

>>> bit.ly/sjf-scraping <<<

Introduktion till screenscraping

Jens Finnäs
Journalism++ Stockholm
@jensfinnas
22 september 2015



Agenda:

Grunderna i HTML

Introducera XPath

Pröva på scraping med pluginet Scraper



Vad är screenscraping?



▼ Ledamöterna

Bokstavsordning

Partiordning
Ersättare
Kultningsmän
Valkrets
Ledamöter som avgått
Riksdagsorgan

► Invalda ledamöter 2014

► Socialdemokraterna

► Moderaterna

► Sverigedemokraterna

► Miljöpartiet

► Centerpartiet

► Vänsterpartiet

► Folkpartiet

► Kristdemokraterna

Bokstavsordning

Bläddra i bokstavsordning

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Å Ö

I riksdagen finns 349 ledamöter. De väljs vart fjärde år då det är allmänna val i Sverige. Här hittar du ledamöter och ersättare i kammaren.

A

 **Abdu, Said (FP)**
Folkpartiet
Västra Götalands läns norra, plats 339
Telefon: 08-7864079
E-post: said.abdu@riksdagen.se

 **Abrahamsson, Maria (M)**
Moderaterna
Stockholms kommun, plats 206
Telefon: 08-7866418
E-post: maria.abrahamsson@riksdagen.se

 **Acketoft, Tina (FP)**
Folkpartiet
Skåne läns västra, plats 220
Telefon: 08-7865963
E-post: tina.acketoft@riksdagen.se

 **Adan, Amir (M)**
Moderaterna
Stockholms kommun, plats 209

Det här gör en ledamot

Som ledamot har du flera olika uppgifter. Ledamöterna beslutar om lagar och statens budget. Dessutom ska de kontrollera regeringen och vara med och påverka utrikespolitiken.

Ledamöterna representerar sitt parti och sin valkrets. Uppdraget innebär mycket kontakter med väljare och representanter från olika delar av samhället. Det kan handla om besök på skolor, arbetsplatser eller organisationer.

- Så arbetar ledamöterna
- Så funkar riksdagen

Ledamöternas villkor

Ledamöterna har sitt uppdrag dygnet runt, året runt. Det betyder att de inte har semester. I stället för lön får ledamöterna ett arvode. En ledamot som avgår före 65 års ålder efter minst tre års sammanhängande tjänstgöring har rätt till inkomstpåslag.

- Ekonomiska villkor
- Åtaganden och ekonomiska intressen
- Så funkar riksdagen



✓	1	Abele, Anton (M)	Moderaterna	Stockholms kommun	141	Telefon: 08-7865962	anton.abele@riksdagen.se
✓	2	Abrahamsson, Maria (M)	Moderaterna	Stockholms kommun	314	Telefon: 08-7866418	maria.abrahamsson@riksdagen.se
✓	3	Acketoft, Tina (FP)	Folkpartiet	Skåne läns västra	153	Telefon: 08-7865963	tina.acketoft@riksdagen.se
✓	4	Adan, Amir (M)	Moderaterna	Stockholms kommun	280	Telefon: 08-7861464	amir.adan@riksdagen.se
✓	5	Adelbo, Christer (S)	Socialdemokraterna	Skåne läns norra och östra	39	Telefon: 08-7864709	christer.adelbo@riksdagen.se
✓	6	Adolfsson Elgestam, Carina (S)	Socialdemokraterna	Kronobergs län	293	Telefon: 08-7864159	carina.a.elgestam@riksdagen.se
✓	7	Ahlberg, Ann-Christin (S)	Socialdemokraterna	Västra Götalands läns södra	302	Telefon: 08-7864398	ann-christin.ahlberg@riksdagen.se
✓	8	Ahlgren, Anders (C)	Centerpartiet	Dalarnas län	22	Telefon: 08-7864190	anders.ahlgren@riksdagen.se
✓	9	Ahlin, Urban (S)	Socialdemokraterna	Västra Götalands läns östra	44	Telefon: 08-7864723	urban.ahlin@riksdagen.se
✓	10	Ake, Christer (M)	Moderaterna	Skåne läns norra och östra	95	Telefon: 08-7861941	christer.ake@riksdagen.se
✓	11	Alfsson, Thoralf (SD)	Sverigedemokraterna	Östergötlands län	119	Telefon: 08-7866459	thoralf.alfsson@riksdagen.se
✓	12	Almqvist, Erik (SD)	Sverigedemokraterna	Västra Götalands läns norra	301	Telefon: 08-7866479	erik.almqvist@riksdagen.se
✓	13	Al-Sahlani, Abir (C)	Centerpartiet	Stockholms kommun	206	Telefon: 08-7864407	abir.al-sahlani@riksdagen.se
✓	14	Amin, Jabar (MP)	Miljöpartiet	Västerbottens län	139	Telefon: 08-7866430	jabar.amin@riksdagen.se
✓	15	Andersson, Anders (KD)	Kristdemokraterna	Kalmar län	255	Telefon: 08-7864447	anders.andersson@riksdagen.se
✓	16	Andersson, Jan R (M)	Moderaterna	Kalmar län	219	Telefon: 08-7861006	jan.r.andersson@riksdagen.se
✓	17	Andersson, Johan (S)	Socialdemokraterna	Östergötlands län	148	Telefon: 08-7865315	johan.andersson@riksdagen.se
✓	18	Andersson, Jörgen (M)	Moderaterna	Kalmar län	290	Telefon: 08-7866419	jorgen.andersson@riksdagen.se




Protokollen

Alla svenska kommunstyrelseprotokoll.
I en öppen, sökbar databas.

Notera att databasen inte är komplett ännu. Protokollen är fortfarande under utveckling.



 Sök på nyckelord

Du kan till exempel söka på [skandal](#), [missförhållande](#), [SVT](#), [sommarjobb](#), [fotboll](#), [bredband](#), [kärnkraft](#), [Reinfeldt](#), [förskola](#), [demonstration](#), [nudistbad](#), [jäv](#), [hot](#).

I den här databasen hittar du kommunstyrelseprotokoll från Sveriges 290

@jenstinnas

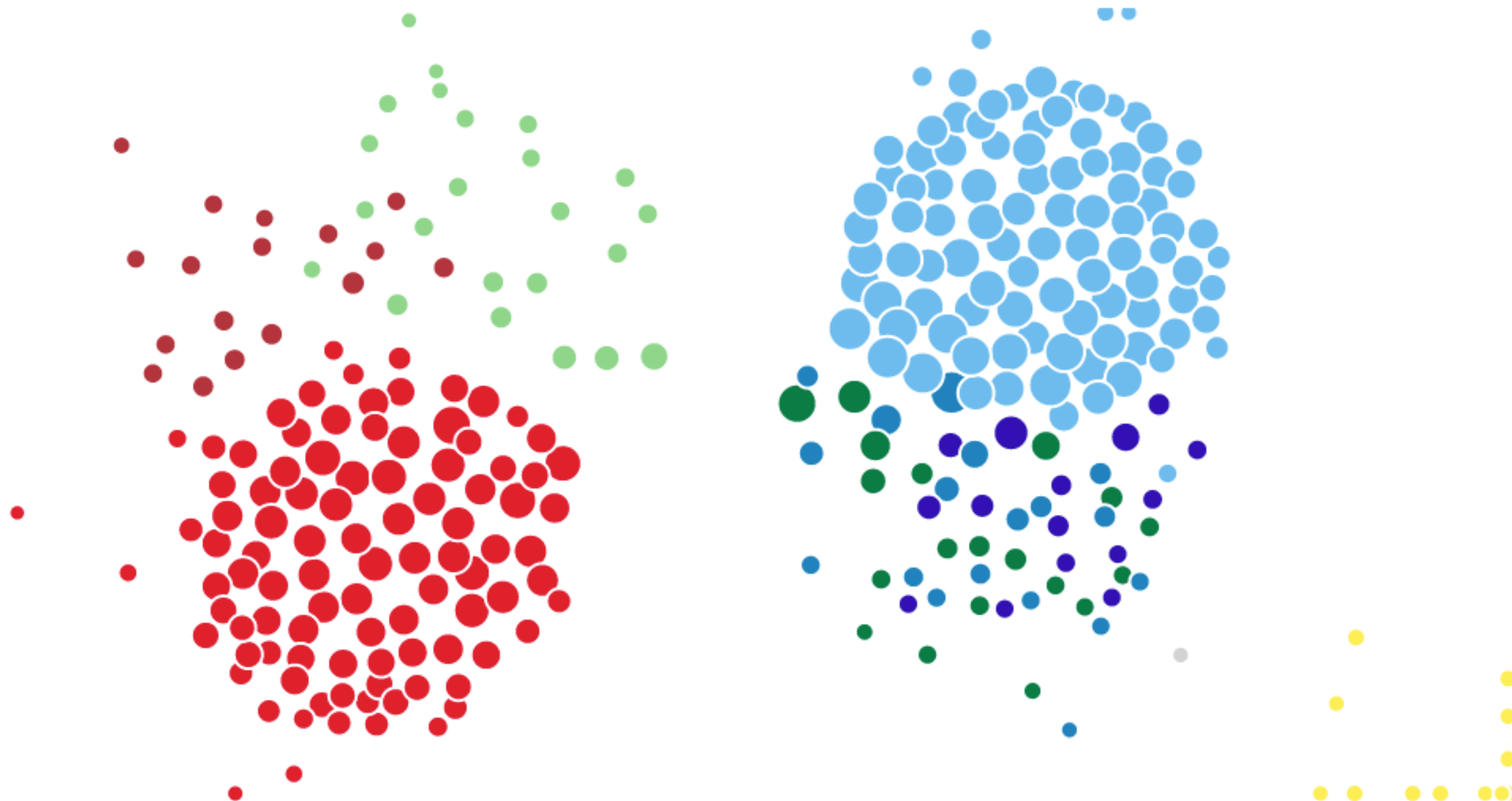


	Petrus församling	
98	EKREM Carola	4
93	SANDSTRÖM Caroline	4
85	KOLLIN Susanne	4
76	AUDAS-KASS Amanda	1
67	BÄCKMAN Christina	4
47	LINDÉN Göran	2
45	DIDRICHSEN Peter	1
38	ÖSTERHOLM Klaus	3
34	FERNSTRÖM Kristina	1
29	JÄRVELÄINEN Jan	2
28	PERRET Christoffer	1
25	NYSTRÖM Mikael	2
24	PIMENOFF Susanna	2
23	FREDRIKSSON Stig-Olof	2
23	VON WEISSENBERG Joachim	3
22	FERNSTRÖM Ben	1
22	PERRET Johanna	1
22	SUNDIUS Carola	2
17	NYSTRÖM Annika	2
15	LINDSTRÖM Peter	1
15	MIKKONEN Tommy	2
13	WALDONEN Tom	1
12	ANTTILA Tuomas	2
10	JÄRVELÄINEN Solveig	2
9	SUNDIUS Tom	2
7	MIKKONEN Catarina	3
5	GRÖNING-JÄNTTI Gun-Britt	1
5	LINDEDAHL Kaj	1
5	MIKKONEN Juha	2

Vem är vän med vem i riksdagen?

P3 Nyheter har kartlagt vänskapsbanden bland riksdagsledamöterna på Facebook. **Klicka på bollarna på kartan** för att se vem som är vän med vem.

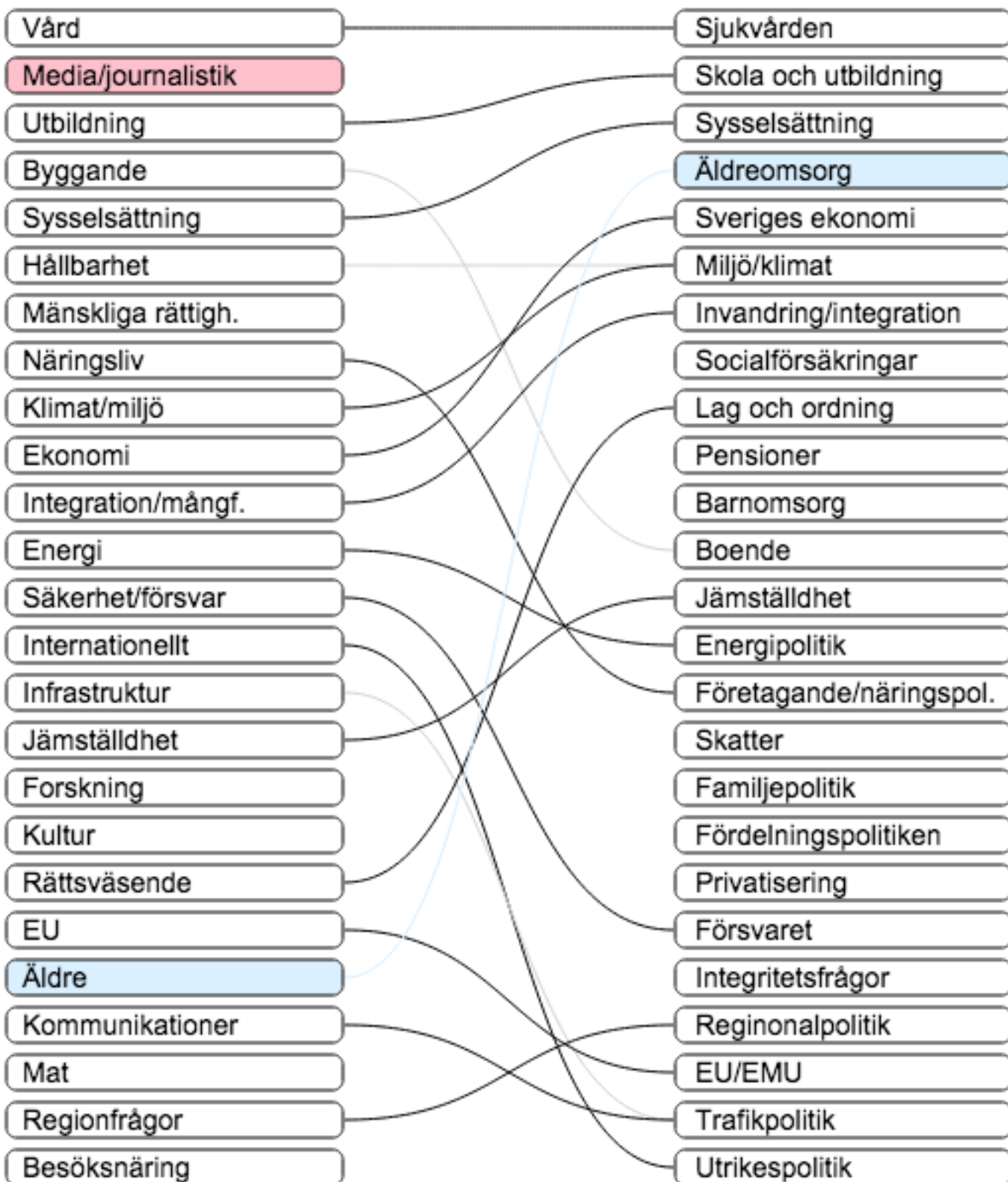
RESEARCH OCH GRAFIK:
JENS FINNÄS, SVERIGES RADIO





Flest arrangemang i Almedalen

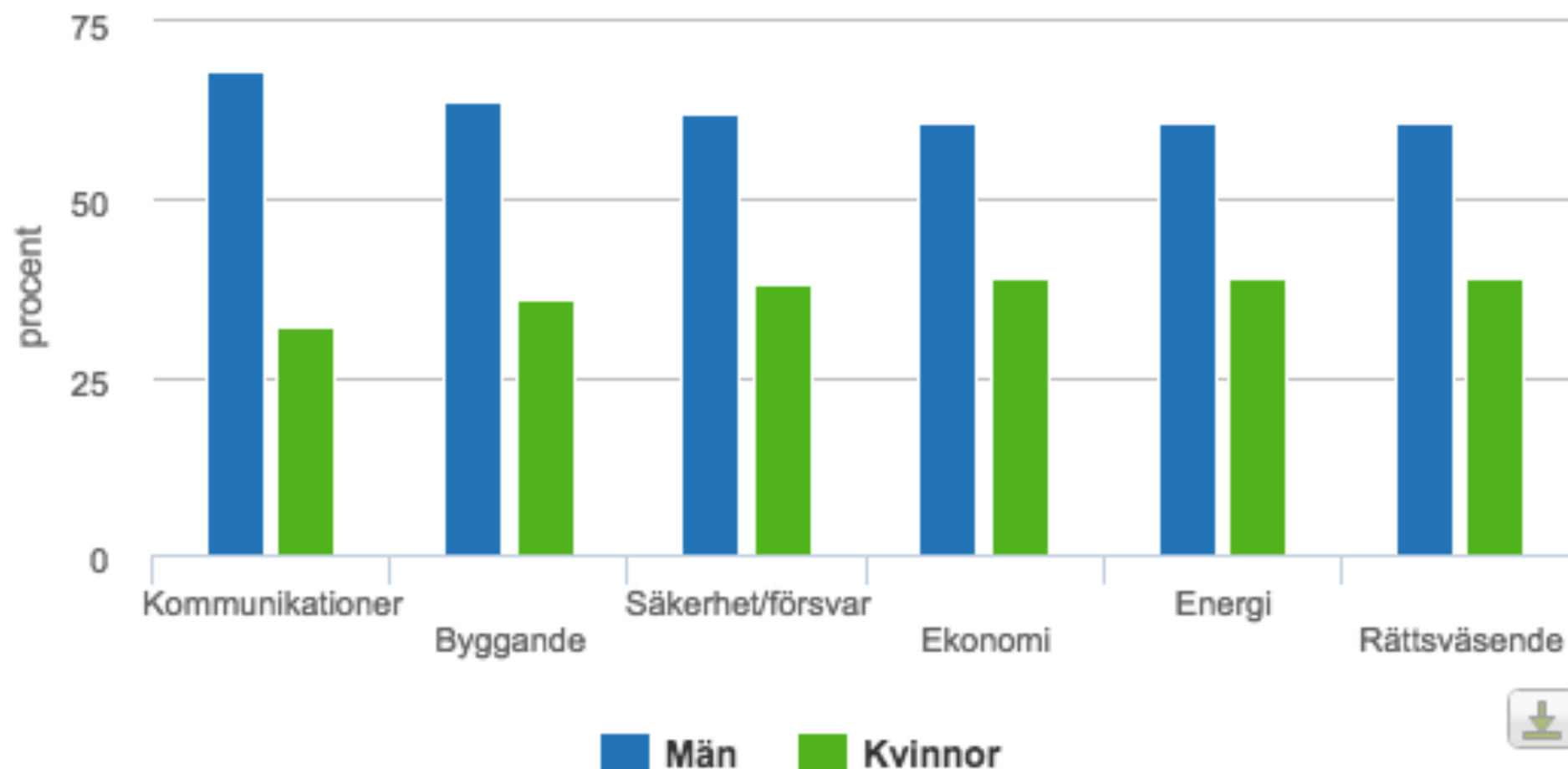
Väljarnas viktigaste frågor





MANLIGASTE KATEGORIerna

Andel män och kvinnor som deltar i evenemangen på Almedalen, i de manligaste kategorierna





Reporting Examples

I asked the folks in NICAR-L about stories that they couldn't have written without scraping and got a nice list:

- Micheal Keller's team at Daily Beast used web scraping to track [presidential candidates' ground games in 2012](#). Bonus: They [wrote about the process](#) when they were done.
- Reporters at The Baltimore Sun scraped Baltimore's finance department website after city officials said it wasn't possible to provide reporters with access to property tax records. Their [series on the unintended consequences of a local tax credit](#) would not have been possible without that data.
 - Megan Twohey and the data team at Reuters scraped five years worth of archives of multiple Yahoo Groups for their five part series on [disrupted adoptions](#). This was an 18 month project -- that included a great deal of data entry to clean and sort the scraped posts.
- ProPublica's Dollars for Docs
- For a Marketplace story on [court debts in Philadelphia](#), I scraped 200,000 records from a public records database so that I could analyze the debt data for Amanda Aronczyk, the reporter on the larger story.
- S. Heather Duncan at The Telegraph (Middle Georgia) got an assist from outside the newsroom to pull thousands of PDFs of state records on day care inspections. They had added work to parse those PDFs, but ultimately found that many centers are allowed to continue operating despite very serious safety violations and deficiencies. The state had insisted that the inspection reports were generated manually and that there wasn't any central database.

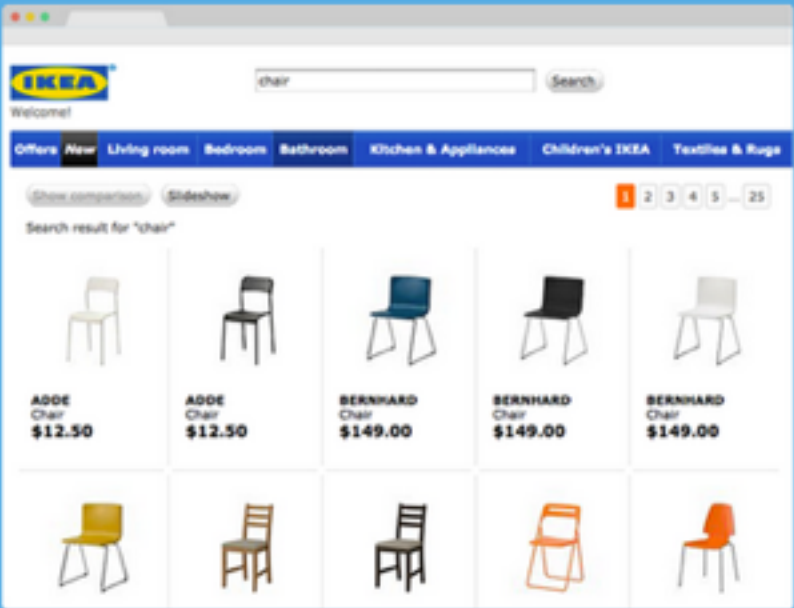


Instantly Turn Web Pages into Data

No Plugin, No Training, No Setup.

Enter a URL to a list page

Try it Out



#	link_1	prodimg_image	prodname_text	prodprice_price
1	ADDE Chair \$12.50 Unit price		ADDE	12.50 USD
2	ADDE Chair \$12.50 Unit price		ADDE	12.50 USD
3	BERNHARD Chair \$149.00 ...		BERNHARD	149.00 USD
4	BERNHARD Chair \$149.00 ...		BERNHARD	149.00 USD
5	BERNHARD Chair \$149.00 ...		BERNHARD	149.00 USD
6	BERNHARD Chair \$149.00 ...		BERNHARD	149.00 USD

COPY TABLE DOWNLOAD... GET API...

Show Examples



kimono

Turn websites into structured APIs from your browser in seconds



Get started, click to install

@jensfinnas



Vad är HTML?



```
<strong class="name">  
    Jens  
</strong>
```