

# Web crawling Technology and benefits

IEA webinar
December 3<sup>rd</sup> 2020

Kasper Mogensen, Big2Great, Denmark



## **Big2Great - Kasper**

Kasper Mogensen
Denmark
Co-founder of **Big2Great**(with Troels Fjordbak)

Spend 24% of my life crawling the web

Master of Business Administration and Computer Science, Copenhagen Business School

Masters thesis about web crawling



## **Big2Great**

Mission: From big data to great value

- Software and statistics consultants
- Web crawling, inhouse large-scale crawler (BigCrawler)
- Statistical models builder
- Energy efficiency & product safety



## Why is web crawling important?



000

from #PostCorona

Online sales grew as much in 8 weeks as it had in the decade before the pandemic

Oversæt Tweet



10.16 PM · 29. nov. 2020 · Twitter Web App

(https://twitter.com/profgalloway/status/1333157834138259456)



Background

#### WHAT IS WEB CRAWLING

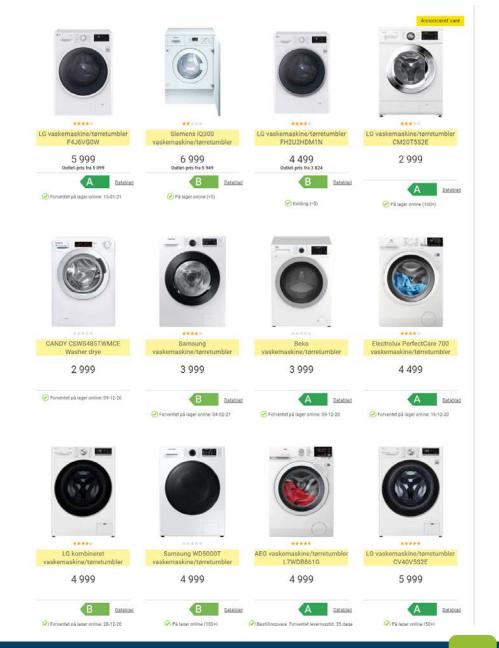


## What is web crawling?





- 1) Page with models
- 2) Visit all links





- Download page as text
- 4) Extract info for text

#### Modelbeskrivelse

Produkttype	Kombineret vaskemaskine/tørretumbler
Type vaskemaskine	Frontbetjent
Placering	Fritstående
Producentens varekode	F4J6VG0W

#### Kapacitet, forbrug og strøm

rapacitot, reibiag og etipin	
Energimærke 🕢	. A:
Vaskeevne (EU) 🚳	· A:
Centrifugeringsevne (EU)	В
Forbrug kWh (vask) 🚱	1,17
Forbrug kWh (vask/terring)	6,12
Forbrug kWh/ár (vask)	234,00
Forbrug kWh/år (vask/tørring) 🚳	1 224,00
Vandforbrug L/cyklus (vask)	140
Vandforbrug L/cyklus (vask/tørring)	140
Vandforbrug L/år (vask/tørring) ❷	28 000
Nominel kapacitet vaskecyklus (kg)	9,00
Tørrekapacitet (kg)	5,00
Tromle volumen (liter)	59
Centrifugeringshastighed (rpm)	1 400
Lydniveau vask (dB) 🚳	55
Lydniveau terring (dB)	56
Lydniveau centrifugering (dB) 😝	75

#### Funktioner

Forskudt start	Ja
Tromlebelysning	Nej
Indikator for resterende tid	Ja
15C koldvask	Ja
Hurtig vask (min.) •	30
Dampfunktion	Ja
Automatisk dosering af vaskemiddel	Nej
Wi-Fi	Nej

```
class="tab-data-row tab-specs-row show" data-tab="tab-specs
div class-"tab-slot" itemscope-""
         <dt class="any-1-1">
            <span class="spec-section-title">Modelbeskrivelse</span</pre>
          <dd class="any-1-1">
              Produkttype
                            <span class="abbr">
                                Type vaskemaskine<i class="fa fa-quest
                                <span>Viser om vaskemaskinen er top-
                        ctd class="any-3-4 5-1-2 xh-highlight" data-mc
                        <td class="any-3-4 5-1-2 xh-highlight" data-md
                           Producentens varekode
                        <td class="any-3-4 S-1-2 xh-highlight" data-md
           cspan class="spec-section-title">Kapacitet, forbrug og str
                             <span class="abbr";</pre>
                                Energimarked class-"fa fa-question-c
                                 <span>Viser produktets energiklasse.
                        <td class="any-3-4 5-1-2 xh-highlight" data-mc
                         <td class="any-1-4 5-1-2" data-md-name-id="31
                               Vaskeevne (EU)<i class="fa fa-question
                                <span>Viser hvor grundigt produktet va
                        <td class="any-3-4 5-1-2 xh-highlight" data-mc
                             <span class="abbr">
                                Centrifugeringsevne (EU) i class-"fa
                                 <span>Viser hvor effektivt produktet
```



# Web crawling extracting raw data

Named taset of		I to Proce	an Prosession	I	lane.	ve l	Crespitate da	ing 13 dish vid	Restletações	Thereware:	(willing acco	Typ in product	1991	hedd (bur (re) (re)	Onongresialistical	Vositorining	Energifications for 60 °C- more standard root hall yielding	Energlishing to 60 °C- termobic granned W logs its left verified in §15 ring	Transplorting 60 °C- tomagning arrest acpe itselet and belt Mitting	Programmighed to 86 °C- monolprogrammid toge Notest and be follows:	Programmed for EE 1C. semajorgrammed block fluids	Programmaryland for 60.10- numerity opinional logo Audiet	Destination of the last of the	thesp
DASS. IDS2109-bits Josephine Report of Participation and American Services					tile Sprods	20 Hut.	Dise Scott	loss to total	aco.	156	Sine Scrowl	Esting colemnia		.50 Non-Surseil		lika todal	kope fluidet vochel lythoug nie trebi	loge-balet verhein häring	acge itself sed halv full ring	and beilipithing	and beer fulfilling	and clobin lyaming	1	lim trad
1969 1922 (1961) in investigation from object which review reference is early firm.  1942 (1961) in investigation from experience in the experimental review of the experimental review	28300 - 177821911 -	FIN FINEDA	eper www.xcory.m	B++1	Q 7	03 Hule 53 Hule	141	72	80,00	re.	12	replaneering	25,00 (0	100 90.00	1 100,00 100 Units	11700 Due hodel	0.65	(3)	E.67	att) oke teler	261	DH SASE	Nortedo	-
SIZE X770 to the apparent figure of their and entered their tells	C124   1238-18-	1,000 Fronts	#8#7 50K140309FU	ý –	tila fancesit	00 Phile	line fundet	Tide lander	0.8	0	the fundet	Komperer unriamence Am	etmbled270 Fil	SE Jose fundet	inia tundet	itte finder	like forcet	lita Sedel	the Lobe	100	To Lede	the Undet	Sira Lane	line tyroet
SMRS (NOTSSHED) their appears from Library and construction with the contract of the Contract	0481 53781461 06599 53781591	In the Prositio	eter Montality	A Canada	las lu	90 Huid 90 Huid	140	72	Nie todel Ni Ob	0.0	Drive Tyrocall	Karonere vicentaccretic	85.00 PG	60 - Dies funder	1 608.00	Nie bode	Interfered	ISS Lode:	DIA Lote:	total fundar	119 1,0 del	tive funcion	tros tuntes	-
1302100Mtd 2 minut siggarden dirproduzieren verbreatsine Web/9 0388 100220Mtd 2 minut siggarden birtroduzieren verbreatsine Web/9 3858 100210Mtd 2 minut siggarden birtroduzieren verbraatsine Web/9 3858 100210Mtd 2 minut siggarden birtroduzieren verbraatsine Web/9	NS40 2778 1 Gen	74% Fronts	elare WD19HS40DV	Α		93 HVE	676 Startootel	Inia tanta	ing Logic	14	Rins Turnost Date Subset	Kamonere' vocama stretu: Komonere' vocama stretu:	reumbied x 10 El	50 50 60	ico sindet	ine tedel	line fundet	ing \$1640 like \$1640	Die Lode	tile funder tile funder	Note \$1000 Dise \$1000	tion funced	1	Non-Solder
Debt. 1000 100 to Joseph State of the Control of th	NI226 BITTS (6a)	1900 Fortow	W. WHISDECHORE	jare .	14 6	00 39/05	215	EJ.	51.00	54	il .	Michigrania (ne	21.00 75	50 47.50	200.00	Tru65	177	174	U.S.C.	DX:	rite Come	id.	1	THE STREET
1004 UCC10 TO TWO OPENING STREET, THE OWNERS OF THE PROPERTY OF THE OWNER, TH	STEER DESTROY	PMAL Fronts	spre handwings	Quest .	97 E	03 3144	194	11	91,00 81,00	88	y .	Packanapins	85.00 BX	0.00 New White	1.000,00	7522P	W 0.5	590	0.44	une len	276. Set	170		
CONTROL CONTROL TO CONTROL OF THE PROPERTY OF	Take STATE	4418 Kerme	at MSPST40077	Arres .		00 HAR 40 HAR	100	19	50.00	ě.	55	Visionesine Weterteiline	14.00	00 MO DO	700.00	930	0.61	18	0.40	08	No	- 100		
				(Area	54 6	50 PH/E	197	13	BC 00	56		Vacuation to Vacua	94.65 85	1.00 SS-00	1 400.06	9556	91.04 91.00	5.5V	D.An.	290	200	DN:	1	-
26/86 17/2/17 http://www.elaparton.sc/product/natcolerrationacces/WMHB 26/98 17/2/19 http://www.elaparton.ch/ps/duct/natcolerrationacces/WMHB 26/98 17/2/19 http://www.elaparton.ch/ps/duct/natcolerrationacces/WMHB	W/748 93762461	7 990 Fromba	etert MARKENTASON	Janes .	43 2	00 94vid. 00 94vid.	190	14	84,00	0.	9	Variancière Variancière	84,50 (3	00 50 00 100 86 00	504.60	91020	9.02	100	(14)	500	270	100	ř.	
				Arre	H7 [73	53 Sari	100	72	80.00	96		Yearnorn	85.00 20	1.00 86.60	trop fundat	9516	hive forcies	the Lorise	lika furtie	Ma Critis	Title Emile	hie traiei	1	
Colors	961/gc (02/82/5/1	S 900 Fronts	esert MOSFINE esert MTDSSSAK	(Annual Control of the Control of th	12 P	DO Hotel	170 Day Lodel	Distributed	64,00 67.00	64 60	Disa Rendel	Vaccanacies Explanel estamacentar		00 60 00		SUGAR Discharge	R32 Dan broad	176 Statement	II 90 III a Louise	Tita Loder	NIS LOAD	190 Biss bendef	2	Ne Noct
SSEE 190719 to mon reparter depodatificament reterration Wildel	Taw (1012Sc)	169 Foots	eper SARRETANION	JANE .		80 Plut	10	9	65.96	63	9	Vision cooking	24.01	50 50.00	1 420 (0	1020	162	1.60	0.02	75	776	170	1	
2006 ISSUED TOWN ADDRESS TO SELECTION AND ADDRESS TO SELECTION OF SELE	ERCOSE TOUTES BUT	17 G37 - BotterBru	WART SEWMENCHES	(Aprel	51 6	00 HVW	795	77	94,00 94,00	(K)	i i	Vacamoskina Vacamoskina	25.00 86	1,50   54.50   00   Residential	1 608.06	15499 15499	D 01	1.57	0.52 0.54	100	P10 P24	196	-	-
Politic 10300 Sylfamin times amounted through other present commences to a Total Co.	47+19 BOTTESSAT	la 1000 Toraman	tent TEMATAZENE	garre.		63 Hvid. 55 Hvid.	974	77	200	Q		Valverselline Sociamentos	#8.00 et	(0) 80,01 (0) Ship bridge	1300.00	5256	0.86	instructi	0.14	10	104	100	D.	
MOST Plate that departer to be at the area and an extended to the area and a second of the	00°w/w 00783 (#1	H 000 Fronts	when previous	Ares	00 0	1.00 PH/M	100	72	51 XX	79	13	Variations	120 0000	00 H0 00	1 500.00	11219	0.95	E34 E30	0.72	D16	228	120	-	
18256 INC220*** The representation of the previous research of the State of the Sta				Janes.		(x) 9946 (x) 1146	594	111	MI,00	0	9	Vac-ancient		00 80.68		70462	6.03	575	0.00	140	200	219	1	
		1600 Forth	when REFERENCES	jaren .	11 (6	AS Street	174	TA .	600	0		Vaccement in	9.5 10 10	St. Die brahe	400.00	99846	0.54	1.64	6.94	100	140	100	Ě	
DISSECTION OF THE PROPERTY OF	H3936 M183801			(A)		02 94,68 53 8190	220	28	NC.00	m Distance	B.	Hacismackins	85.00 KG	00 89 00 00 89 00	1,200,00	52(58) 56(340)	Dide fundet	130 tests	ELECTRICAL STREET	in total	Distance	Ina timor	E	
2004 100020 http://www.eligenten.dr/poduth-berare/retire/MD79 2004 1000211 http://www.eligenten.dr/poduth-berare/retire/MD79	85445 \$37843et	5.430 Fronts	enert MCCOMeditive	15A	José funcado.	00 Huit A0 Huit	Side funded	ine base	61,00	itte tunner	sex tyrost	Kananeri vedi erasonaten	WILLIAMS CO. EX	.00 55.00 .00 New Number	NIVE SAGE	metiner	He forcet	son Sinder	ice tinde	instante	tion s, made	los facet	*	the target
Tacket 1 (0.7777 mm) consequent forces or further communicate destruct BETC	ALTA: CHTRASCT.	HARL Fronts	wywrt &BD16AZ	AA.	59 (1)	52 Wá	222	R	b1.00	IN BOOK	-	1/80/90/00/100	28.00 56	00.00	90,008.7	70340	one total	INVESTIGATION .	ION SAME	Die tyte	HER TOOK	she timber	1	
PARTY IN SECURITION TO A PROPERTY OF THE PARTY OF THE PAR	MANY DOTAGET	Date Franks	etert Philipsons	9041	No. le	00 Hult	196	34	84,00	NI.	9	Visitemastine Visitemastine	85.00 EU	.50 Sine funder	1.004.00	11600	0.06	1.60	0.56	(A)	100	170	1	
		1 kill Kome	## ENGINCOS	(Ares		03 Hyut	10	7	57,00	ü		Modernokine	85.60 50	756 S7.7e	200.00	5000	176	197	0.41	90	Oi .	100	1	
1966 (ALIZZANIS PRO REPRESENTATIONAL PROPERTY AND PROPERT	700301   SSTB4561 14000   SSTB4561	1400 Rocks	elere WASHARMED	A+++		03 94/0E 03 81/0E	Trial	FIRE SANSE	84.00 51.00	C) Die bride	Total Syndet	Vacionación Kanadere vacionaciónstan	represents on the	100 45 08	Lice funder	Disc tanks	prov.	1.71 (0) \$446	Dist Lines	TITLE TURBET	10x 1046	Tro Rese tunces	1	Fire Sandet
1996 University Investigation for very local program of participation of VPT 1996 UNIVERSITY of the appeals for the Arthonism several participation of VPT INVESTIGATION OF THE ARCHITECTURE CONTRACT CON	Wrose Billsaler		elet PHSMD -	line fundet		50 file 50 Head	like fundet	ion wade	W (X)	66	tine funcet	Veriandina	187.667	00 this first	1200.00	DIA ECCY	13	III LANG	It's Lode	N.	15	tee forcet	N/9 1,666	Ree fundet Ree fundet
18-25 18-222-9-55 Print disperse from the second and an extended to the second and the second an	Wall Killson	2 900 Proche	adart Lotzwelleg.	jare :	les to	45 3544	1/1	79	81.00	81	8	/voursmodeine				9609	1.00	2.62	0.95	900	No.	100	-	
CARRIED TO SECURE A CONTRACT OF THE PROPERTY O	20482 (2789 fet	CAND FRANK	ASSET FOUNTAINE	SALES .		00 Ned	140	54	N2.00	64	10	Notionacting	\$6.00 %	(50 54 00	1 400.00	SQ10	0.48	(3)	0.07	100	290	16	1	
1698 180771 etc. Prince opposes organization are maken as the 5% to prince 16,222 etc. Prince opposes or productivity construction and part of the 1986 180222 etc. Prince opposes or productivity organization of the 18022	#4101 (TRES)	290 Kanpa	er Birthcolts	20	10 5	SI Host	181	27	60.00	10	1	raverusery	85 (0 PK	80 33 66	100 00	5000	0.68	155	145	01	155	W.	Ě	
2006 120222 Max men opposen or pro-abtrace envariance and party 2006 120222 Max men opposen or pro-abtrace envariance of 0.70	1892 13785 Get	190 Fronts	eter ESCHUIZO	Garage Control	10 11	93 HVB	199	78.	MJ.(K) BH 00	title funder	7	Variencoine Variencoine	95.00 \$5	1,50 N1.00	1 600,00	753AV	one toract	11/4 1/1001 1: 82	016 N/108	HAR TUNKS	EVE TABLE	Prior spread	1	
				Arri	10 10	0) Shell No most	182	75	e4.00	0	1	Vincentalism	85.60 86	(00 (66 66 (00 (66 66)	400.00	12700	071	Si	0,61	103	107	100	Ď.	
INVALOR TO THE TAX THE SECOND STREET OF THE SECOND	THERETER SERVIN	2 420 Profibe	which is 1459528	Area Area	05 6	00 Heat	121	11.	N.00	50	-	rasiancomi	35.00 20	1,00 Hos White	1.400.00	0200	0.62	1.0	Det	No.	240	100	-	
1808 12222 Inter Investigation Separate Section Interest	73168 12786Wrt	1900 Freeha	etert \$315/21623	Arec.	50 7	00 Hule 00 Hule	100	Tile Stated	like Legal .	100	No. Second	Violenceiro	\$4.00 K	00 54.08	1 500,00 Line Suided	0000 Usa Sector	T DE	271	D7) Distant	NO.	104 Cartes	196 Mar Folder	1	
9892 00271 N.s. montpyele: Dipolatistines/resenterativeF888	SACHO STRANGY	54% Fronts	war PHSSSCHON	Ģ.	the Sator	00 (44)	1.14	no unid	ac 00	n	tee Level	Kertner solespotedur		00 84.00	no trible	in tree	tion forcial	114 3.000	tie tobe	He lank	Nie Sinder	the Seriel	1	
JAME LAULI METAL I INVA departer bi presudit volvale i aliest que re vivido. SARE IXCENTRALE Unior departer pripaduti hidonare i valente la COVE	039600   E2789391 95606   83786361	3400 Freebe	etert COW 600D-0	JAPPE A	title funded?	00 Huit 00 Huid	193 Bide fancet	Frie lands	6H, 00 5C 00	(i) (i) the del	Directured	Kyronesere vectorescenter	95.40 EX	100 80 00	1 BEGICHI Trio fundet	(No finds)	00,79 Dide famost -	1,03 100 \$550	Dise Locks	Dite fancer	(5) (0) (4.66)	tin tinter	1	Size tanget
SSSE 1807 Entire Investigation (inproductivation are contemporarisms) STO	0825 909660	P. 639 Teraporo	ent 1.31 0355-5	jaren .	li s	50 Hvie 00 Hvie	140	17 109 Sedal	81.30	Ö	it tota hoped	Newsenson	85 25 16	Of the firsts	1300.00	SIGN Sign funder	1179 Interferrid	1.50 Ina tode	1.65 too Long	PRO ense famour	ict ice Lone	ine tuncer		see target
TAXABLE TOTAL STREET, COLOR CONTRACTOR OF THE PROPERTY AND ADDRESS OF THE PROPERTY ADDRESS OF THE PROP	9534 B3786941	1990 Fronts	egare MUSCICHOSUM etare MUSCICHOSUM etare MADHITHOSUM	Janes .	52 7	02 3446	151	75	94,00 94,00	01	u.	Nacionalities	85.00 BG	00 (60 M) 00 (61 M)	1 604,00	Dion	839	1195	E140	250	200	160	-	land pulling
DESE 13/224 fettir //www.elpigarten_firtireduct-hidevaren/actionesitire.AGD11	17106 1278D%1				ida tana P	60 years	673	Jive landel	64,00 ·	79	Pire fundet	Kartonerel vaci empurant for		00 8169		Rie tardel	Dieg fundet	Title Sorber	the Lode	tite funds	Arte Sander	His factor	1	
		1 430 Profile	WHAT BUILDS AND THE PERSON	g, and	12 0	00 96/6	100	T9	81,00	63	-	Visitersolitie.	85.00 00	1,00   55.00	1 405,00	D100	5.91	579	0.65	261	202	the target	1	
				Quest .	67 (6	throughout	155	/A  T2	84.00 84.00	NI		Hackenaphine Meriencohine	84.50 80	.00 See funding	1.608.66	8710 65224	90.0.0	E-11	D.42	216 216	276	270	£	
DSP6 IDSCOPPER in house stopperson defends the personal automatical wild for	11478 1239.7341	1990 Foots	ware MARIATATION	Atres.	43 7	(c) 3+vid	18	18	0.00	(1)		(Varytain text) into-	34.00 77	1.60 59.60	1.400.06	1039	60	10	0.0	19	716	IW.	1	
[15/56 ] [LX:224@http://mww.ologoatten.firth/dex.st/hittorians/n/stitutes/states/www.	033672 92787601	1100 Freebe	egent SIMILITATION egent SIMILITATION ellert SIMILITATION	Direct Control	52 K	00 91/18 00 91/18	196	19	B3.00 B3.00	63	1	Vacuumusines Vacuumusines	85.00 (3)	1,00 35.00	1:400.06	2700 2700	1,91	170	0.80	293 .	P3 :	- 102	1	- 0
2002 SCCS Intra Inter eligation followed at hyderare in alternative Wild Contract of the Contr		2290 Fronts	elert MHISBIRGN	Acres	2 6	63 Huid 62 Hold	133	76	0.00	42		Visitamenting Visitamenting	84.00 (1	1.00 SS 00 -	1 200.00 1 000.00	90565	life forget	ilia Lodel	the Lode	ine farder	line Sunder	like fundet	1	
15556 1202225 https://www.elepparters.distributed.com/www.houtemarking.com/states	W286 - 92597790	1 437 Fronts	NOTICE LANGE LANGE	Arre	20 In	a función de	198	74	81,00	es.	8.	Nomina	84.80 50	80 59.00	400.00	11229	0.64	1,00	0.70	216	218	198	¥.	
2006 1202220 m.u. objection provide the construction of construction of the constructi	133WS - 837973WT :	F777 Fronthe	esert VIXIOUS	Mitter Mitter	(4) (b)	02 Huit 02 Huit	(3)	73	84,00	04	8	Vacianosino Vacianasina	85.00 Es	.00 Not funder	1000.00	9900	912	5.90 5.90	0.60	175	104 104	10	P.	-
35368 12522256 et la Tirus Aleganderi Schrischa Effekteranski sektematik in WASH	Tibes 1277 tacket	Q 100 Fronts	eper MANIST	Anni	65 90	93 Half 93 Half 93 Half	155	72	pt 30	Sa Sa	9	Variationalities Vapiationismi	85.00 50	50 little forcing	1500.00	11000	0.26	6.00	6.18	os	(7%)	176	Ť.	
DSSS TOUZES IN A STREET THE STREET ST	PERSON PUTPERSON	1437 Frombe	WHITE BUT THE PROPERTY OF	Direct.	14 6	92 99VM	122	76	60,00	54	7.	(Vagnamoskina	35 00 50	00 PHI NYSE	1.400,00	D400	0.60	1.52	0.14	) Q	214	174		
5000 USC225 first in Invaridappeter, Schröduchhöber and resternation (White Intelligence of Control of Cont	WITERS STREET	1850 Fuells	where Subtrate Colors where Subtrate Calors	ja	the books	0) Hud. 42 Hud.	5.21	Time bridge	0.00	0	his trebt	Makenesins Extrace constructable		1.00 59.00	1 420,00 11to Sci24	Day bardel	MSS Dise forcial	1.79 Itte bi-de	0.95 Ne Links	ote fede	No hade	Ma Spicel	1	
2006 VACCES IN SECURITY STREET STREET, STREET STREET, STREET STREET, S	HERRITE MAN	13% Forbe	ener Mortaletto	A		40 H/M	3.61	tree funded.	91.00	52	nee timest		wilmessforce or		1009 Sunidad	instanti -	114 1270K	ION COME	Dia Linde	tile brook	On Cook	the target	¥.	
DSSS 1202250 filter Timur elopation Symposysty Sevaren verticina in the N/O425	A5650 83786761 ·	5300 Freebe	eter MD42X95158	JA .	tile funded)	27 Style	F-122	Inte fundet	M6.00 M4.00	100	Dis fundet	Kanonerel veden aukreiter	recurrence S co. 13	00 83.90		like tindel	20 to 4 Bit 4e Notabel .	Rise Stodes	Die Lote	sie tade	like turder	elle fundet	P.	
1875 ISC220 into Your elegation to product his branch restor (ISC).  Initial ISC220 into Experient Digmonth converse representation of	15a55 USTBANGO!	1900 Feebs	ejerz MDS0254234W ejerz Szentzkatet	8	ina tyropic	N3 Hold: 92 Hold:	574	inner landed	84.00	U.	lie trest	Karcowye yesteracionali: Packamachina		0.00 80-65 0.00 80-6 N-1360	now to side	Dise Speciel TSC000	lite total	the Lode	tita Lobe	the Leviel	tica fundar	tite farcet	1	
1997 1997 The State of the Stat	176216 92788961 41993 52789661 Victor 92384641	1 872 Fronte	etert fungzareszti atere fungzareszti	(Arre	57 89	00 Hunt	153	79	62.00	E .	K.	Visionopine		00 55.00 00 44.00		10162	0.76	179	0.45	gio	276	100	5	
19595. ICEL STREET IS Treated associated the land of the president and a Party	CHICAGO BESTERNAS		etert SiES435 etert Elforgroots	Ares Ares		SS Hyle 17 Host	143	75	50.50 50.50	e no	N N	Naniemenine Neparation	25.00 Ec	00 84.00	1400.00	0008	0.79	E 51	0.45	951 950	106	160	£	
DISSES 1,59227 BYESE SPINISH RESERVED SKYP VELICITY SERVEN SHAD SERVED SHADOW S	METOE PETERSON	2900 Septemb	SAMS TRYUDOSTOS	April .	57 59	to me	167	14	65.00	tu .		Vaccemonists.	89.00 30	.00 No.00	1-000,00	8990	0.54	HIR SCOTE	0.90	290	319	108	k.	
SSICE 1092227 Inter Your Happerter, Brighted at Shider are lite them as EnacTNO to	NS128 93789561	[7 995   Tuport	Sent TWO DECISE	gare.	107 10	03 91vis 85 91vis	193	76	50.00	0	0	Vspiemopilos (Veniementine	EN 109-340	.00 99 98 .50 99 98	1,300,00	0000	8.00	1.55 His Krider	B.10 B.50	(A)	T/F T48	150	6	
3684 (302295) in the appeter trip of all retains a construct of FUE	SACHE STREET	7.490 from to	ese arvesanes	Green Table	61 6	63 H4/65 00 H4/65	166	H.	e4.00	90		Variationists	25.00.00	00 80 50	1 500,00 1 500,00	15499	0.64	1M	0.64	loe .	lid.	18	1	
DAME LAULDY INTO THE DAWN ENGAGED SERVED AND AND THE PROPERTY OF THE STREET AND A S	4735/6 B37890/1	PLANC Frontier	ASSET TAMOUR STORY	Que.		31Vif.	1 154	78	82.00	80	7	Hackgrookins Makemookins	85.00 80	00 ST68	1.400.00	15400	0.67	5.00	0.53	240	154	127	6	
DADAS DOCUMENTA TOTAL STREET, PRINTED AND STREET, AND	recurs entraced	199 Forte	wint \$500000000	A		00 Style:	5.87	ine untel	No tropic	His todal	Res Suncet	Carcinerer variant accretion	resumbled 7.00 KK	50 H3 69	ion tradel	Rie Lodel	interfactori.	Title School	Data Ender	tile table	NUFT TOTAL	tile fundet	NIS 1.609	
poses 130222048tts: Thrux dippartent provide x5 history and history pre-prison a FoCO	FORB. \$117991(4)	S 600 POCOS	CORRESPONDING CONTRACTOR	Diec.		40 - 34/M	215	181	94,00	52	W.	Int funct		.50 Red funder		32294	7.10	L/N	0.70	208	105	190	*	
NSE 130233455 Investigation Signature Superior Strategy (NSS 130739455 Investigation Strategy (NSS 130739455 Investigation Strategy)	\$116. \$27903e* 578.54 \$3790561	1 500 States	lant EST SEC-S	jaren .	51 P	50 Hvid	155	79	EU 50	(2) (2)	0	Honorookine Newtoniestine	25.00 \$40	0.00 Mo.00	1000.06	5/06 5/06	pres futors pt.78	Dist Loger	Inta Lode 0.50	ink tiree	ANE TUNESS	His Surger	E .	
1500 1500 to once appeter depends of history per value and fill to	6765 E37986(1		ener Kortsbactivi	2500		43 39,45	172	77	F0.50	80	7.	Transacturatery.	25 (0.50	100 53 28 100 54 bu		0000	Site Schild	tie tone	lite Lote:	ine lane	itse Lone	ine tone	1	
2006 1,0,2,2,6 stp: new opgates to pre-activities over-activities 6321 3086 120227 stp: new disposities the educativities of incidental research (144 3088 120227 stp: new edgates for odd thick sension of contract of 8714	M\$165 33790947			D)++-	01 0	ES HUE AS Style	190	10	84.00	53	8	Vanenceine Verenceine	85.09 80	0.00 (55.60	1 604.06	0000	0.54	like kindst	Die Lede	24	the Lodg.	Pix funds	1	
1995 1997 Petro Propriét Dipologia de la Caracteria de la	97145 EC76561	1460 Freche	eges PVISIOSES	garen Garen	N K	40 Hotel 02 Hotel	156	78	84.00 84.00	20		Visite funds		00 55 90 00 55 00		10000	0.00	E 24	0.00	Pin .	000 000	100	1	



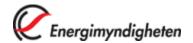
#### **NordCrawl**

**Nordic Council of Ministers** 



The Nordic energy/market surveillance agencies









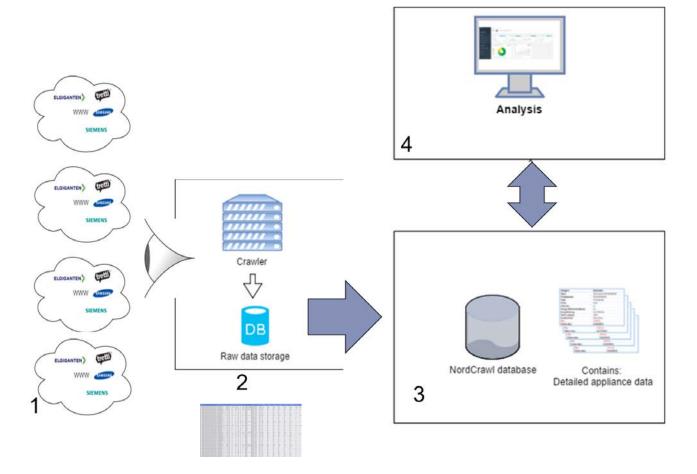






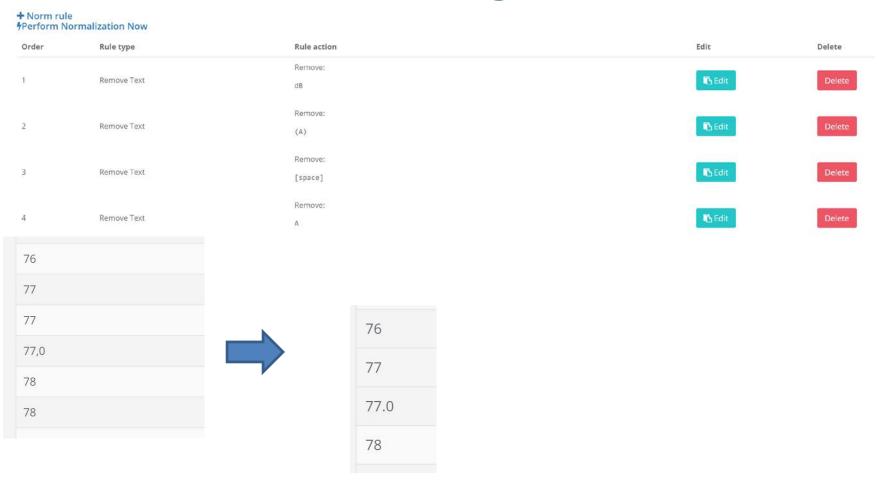
## **System overview**

- Data on products, stored at various publicly available web sites.
- 2. The web crawler engine and a temporary raw data storage.
- 3. A product data base containing processed data.
- 4. The analysis module (a program), which access and display the data in various ways.





## QA – Attribute cleaning & normalization





Examples of

# NORDCRAWL FOR MARKET SURVEILLANCE

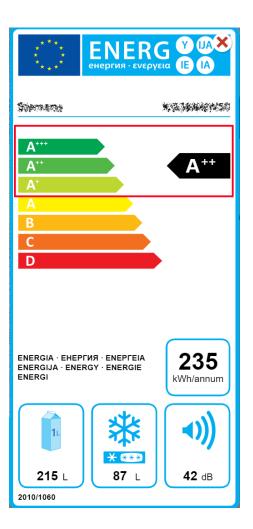


#### **Purpose**

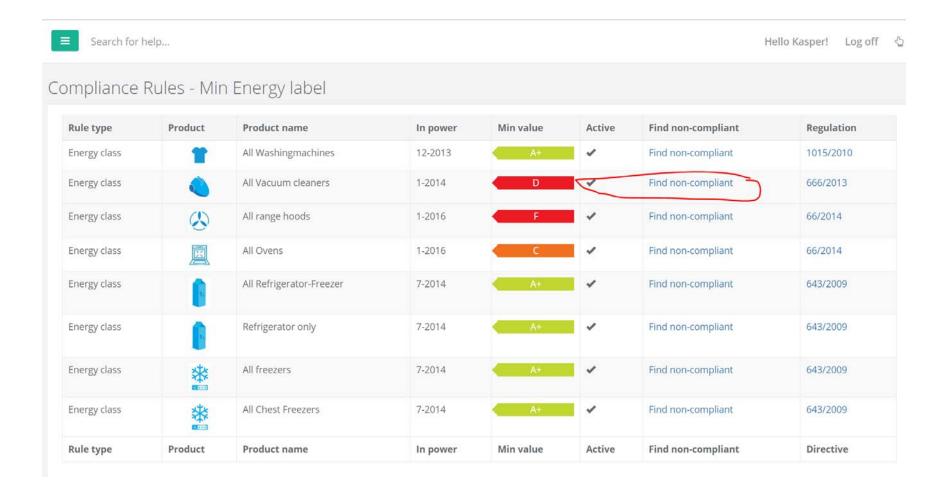
- The Ecodesign directives sets minimum performance standards for energy consumption
- Identify product below the minimum standard

#### Method

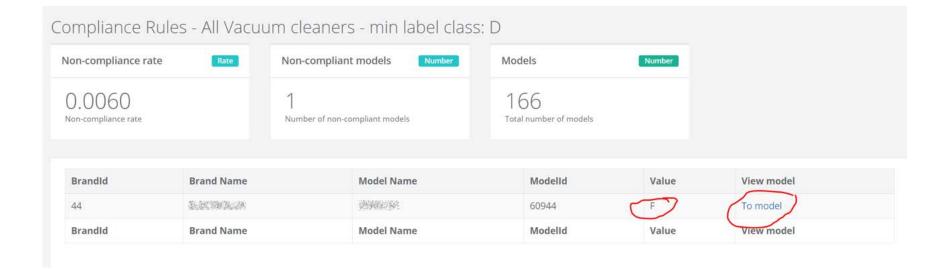
- Minimum Energy Efficiency Index (EEI) or minimum consumption per year in kWh => minimum energy class
- Search database for recently seen model below minimum energy class
- Remove "false positive" products that is not under the regulation
- Some might be on market before regulation













	Minimum	<b>■</b> DK	<b>■</b> SE	₩ NO	+ FI	₩ ICE
Washing machines	A+	0,75% (4)	1,36% (7)	0,44% (2)	0,67% (1)	0% (0)
Vacuum cleaners	D	1,86% (6)	0,6% (1)	0% (0)	5,22% (7)	0% (0)
Range hoods	F	0% (0)	0% (0)	0% (0)	0% (0)	-
Ovens	С	0% (0)	0% (0)	0% (0)	0% (0)	-
Refrigerator-Freezer	A+	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)
Refrigerator	A+	1,19% (9)	0% (0)	0,83% (4)	0,36% (1)	0% (0)
Freezers	A+	0,22% (1)	0% (0)	0% (0)	0% (0)	0% (0)
Chest Freezers	A+	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)

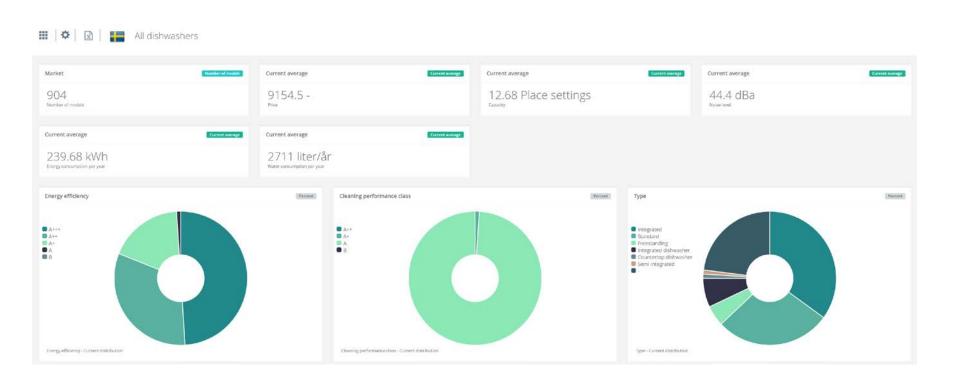


Examples of

# NORDCRAWL FOR POLICY EVALUATION

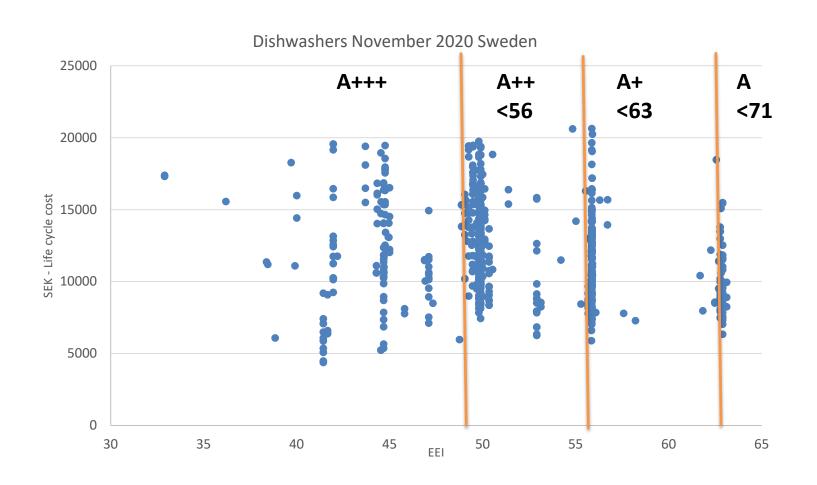


### Market snapshop for policy evaluation





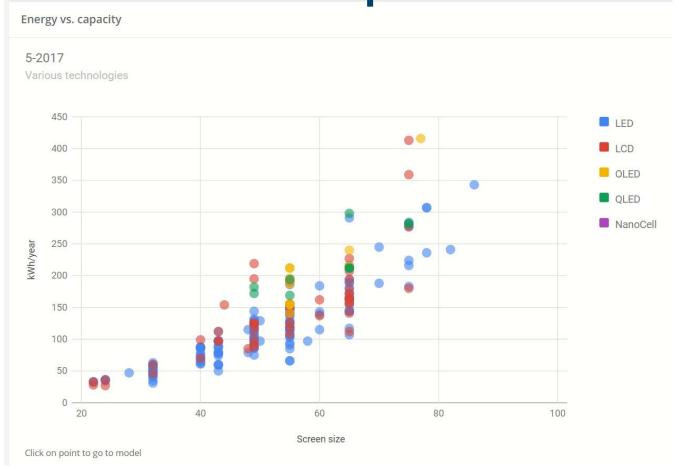
## **Policy evaluation – trends**





Policy evaluation – trends. Stop motion: market development

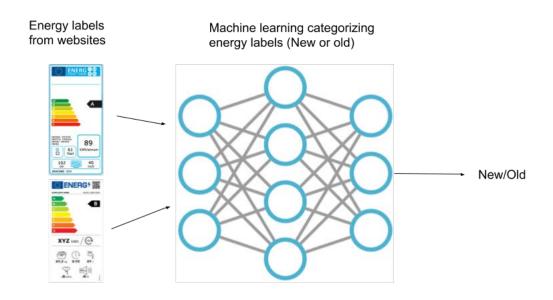






#### **NordCrawl future**

- New energy label tracker
  - Energy label rescale March 2021 (Washing machines and washer-dryers, Dishwashers, Electronic displays, and Refrigerators)





#### Other countries





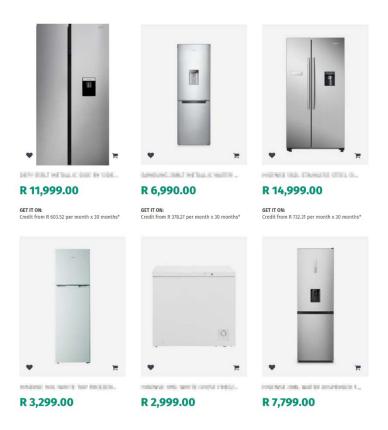
# IEA project air-conditioner and refrigerators in South Africa and Indonesia

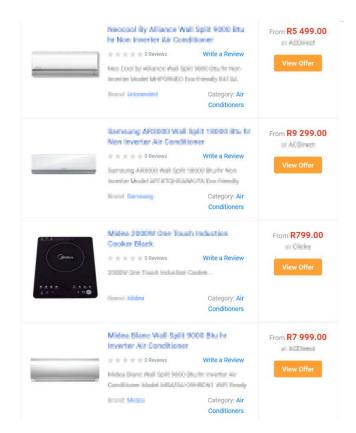
- Fall 2019
- Major online retailer and manufacturers (provided by local partners)
- Get, clean and merge data





## Information quality







# Information quality

					Spesifikasi Sharp AH-A	7SAY AC Split 3/4PK Standard Putih
Product D	etails	Reviews (0)	Questions (0)	Stats	Produk SKU	AH-A7SAY
					Туре	AC Standard 10
Features		Company and Company	oms with this AC and enjo		Putih	
and enjoy fresh and cold. Covering every corner of the  Manufacturer Jet-air room and cooling it evenly is its primary feature. The Jet-				CONTRACTOR	Daya Listrik	595 Watt 0
			le Air Conditioner is engin		Dimensi	88 x 29 x 22 cm
Model	design and can be operated from a user-friendly remote control. Features: - Remote control - Elegant Modern Panel				Berat	9 Kg
Number 254	254		th LED display - Free Mov	Same and the second	Garansi	1 Tahun Resmi (1)
			Air flow: 400 m3/h - Appli		PK	3/4 PK 🐧
		m2 - Noise level: 64 dB(A) - Warranty: 3 years. [T Apply] Dimension(mm): - Net: 480 x 400 x 795 - Gross: 520 x 450 x			BTU/h	7000 BTU/h 🐧
		885 More Specs: - Rating Cooling/Heating(KW/h):	Refrigrant	R-32 🙃		
		12000/10000 - Rating Cooling/Heating Current (A): 5,95/4,88 What's in the box 1 x Jet-Air Portable Air			Uk. Pipa	∞1/4 ∞3/8 🕦
			er 1 x Remote Control 1 x		Dimensi Ou.	60 x 50 x 27 cm
		Window Kit	ndow Kit Berat Ou. 19 Kg		19 Kg	
					Made In	thailand
					Customer Reviews	*** 2 review



#### Ex. level of information

#### **South Africa – Air-con**

Cooling capacity btu/h	90%
Refrigerant	80%
Room size m2	37%
Energy class	17%
Noise Level (dB) outdoor unit	9%
EER	5%

#### **Indonesia – Refrigerators**

Price	89%
Capacity total liter(net)	78%
Type of product	17%
Electricity per year (kWh)	4%
Noise level dB (A)	2%



#### **Observations**

- Information on energy and energy efficiency is minimal
- Hard for consumers to make well-informed decisions
- "EU like" requirements for consumer information would help



Thank you!

Kasper Mogensen, CTO Big2Great

ksm@big2great.dk

Feel free to contact me