results!

Proposal

Manchester United F.C.
From Wikipedia, the free encyclopedia

http://dbpedia.org/resource/Manchester_United_F.C.

dbp:currentclub

No.	Position	Player
1	GK	David de Gea

http://dbpedia.org/resource/David_de_Gea

dbp:position

http://dbpedia.org/resource/Spain → http://dbpedia.org/resource/Goalkeeper_(association_football)

2	DF	Rafael
3	DF	Patrice Evra (vice-captain)
...		
10	FW	Wayne Rooney

http://dbpedia.org/resource/Wayne_Rooney

dbo:birthPlace

dbp:position

http://dbpedia.org/resource/England → http://dbpedia.org/resource/Forward_(association_football)

| 22 | DF | Fábio |

http://dbpedia.org/resource/Fabio_Pereira_da_Silva

dbp:position

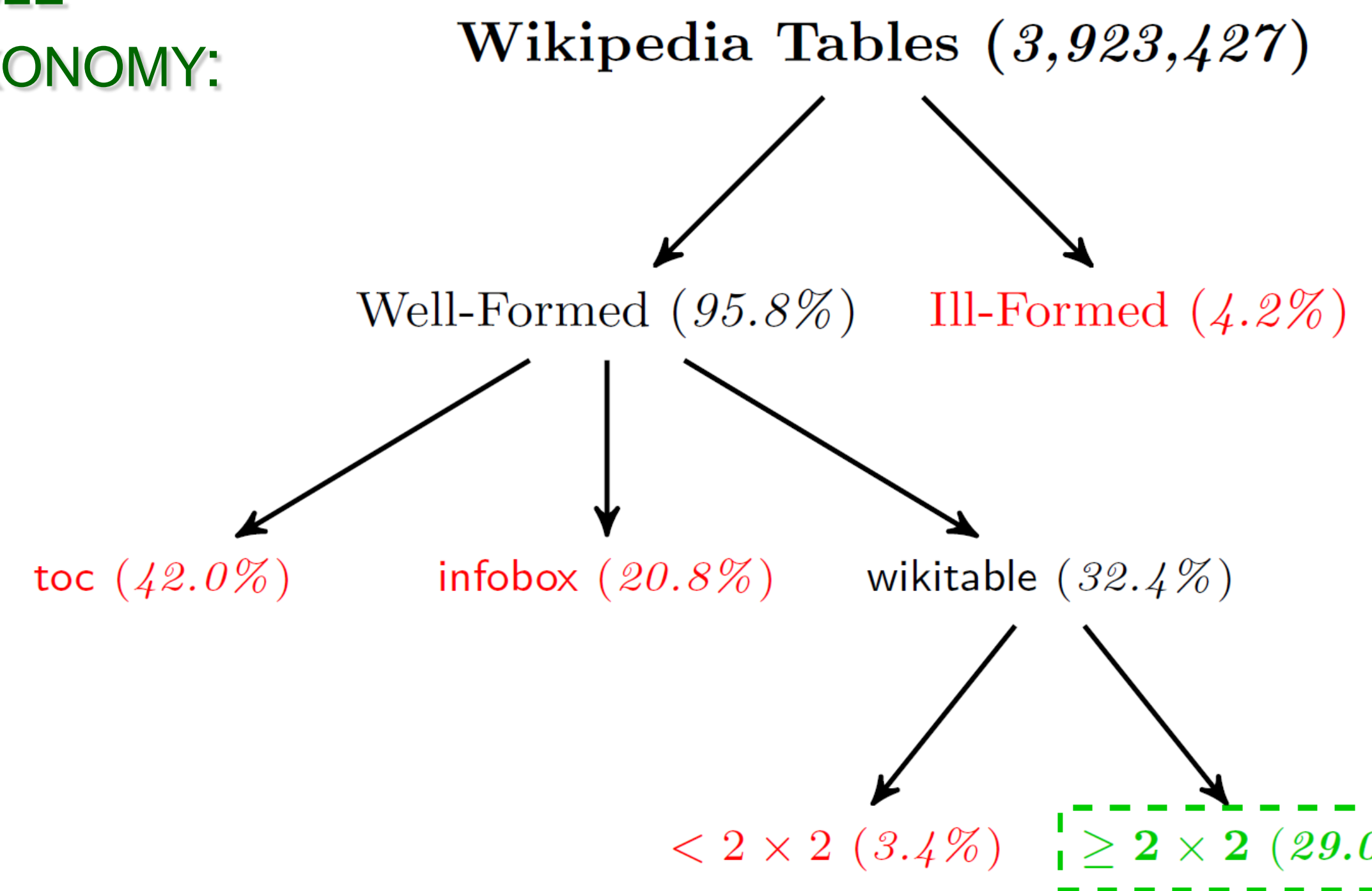http://dbpedia.org/resource/Brazil → http://dbpedia.org/resource/Defender_(association_football)
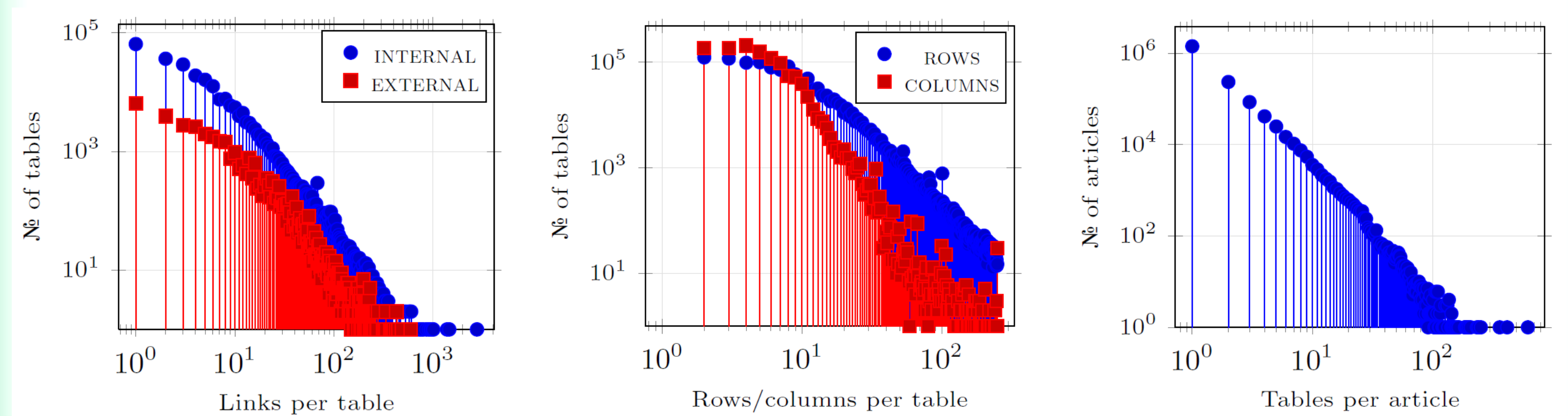
Suggested Triples:
(1) dbr:David_de_Gea dbo:birthPlace dbr:Spain .
(2) dbr:Fabio_Pereira_de_Silva dbo:birthPlace dbr:Brazil .
(3) dbr:Fabio_Pereira_de_Silva dbp:currentclub dbr:Manchester_United_F.C .

Wikitable survey

Table Taxonomy:

Wikipedia Tables $(3,923,427)$

Well-Formed (95.8%) — Ill-Formed (4.2%)

toc (42.0%) — infobox (20.8%) — wikitable (32.4%)

$< 2 \times 2$ (3.4%) — $\geq 2 \times 2$ (29.0%)

Distributions:

Results

(1) Extracted 34.9 million unique & novel triples from 1.14 million Wikitables
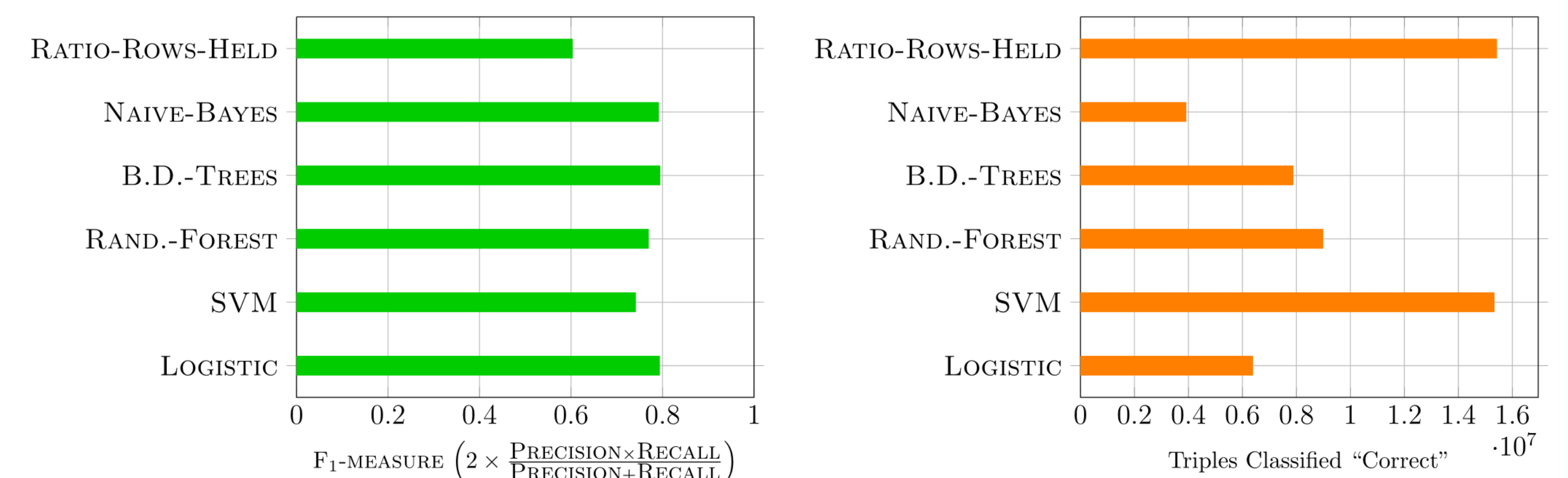(8 machines: 4GB RAM, 2.2 GHz single core; 12 days)

(2) Initial Evaluation:
(Manual Annotation; Three Judges; 750 Triples each)

All *correct*	234	31.2%
Some *correct*, Rest *unknown*	27	3.6%
Positive consensus	**261**	**34.8%**
All *incorrect*	195	26.0%
Some *incorrect*, Rest *unknown*	48	6.4%
Negative consensus	**243**	**42.8%**
Some *correct*, Some *incorrect* (removed)	246	32.8%

(3) Machine Learning classifiers:
(Consensus Gold Standard; Variety of Features)

F_1-measure $\left(2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right)$

Triples Classified "Correct"

From 1.14 million Wikitables:
Bagging Decision Trees: 7.9 million triples @81.5% prec.
Support Vector Machines: 15.3 million triples @72.4% prec.

Demo ... http://emunoz.org/wikitables

Enabling **Networked** Knowledge