



# Automating News Content Analysis

## An Application to Gender Bias and Readability

Omar Ali, Ilias Flaounas, Tijn De Bie, Nello Cristianini,  
Nick Mosdell, Justin Lewis.



# Overview

- Examine topics in the media.
- Examples of research by social scientists.
- How we can automate such work?
- Demonstrate biases by topic:
  - Gender.
  - Readability.



The New York Times

Thursday, May 27, 2010

Local Times: a collection feed of all the business headlines.

**Top Kill Effort Seems to Be Working, U.S. Says Cautiously**

Oil spill cleanup officials said on Thursday that the flow of oil from the well was almost three weeks off schedule.

**Obama Debates Handling of Oil Spill**

President Obama said on Thursday that he is "proud and breathless" over the effort to stop the spill, and considered a nomination on new powers to fund new deepwater wells.

**Estimates Suggest Spill is Impacting U.S. History**

A National Geographic magazine article said the spill is "likely to have a profound impact on the nation's history."

**New U.S. Security Strategy Focuses on Managing Threats**

The Obama administration said on Thursday that it is shifting its focus on security to managing threats.

**Obama's New Executive Order**

President Obama signed an executive order on Thursday that would allow the federal government to use more power to regulate the environment.

**Without being connected to the Web.**

Times Reader 2.0. The parallel, portable, mobile reader.

FINANCIAL TIMES

Five over Exxon: Dated \$1.6B on dark side of China factories.

John Gapper: Murdoch has to become online boss.

Deen on Berlusconi: 'Overpriced' of 'Bash' host.

Market: Japan share price at 6-year low as a result of a weak recovery.

FT.com

UK World Companies Markets Global Economy Law Comment Management Personal Finance Arts & Culture

**Pru tries to renegotiate AIA price**

Bid to cut purchase to \$yobos to win over bigwigs investors.

**Obama unveils offshore drilling rules**

Oil will need to be larger than estimated.

**Duncan Smith pledges benefits shake up**

Radical overhaul of benefits to ensure 'work pays'.

**Euro's rebound helps to boost stocks**

Hope for global recovery trump fears over uncertainty.

**China fund admits 'concern' over eurozone**

Manager of China's sovereign wealth fund says 'concern' over eurozone.

**First Handed Clowns are Found in Britain**

First handed clowns have been spotted in the UK.

Markets data

INDEXES	COMMODITIES	CURRENCIES
FTSE 100: 5,195 (+1.2%)	Oil: 84.28 (+0.1%)	USD/GBP: 1.547 (+0.01)
DAX: 7,048 (+0.8%)	Gold: 1,218.50 (+0.1%)	EUR/GBP: 0.736 (+0.01)
Nikkei 225: 9,840 (+1.2%)	Platinum: 1,565.00 (+0.1%)	GBP/USD: 0.647 (-0.01)
S&P 500: 1,188.14 (+1.1%)	Iron Ore: 147.50 (+0.1%)	GBP/JPY: 161.15 (+0.1%)

**FIND OUT AT OUR VITAL TOPICS LECTURE SERIES.**

Science

NEWS SCIENCE JOURNALS CAREERS BLOGS & COMMUNITIES MULTIMEDIA COLLECTIONS

**SYNTHETIC BIOLOGY BREAKTHROUGH**

University of California researchers have created a synthetic cell that can grow and divide.

**Obama's New Executive Order**

President Obama signed an executive order on Thursday that would allow the federal government to use more power to regulate the environment.

**Without being connected to the Web.**

Times Reader 2.0. The parallel, portable, mobile reader.

**Pipetting, it's easier with us.**

Introducing the new pipette.

Celebrity News, Photos, etc.

**Cheryl files for divorce**

Phil and Ashley's divorce is finalized before the World Cup.

**1/2 PRICE Laser Eye Surgery**

Optimax

**Today on MSN Celebrity**

Louise Wulfeaton's opening: There were some surprise appearances from the original stars of the original 'The Louie Wulfeaton Show'.

**World premiere of SALES - pick it up to go**

World premiere of SALES - pick it up to go.

**World premiere of SALES - pick it up to go**

World premiere of SALES - pick it up to go.

BBC SPORT

The Olympic Debacle

**World Cup 2010**

South Africa World Cup Journey

**Live - Murray v Chela**

Andy Murray returns for second-round clash with Juan Ignacio Chela at a new-entrained Rio de Janeiro Open.

**French Open, 22 May 6 June**

BBC coverage | Match diary | Highlights | Schedule | Tickets | Official site

**World Cup 2010**

World Cup 2010: South Africa World Cup Journey

**World Cup 2010**

World Cup 2010: South Africa World Cup Journey

guardian.co.uk

News | Sport | Comment | Culture | Business | Money | Life & Style | Travel | Environment | TV | Blogs | Video | Community | Jobs

**politics**

**David Cameron says coalition may compromise over capital gains tax**

Prime minister refuses to sign accompanying paper on CGT which would help to raise money for the NHS.

**Pulling stations open for delayed election in Thailand and Malawi**

Thailand's election is delayed as a result of a military coup.

**Why has the BBC given Hazel Blears a place on Andrew Neil's sofa?**

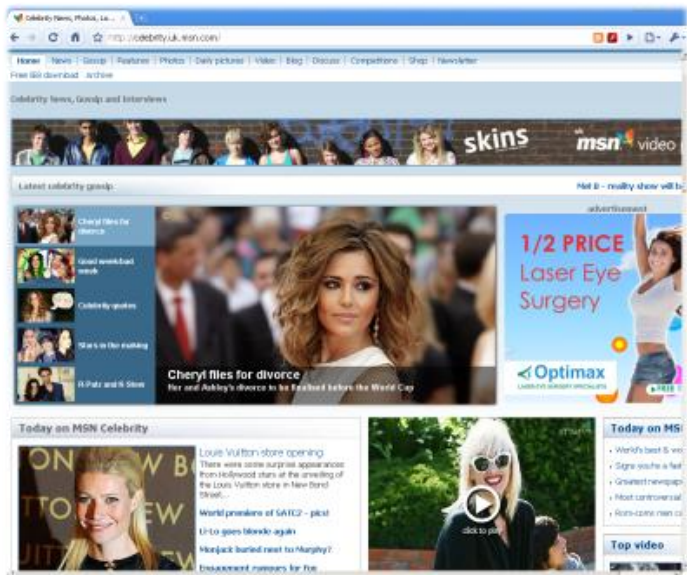
Andrew Neil's sofa.



Readability:  
**FRES = 51.8**

References to:  
**Males = 6**

**Females = 2**



Readability:  
**FRES = 66.9**

References to:  
**Males = 2**

**Females = 2**

# Results from the Social Sciences

- The readability (FRES) of news articles is higher for fabricated stories (Dalecki, Lasorsa, Lewis, 2009).
- Gender bias in news articles varies across topics (Len-Rios et al., 2005):
  - Entertainment is the least male-biased.
  - Front Page.
  - Business.
  - Sport displays the most male-bias.



# How can we automate this?

## MSN Celebrity Blog »

Refresh

- ☆ **Brittany Murphy's widower found dead** - Simon Monjack has died jus
- ☆ **Rachel Stevens is expecting!** - Rachel Stevens has delighted her fans
- ☆ **Russell and Rossy to reunite** - Jonathan Ross is reportedly ending his
- ☆ **Katy Perry tops Maxim Hot 100 list** - Katy Perry might be about to tie th
- ☆ **Kylie, shy and retiring? Pah!** - Kylie Minogue might be one of the most
- ☆ **Jude and Sienna: it's official!** - Jude Law and Sienna Miller have been
- ☆ **Sandra Bullock files for divorce, adopts baby** - Sandra Bullock has fil
- ☆ **Jack Tweed cleared of rape charges** - Jack Tweed, the widower of Ja
- ☆ **Mark Owen reunites with wife** - Mark Owen recently admitted to having
- ☆ **Brian apologises for Twitter rant** - It's been a tough couple of years for
- ☆ **Diana Vickers on life after X Factor** - Our music editor James Hurley in
- ☆ **Kim Kardashian dating Cristiano Ronaldo?** - A couple of weeks ago,
- ☆ **Kerry Katona on her 'car crash' life** - Kerry Katona has finally opened t
- ☆ **Katie Price's death fears** - In an interview with OK magazine, Katie Pric
- ☆ **Kylie for Fashion Targets Breast Cancer** - New pictures for the campe
- ☆ **Cheryl's voicemail rape threat drama** - Detectives have launched an in
- ☆ **Has Isla Fisher become Mrs Borat?** - They've been engaged for a who
- ☆ **Sandra Bullock hit by Oscars curse?** - Sandra Bullock has pulled out o
- ☆ **Katie to launch kids make-up range?** - Katie Price is no stranger to cc
- ☆ **Cheryl and Ashley finally meet** - After weeks of being apart, Cheryl and
- ☆ **Kerry Katona signs Peter Andre's manager** - It's not been an easy co

## BBC Sport | Sport Homepage | UK Edition »

Refresh

- ☆ **Red Bull out on own - Schumacher** - Former Formula 1 world champi
- ☆ **Montgomerie plans Garcia pep talk** - Colin Montgomerie plans to talk t
- ☆ **Ferrari set to hold on to Massa** - Ferrari say they are almost certain to r
- ☆ **Mourinho 'a perfect fit for Real'** - Jose Mourinho's style of play will "fit in
- ☆ **Live - England v Bangladesh** - England reach 100-1 after being put in l
- ☆ **Live - French Open** - Play finally begins at a showery Roland Garros wh
- ☆ **Hill to be named Worcester boss** - Worcester Warriors are expected to
- ☆ **Exeter seal Premiership promotion** - Gareth Steenson kicks 24 points
- ☆ **Cardiff's £6m takeover approved** - Cardiff City shareholders rubber-st
- ☆ **Hooker Brits wins players' award** - Saracens hooker Schalk Brits is vc
- ☆ **Cotterill leaves Notts County job** - Steve Cotterill, who has been linked
- ☆ **Woods to return at Memorial event** - World number one Tiger Woods v
- ☆ **Thursday's gossip column** - Inter Milan in for Fabio Capello, plus other
- ☆ **Jockey Holland to miss Derby ride** - Jockey Darryll Holland will miss th
- ☆ **Radcliffe looks beyond 2012 Games** - Paula Radcliffe raises the prosp
- ☆ **Pellegrini sacked by Real Madrid** - Real Madrid sack coach Manuel Pe
- ☆ **Second player targeted by fixers** - The England and Wales Cricket Bo
- ☆ **Pompey creditors offered 20% deal** - Portsmouth's creditors are to be
- ☆ **Trial by television - BBC to show stewards' inquiries** - Stewards' inq
- ☆ **Red Bull can be beaten - McLaren** - Championship leaders Red Bull ca
- ☆ **Strauss warns players over fixing** - England captain Andrew Strauss u

# How can we automate this?

## MSN Celebrity Blog »

Refresh


- ☆ **Brittany Murphy's widower found dead** - Simon Monjack has died jus
- ☆ **Rachel Stevens is expecting!** - Rachel Stevens has delighted her fans
- ☆ **Russell and Rossy to reunite** - Jonathan Ross is reportedly ending his
- ☆ **Katy Perry tops Maxim Hot 100 list** - Katy Perry might be about to tie th
- ☆ **Kylie, shy and retiring? Pah!** - Kylie Minogue might be one of the most
- ☆ **Jude and Sienna: it's official!** - Jude Law and Sienna Miller have been
- ☆ **Sandra Bullock files for divorce, adopts baby** - Sandra Bullock has fil
- ☆ **Jack Tweed cleared of rape charges** - Jack Tweed, the widower of Ja
- ☆ **Mark Owen reunites with wife** - Mark Owen recently admitted to having
- ☆ **Brian apologises for Twitter rant** - It's been a tough couple of years for
- ☆ **Diana Vickers on life after X Factor** - Our music editor James Hurley in
- ☆ **Kim Kardashian dating Cristiano Ronaldo?** - A couple of weeks ago,
- ☆ **Kerry Katona on her 'car crash' life** - Kerry Katona has finally opened t
- ☆ **Katie Price's death fears** - In an interview with OK magazine, Katie Pric
- ☆ **Kylie for Fashion Targets Breast Cancer** - New pictures for the campe
- ☆ **Cheryl's voicemail rape threat drama** - Detectives have launched an in
- ☆ **Has Isla Fisher become Mrs Borat?** - They've been engaged for a who
- ☆ **Sandra Bullock hit by Oscars curse?** - Sandra Bullock has pulled out o
- ☆ **Katie to launch kids make-up range?** - Katie Price is no stranger to cc
- ☆ **Cheryl and Ashley finally meet** - After weeks of being apart, Cheryl and
- ☆ **Kerry Katona signs Peter Andre's manager** - It's not been an easy co

## BBC Sport | Sport Homepage | UK Edition »

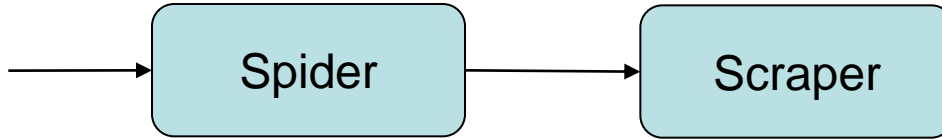
Refresh

- ☆ **Red Bull out on own - Schumacher** - Former Formula 1 world champi
- ☆ **Montgomerie plans Garcia pep talk** - Colin Montgomerie plans to talk t
- ☆ **Ferrari set to hold on to Massa** - Ferrari say they are almost certain to r
- ☆ **Mourinho 'a perfect fit for Real'** - Jose Mourinho's style of play will "fit in
- ☆ **Live - England v Bangladesh** - England reach 100-1 after being put in l
- ☆ **Live - French Open** - Play finally begins at a showery Roland Garros wh
- ☆ **Hill to be named Worcester boss** - Worcester Warriors are expected to
- ☆ **Exeter seal Premiership promotion** - Gareth Steenson kicks 24 points
- ☆ **Cardiff's £6m takeover approved** - Cardiff City shareholders rubber-sta
- ☆ **Hooker Brits wins players' award** - Saracens hooker Schalk Brits is vc
- ☆ **Cotterill leaves Notts County job** - Steve Cotterill, who has been linke
- ☆ **Woods to return at Memorial event** - World number one Tiger Woods v
- ☆ **Thursday's gossip column** - Inter Milan in for Fabio Capello, plus other
- ☆ **Jockey Holland to miss Derby ride** - Jockey Darryll Holland will miss th
- ☆ **Radcliffe looks beyond 2012 Games** - Paula Radcliffe raises the prosp
- ☆ **Pellegrini sacked by Real Madrid** - Real Madrid sack coach Manuel Pe
- ☆ **Second player targeted by fixers** - The England and Wales Cricket Bo
- ☆ **Pompey creditors offered 20% deal** - Portsmouth's creditors are to be
- ☆ **Trial by television - BBC to show stewards' inquiries** - Stewards' inq
- ☆ **Red Bull can be beaten - McLaren** - Championship leaders Red Bull cr
- ☆ **Strauss warns players over fixing** - England captain Andrew Strauss u

# How can we automate this?



Source	Country	Headline
BBC Sport	UK	BBC Sport   Sport Homepage   UK Edition
MSN	USA	Red Bull out on own - Schumacher - Former Formula 1 world champion
Reuters	Spain	Montenegro plans Garcia pay talk - Colin Montgomerie plans to take
BBC	UK	Ferrari set to hold on to Massa - Ferrari say they are almost certain to
BBC	UK	Mourinho 'a perfect fit for Real' - Jose Mourinho's style of play will fit
BBC	UK	Live - England v Bangladesh - England reach 1000 after being put in
MSN	UK	Live - French Open - The lady begins at a shrewd Roland Garros with
Reuters	UK	Hill to be named Worcester boss - Worcester Warriors are expected to
BBC	UK	Exeter seal Premiership promotion - Gareth Stevenson scores 24 points
Rac	UK	Cardiff's 60m takeover approved - Cardiff City shareholders rubber-stamp
Live	UK	Hooker bids to win players' award - Connors hooker Jonah Dillnott is
Rus	UK	Cotterill leaves Notts County job - Steve Cotterill, who has been linked
Kat	UK	Woods to return at Memorial event - World number one Tiger Woods
Kyl	UK	Thursday's gossip column - Inter Milan in for Fabio Capello, plus other
Jud	UK	Jockey Holland to miss Derby ride - Jockey Daryl Holland will miss the
Sar	UK	Rodriguez looks beyond 2012 Games - Paula Rodriguez rejects the prospect
Jac	UK	Pellegrini sacked by Real Madrid - Real Madrid sack coach Manuel Pellegrini
Mar	UK	Second player targeted by fixers - The England and Wales Cricket Board
Bre	UK	Pompey creditors offered 20% deal - Portsmouth's creditors are to be
Die	UK	Trial by television - BBC to show stewards' inquiries - Stewards' inquiry
Kim	UK	Red Bull can be beaten - McLaren - Championship leader Red Bull
Ra	UK	Strauss warns players over fixing - England captain Andrew Strauss
Kat	UK	Cheryl and Ashley finally meet - How much of being apart, Cheryl and
Kyl	UK	Kerry Katona signs Peter Andre's manager - It's not been an easy job
Chi	UK	Red Bull can be beaten - McLaren - Championship leader Red Bull
Has	UK	Strauss warns players over fixing - England captain Andrew Strauss
Sar	UK	Cheryl and Ashley finally meet - How much of being apart, Cheryl and
Chi	UK	Kerry Katona signs Peter Andre's manager - It's not been an easy job





# Spider

## Title:

Red Bull can be beaten – McLaren

## Description:

Championship leaders Red Bull can be caught this season, according to McLaren managing director Jonathan Neale.

## URL:

<http://news.bbc.co.uk/sport1/hi/...>

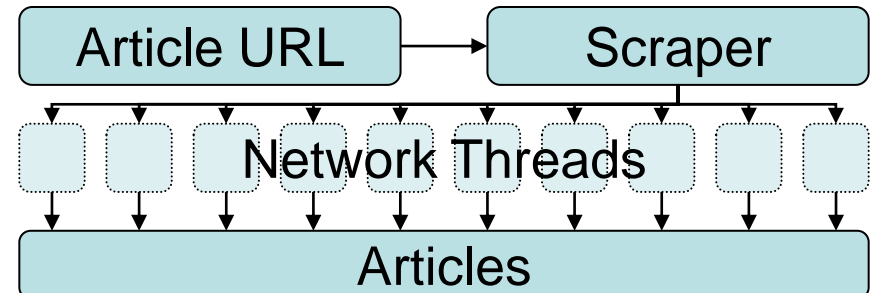
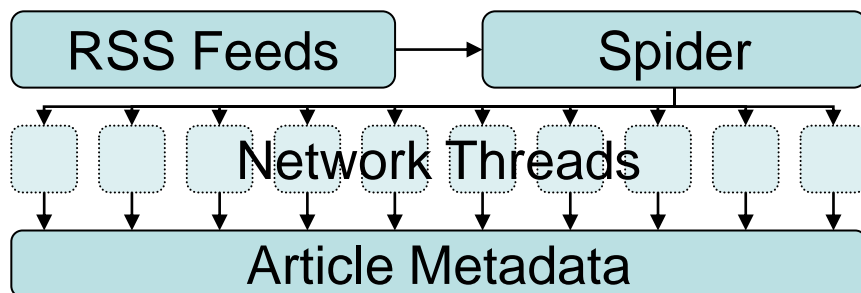
# Scraper

## Content:


Red Bull have been in scintillating form this campaign, claiming pole in all six races and winning three.

McLaren and the other teams have been left playing catch-up but Neale said: "They can be beaten.

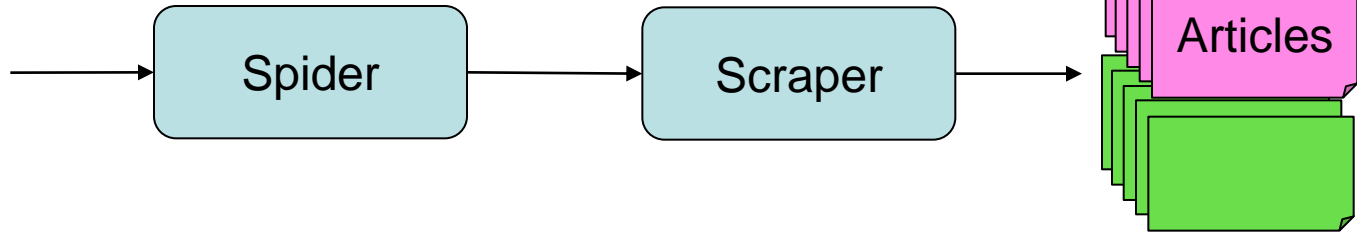
"Red Bull have a have a quick car and they have the upper hand but I am confident that won't be the case by the middle of the season." Red Bull lead both the drivers' and constructors' standings with Australian Mark Webber...



# How can we automate this?



BBC Sport	Sport Homepage	UK Edition
MSN	Red Bull out on own - Schumacher	Former Formula 1 world champion
BBC	Montepierre plans Garcia pep talk	Colin Montepierre plans to talk to
BBC	Ferrari set to hold on to Massa	Ferrari say they are almost certain to
BBC	Mourinho 'a perfect fit for Real'	Jose Mourinho's style of play will fit
BBC	Live - England v Bangladesh	England reach 1000 after being pulled
BBC	Live - French Open	French Open begins at a shrewdly played
MSN	Hill to be named Worcester boss	Worcester Warriors are expected to
BBC	Exeter seal Premiership promotion	Garth Stevenson scores 24 points
BBC	Cardiff's Kim takeover approved	Cardiff City shareholders rubber-stamp
BBC	Hooker offers Wales players' award	Cardiff hooker Jonah Williams to
BBC	Cotterill leaves Notts County job	Steve Cotterill, who has been linked
BBC	Woods to return at Memorial event	World number one Tiger Woods' plus other
BBC	Thursday's gossip column	Insider Milan in for Fabio Capello, plus other
BBC	Jockey Holland to miss Derby ride	Jockey Danny Holland will miss the
BBC	Rodcliffe looks beyond 2012 Games	Paul Rodcliffe reopens the possibility
BBC	Pellegrini sacked by Real Madrid	Real Madrid sack coach Manuel Pellegrini
BBC	Second player targeted by fixers	The England and Wales Cricket Board
BBC	Pompey creditors offered 20% deal	Portsmouth's creditors are to be offered
BBC	Trial by television - BBC to show stewards' inquiries	Stewards' inquiries
BBC	Red Bull can be beaten - McQueen	Championship leader Red Bull
BBC	Strauss warns players over fixing	England captain Andrew Strauss
BBC	Kallis to return to international cricket	South African batsman Jacques Kallis
BBC	Cheryl and Ashley finally meet	West Indies' Cheryl and Ashley
BBC	Kerry Katona signs Peter Andre's manager	It's not been an easy job




The system monitored:

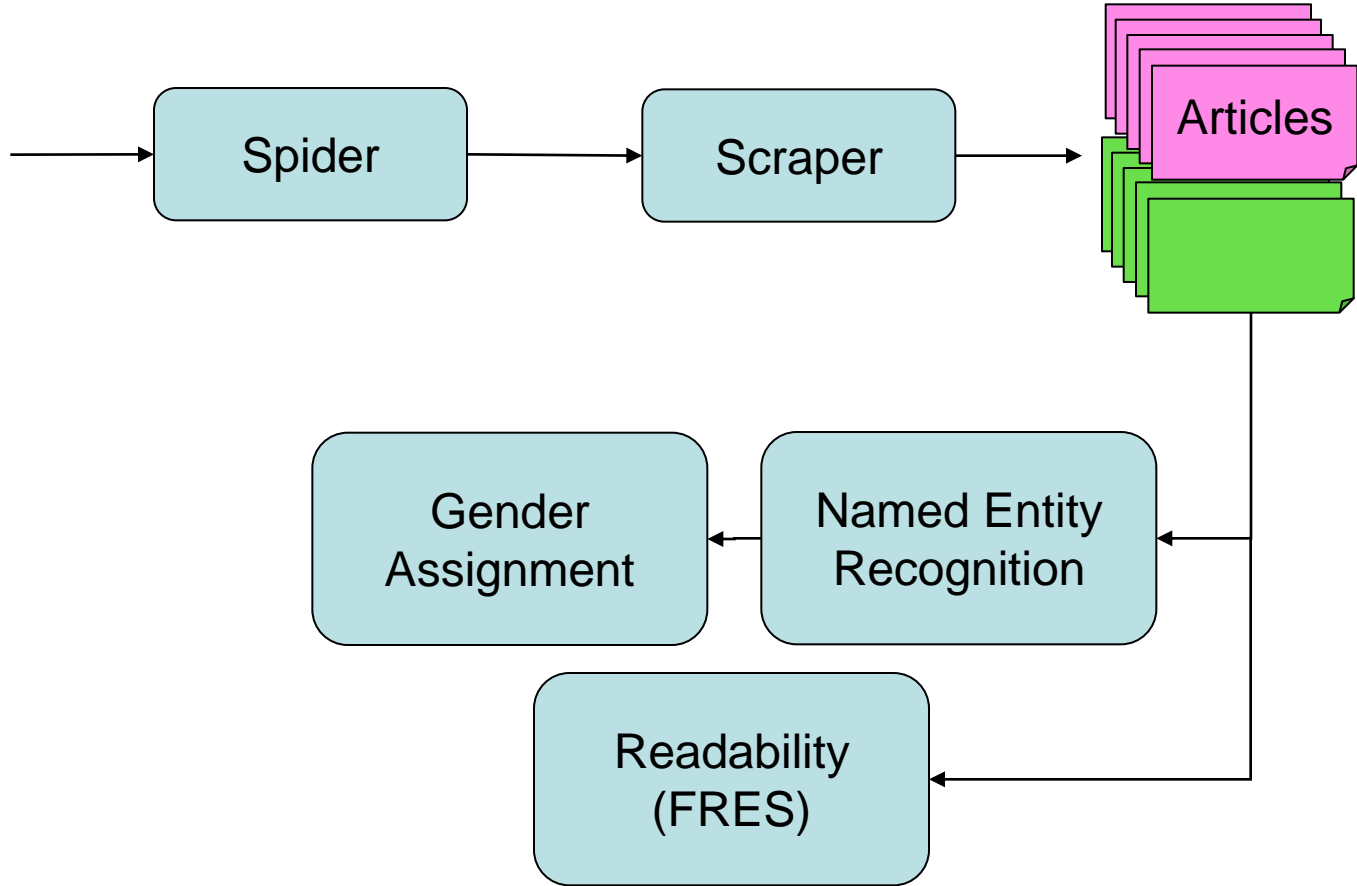
- 6 topics
- 1,121 outlets
- 14 months
- 3.5 million English-language articles



# How can we automate this?



BBC Sport   Sport Homepage   UK Edition	
MSN	Red Bull out on own - Schumacher - Former Formula 1 world champion
Rubus	Montepierre plans Garcia pay talk - Colin Montepierre plans to talk
BBC	Ferrari set to hold on to Massa - Ferrari say they are almost certain to
MSN	Mourinho 'a perfect fit for Real' - Jose Mourinho's style of play will fit
Rus	Live - England v Bangladesh - England reach 1000 after being put in
MSN	Live - French Open - The lady begins at a shrewd Roland Garros with
Rubus	Hill to be named Worcester boss - Worcester Warriors are expected to
BBC	Exeter seal Premiership promotion - Gareth Stevenson scores 24 points
Rac	Cardiff's 60m takeover approved - Cardiff City shareholders rubber-stamp
Liv	Hooker 60th with players' award - Captain Hooker 50th birthday
Rus	Cotterill leaves Notts County job - Steve Cotterill, who has been linked
Kat	Woods to return at Memorial event - World number one Tiger Woods
Kyfl	Thursday's gossip column - Inter Milan in for Fabio Capello, plus other
Jud	Jockey Holland to miss Derby ride - Jockey Danny Holland will miss the
Sar	Rodriguez looks beyond 2012 Games - Spain's Rodriguez reopens the possibility
Jacj	Pellegrini sacked by Real Madrid - Real Madrid sack coach Manuel Pellegrini
Mar	Second player targeted by fixers - The England and Wales Cricket Board
Bbc	Pompey creditors offered 20% deal - Portsmouth's creditors are to be
Die	Trial by television - BBC to show stewards' inquiries - Stewards' inquiry
Kim	Red Bull can be beaten - McLaren - Championship leaders Red Bull can
Ra	Strauss warns players over fixing - England captain Andrew Strauss
Ker	Kallis to return to international range - India's Kallis to be available for
Kat	Cheryl and Ashley finally meet - With weeks of being apart, Cheryl and
Kyfl	Kerry Katona signs Peter Andre's manager - It's not been an easy job
Chi	Red Bull can be beaten - McLaren - Championship leaders Red Bull can
Has	Strauss warns players over fixing - England captain Andrew Strauss
Sar	Kallis to return to international range - India's Kallis to be available for
Ker	Cheryl and Ashley finally meet - With weeks of being apart, Cheryl and
Kyfl	Kerry Katona signs Peter Andre's manager - It's not been an easy job



# Flesch Reading Ease Score

$$206.876 - 1.015 \left( \frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left( \frac{\text{total syllables}}{\text{total words}} \right)$$

Scores:

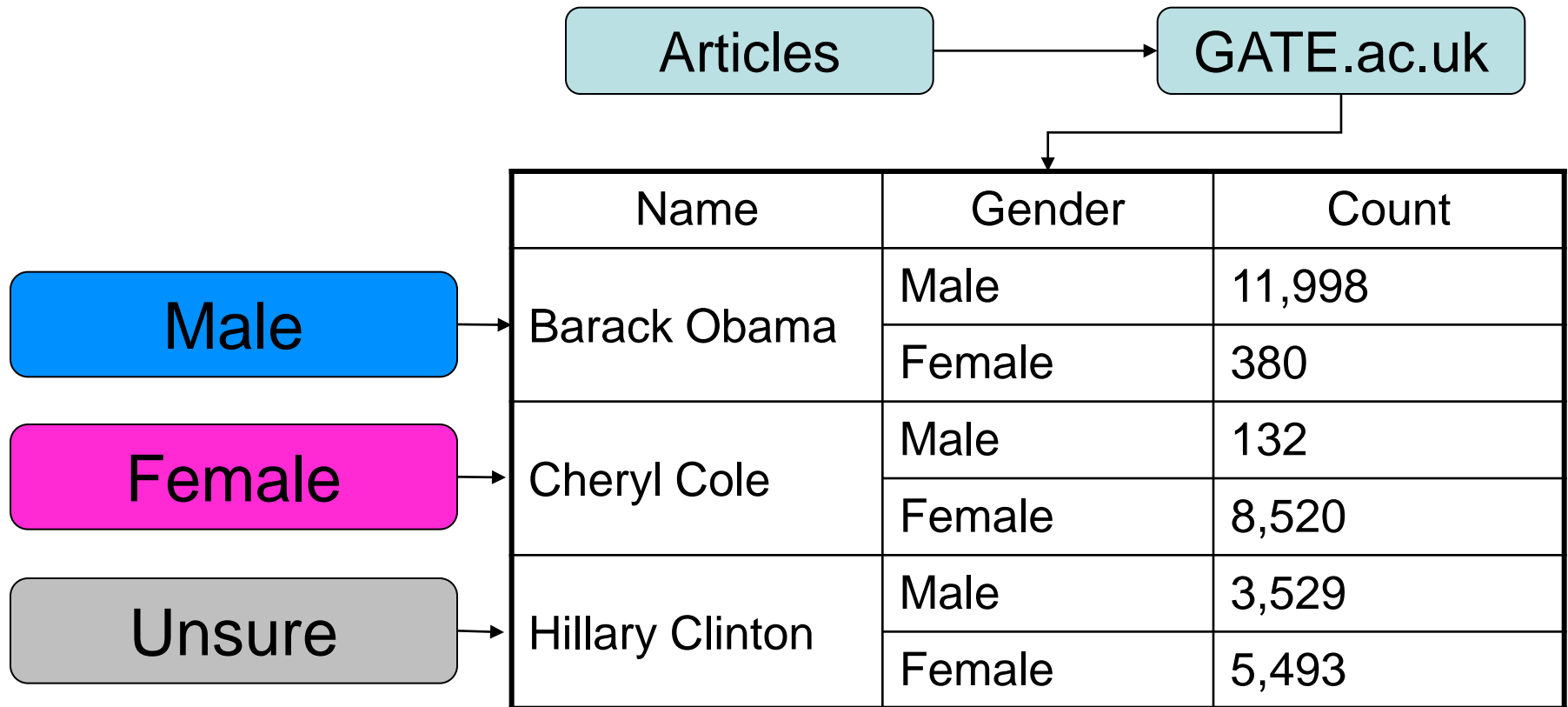
90.0–100.0: easily understandable by an average 11-year-old student

60.0–70.0: easily understandable by 13- to 15-year-old students

0.0–30.0: best understood by university graduates



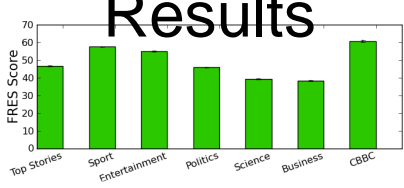
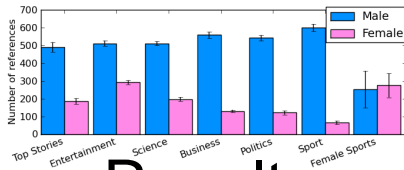
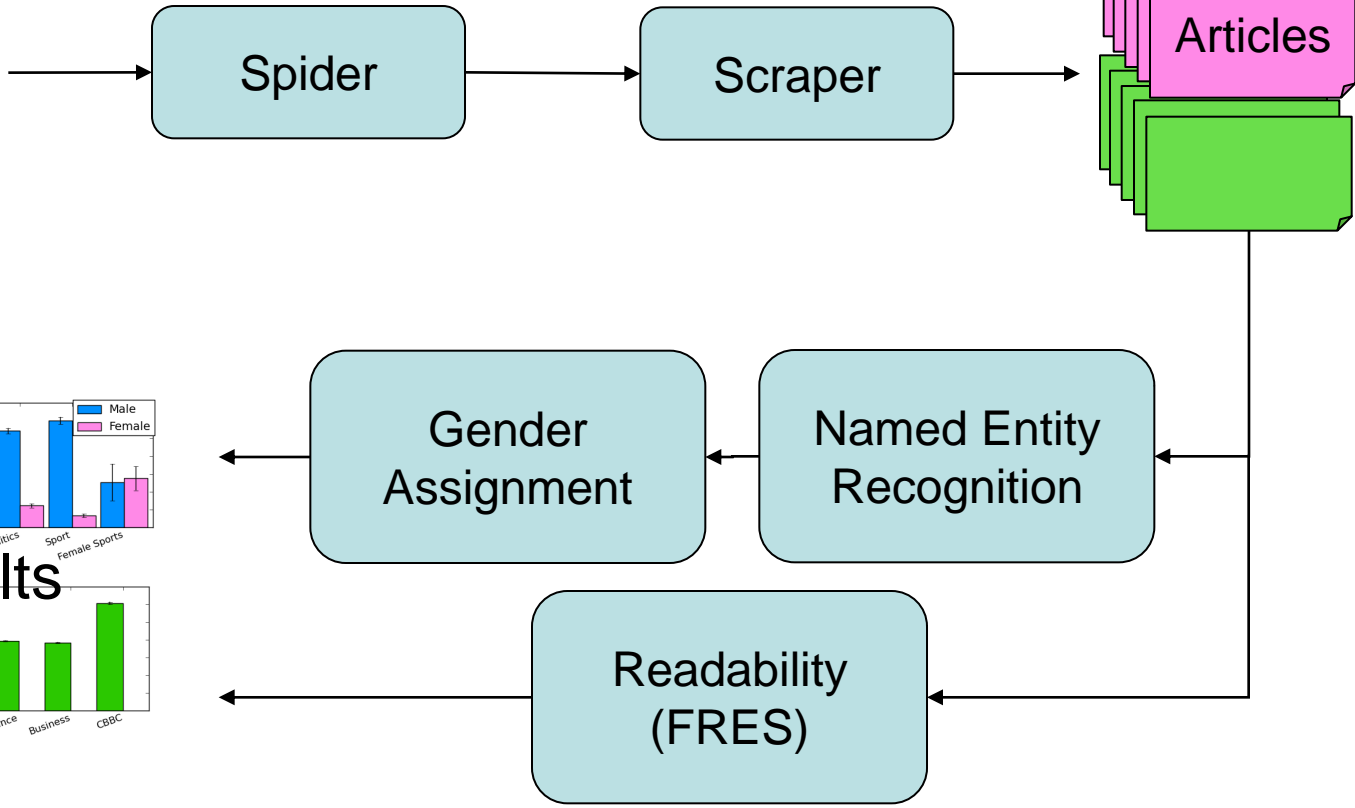
# Gender Labelling

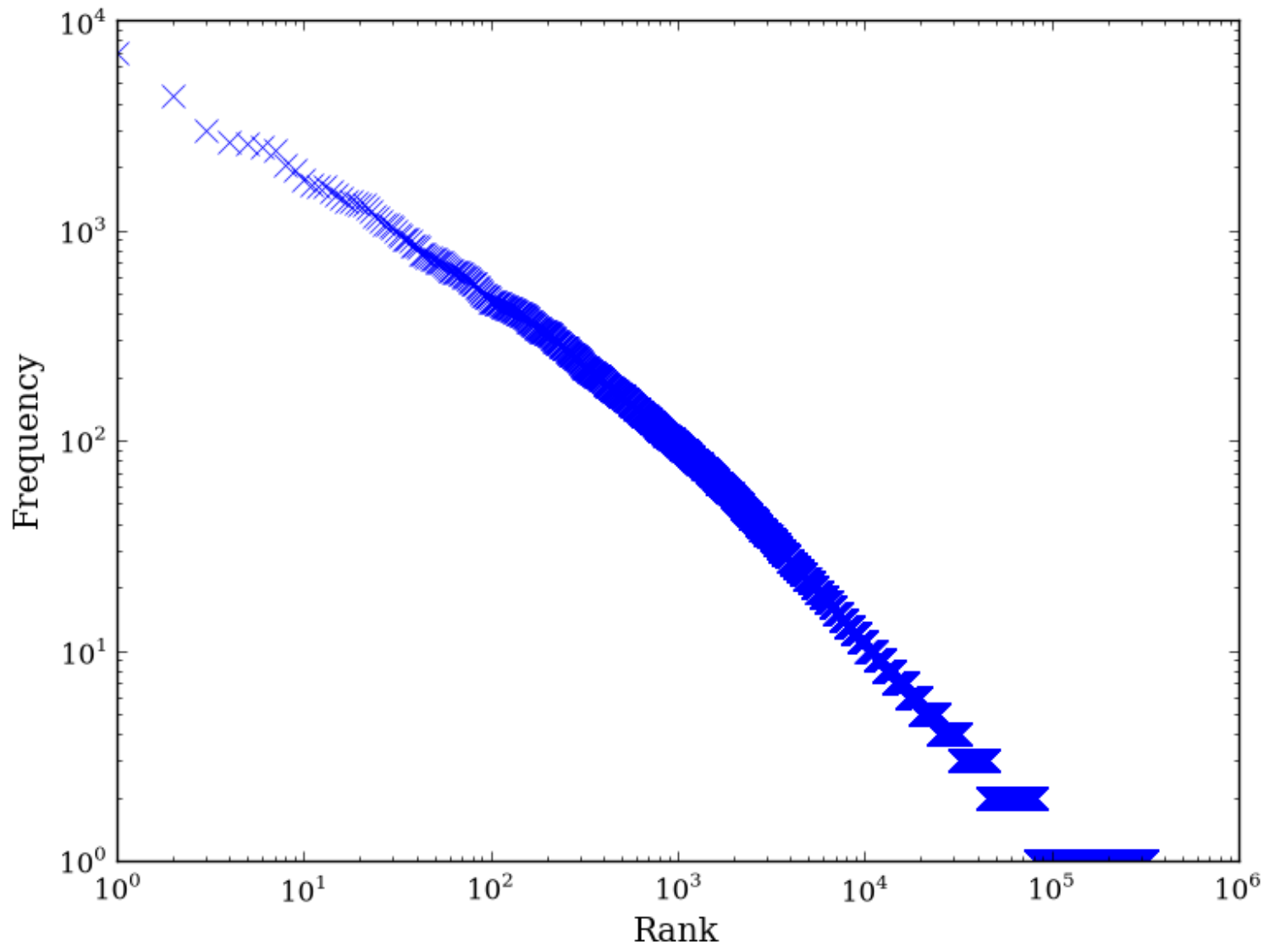


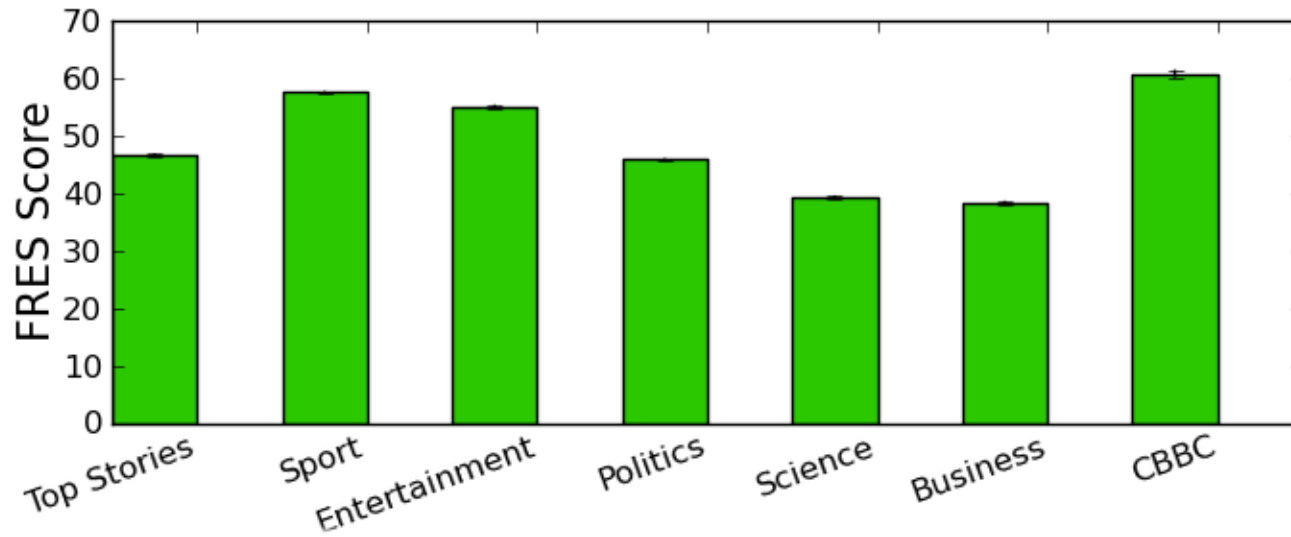
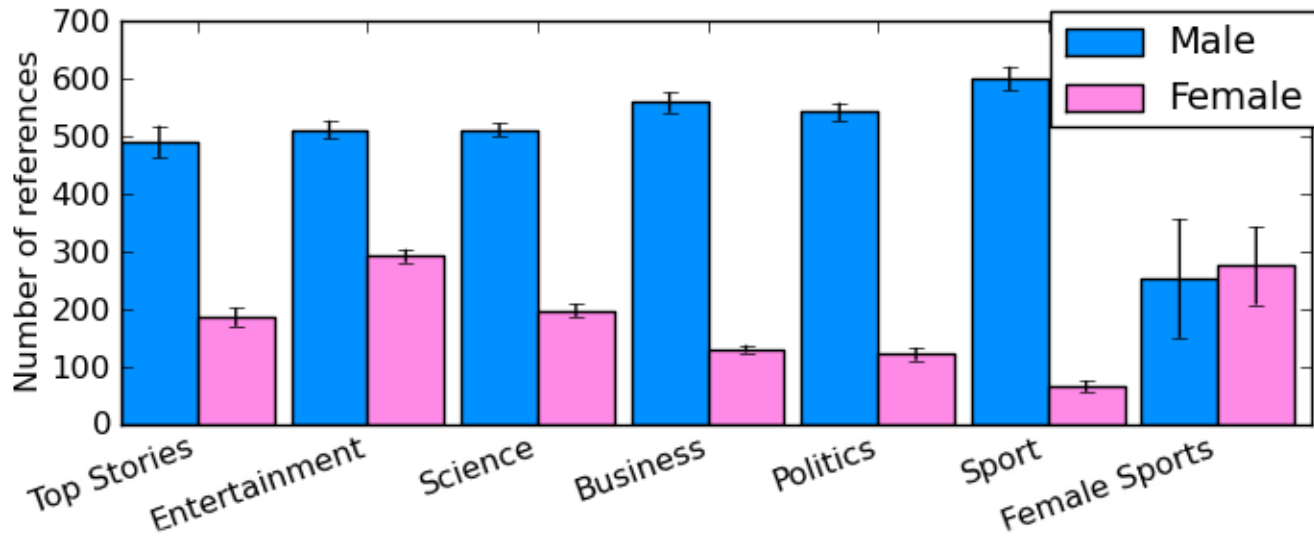


# How can we automate this?

BBC Sport   Sport Homepage   UK Edition	
Red Bull out on own - Schumacher - Former Formula 1 world champion	
Montepierre plans Garcia pay talk - Colin Montepierre plans to talk	
Ferrari set to hold on to Massa - Ferrari say they are almost certain to	
Mourinho 'a perfect fit for Real' - Jose Mourinho's style of play will fit	
England v Bangladesh - England reach '100' after being pulled	
French Open - The lady begins at a shrewd Roland Garros	
Hill to be named Worcester boss - Worcester Warriors are expected to	
Exeter seal Premiership promotion - Gareth Stevenson scores 24 points	
Cardiff's 60m takeover approved - Cardiff City shareholders rubber-stamp	
Hooker 60th while players' award - Gloucestershire's 20th birthday	
Cottrell leaves Notts County job - Steve Cottrell, who has been linked	
Woods to return at Memorial event - World number one Tiger Woods	
Thursday's gossip column - Peter Milan in for Fabio Capello, plus other	
Jockey Holland to miss Derby ride - Jockey Danny Holland will miss	
Radcliffe looks beyond 2012 Games - Paula Radcliffe rejects the prize	
Palingini sacked by Real Madrid - Real Madrid sack coach Manuel	
Second player targeted by fixers - The England and Wales Cricket Bo	
Pompey creditors offered 20% deal - Portsmouth's creditors are to be	
Trial by television - BBC to show stewards' inquiries	
Red Bull can be beaten - McLaren Championship leader Red Bull	
Strauss warns players over fixing - England captain Andrew Strauss	
Cheryl and Ashley finally meet - Who's missing being apart, Cheryl an	
Kerry Katona signs Peter Andre's manager - It's not been an easy ca	









# Summary

- Social scientists study the media.
- Online media is easily accessible to machines.
- We have automated collection and analysis of news.
- Demonstrated biases by topic:
  - Gender.
  - Readability.





University of  
BRISTOL



# Evaluation against Freebase

		Our Label		
		Male	Female	Unsure
Freebase Label	Male	25,142	411	6,108
	Female	215	5,427	1,177



University of  
BRISTOL



# Links

- <http://www.cmswire.com/cms/web-publishing/who-reads-technology-blogs-a-dirty-analysis-of-reader-demographics-002631.php>

