



Centraal Bureau voor de Statistiek

Programma Impact ICT

Onderzoeksrapport nr. 7

Medegefinancierd door het Ministerie van Economische Zaken en Prima

**EINDRAPPORT INTERNET ROBOTS WONINGMARKT
OVER 2011 EN 2012**

Olav ten Bosch en Dick Windmeijer

Kennisgeving:

De in dit rapport weergegeven opvattingen zijn die van de auteurs en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

Projectnummer:

210394

Datum:

11 maart 2013

Samenvatting

Het internet robots project van het CBS heeft in 2011 en 2012 voor analysedoeleinden dagelijks data waargenomen over één provincie van 5 verschillende woningmarkt sites. Wijzigingen aan de sites zijn binnen de looptijd van het project in de robots verdisconteerd. Het CBS-speerpunt woningmarkt heeft in 2011 de toen beschikbare data geanalyseerd met als conclusie dat de mogelijkheden voor statistiek veelbelovend waren, maar verder onderzoek noodzakelijk werd geacht. De robots zijn in 2012 voortgezet en iets uitgebreid. Het internet robots project heeft het speerpunt woningmarkt geassisteerd bij gesprekken met houders van woningmarkt sites over een eventuele opschaling naar heel Nederland. Dit project heeft aangetoond dat het goed mogelijk is meerdere internetbronnen die hetzelfde domein beschrijven periodiek over langere tijd uit te lezen en de bronnen in te dikken tot een voor analyse bruikbare dataverzameling en heeft aldus als katalysator voor vernieuwing gefungeerd. Wij denken dat internet en internet technieken nog breder kunnen worden ingezet voor de statistiek. De ervaringen van dit project en de ontwikkelde generieke waarneem software (“het framework”) kunnen daarbij als basis dienen

Trefwoorden: internet robots, dataverzameling, woningmarkt

Inhoud

1. Inleiding.....	4
2. Het verloop van het project.....	5
2.1 De voorbereiding	5
2.2 De uitvoering	6
2.3 De verwerking van data voor analyse.....	7
3. De bevindingen van het speerpunt woningmarkt.....	10
4. De rol van internet robots voor de statistiek.....	11
5. Juridische aspecten	13
6. Conclusie	14
7. Referenties	15

1. Inleiding

Als onderdeel van het innovatie programma bij het CBS worden nieuwe vormen van informatietechnologie verkend voor statistische doeleinden. Een van de onderdelen is het gebruik van internet robots voor het verzamelen van internet data voor de statistiek. In 2009 en 2010 is in dit project op bescheiden schaal geëxperimenteerd met internet robots voor het waarnemen van brandstofprijzen en prijzen van vliegreizen. De ervaringen zijn beschreven in [1, 2]. Daarnaast worden bij het CBS projecten uitgevoerd op het gebied van het gebruik van internet data voor statistieken over vacatures en toerisme.

Het CBS organiseert vernieuwingsprojecten in zogenaamde speerpunt projecten. Voor de vernieuwing van de woningmarkt statistieken is het speerpunt woningmarkt in het leven geroepen. Voor het stimuleren van Internet als nieuwe databron (IaD) voor het maken van statistiek is het speerpunt Impact ICT in het leven geroepen. In overleg met het speerpunt woningmarkt heeft het speerpunt Impact ICT besloten in 2011 en 2012 te onderzoeken of internet robots zinvol kunnen worden ingezet ten behoeve van statistieken op het gebied van de woningmarkt. Een van de wensen van het speerpunt woningmarkt was om nieuwe, snelle indicatoren over de woningmarkt te creëren, zoals een woningmarkt spanningsindicator. Het idee was om de data over vraagprijzen van woningen te confronteren met andere data zoals de werkelijke transactiepreizen van woningen om iets te kunnen zeggen over de bewegingen op de woningmarkt [3, 4]. Het speerpunt woningmarkt heeft dit met behulp van de door de robots vergaarde data in 2011 inderdaad kunnen onderzoeken, resulterend de conclusie dat de mogelijkheden voor statistiek veelbelovend waren, maar verder onderzoek noodzakelijk werd geacht. Daarnaast bestond er belangstelling om internet data in te zetten voor het in beeld brengen van ontwikkelingen op de huurmarkt. Ook daar is in dit project een test mee gedaan, echter deze data is nooit geanalyseerd. Wij laten dit verder buiten beschouwing in dit rapport.

Het eindrapport dat nu voor u ligt geeft een overzicht van het internet robots project voor de woningmarkt over de periode 2011 en 2012. Hoofdstuk 2 van dit rapport beschrijft het verloop van het project in algemene zin. In hoofdstuk 3 geven we de conclusies van het speerpunt woningmarkt na analyse van de data weer. Hoofdstuk 4 gaat in op de rol die internet robots kunnen hebben voor de statistiek. Hoofdstuk 5 beschrijft het juridisch kader. Hoofdstuk 6 ten slotte geeft de conclusies.

2. Het verloop van het project

2.1 De voorbereiding

Allereerst is een plan van aanpak geschreven. Daarin is de kennis en ervaring uit de eerdere robot projecten [1, 2] verwerkt. Veel was nog onduidelijk, bijvoorbeeld hoe overlappend het aanbod van de huizensites was, hoe eenvoudig de sites waren uit te lezen, hoe dynamisch de sites waren, welke variabelen er te halen waren, welk volume moest worden verwacht en hoe beweeglijk de data was. Veel van deze vragen waren alleen te beantwoorden door de situatie beter te analyseren, zowel handmatig als automatisch.

Daarom is eerst een voorstudie uitgevoerd naar de kenmerken van zo'n 32 woningmarkt sites. Deze kenmerken zoals aantal woningen, de gebruikte bronnen, de bestaansduur en de technische opzet¹ zijn zoveel mogelijk ingevuld, maar dat kon lang niet in alle gevallen worden gedaan. Vervolgens is onderscheid gemaakt tussen sites die 'eigen content' hebben en sites die hun content voornamelijk van andere sites halen. Deze indeling was belangrijk om te bepalen welke bronnen voor statistische doeleinden het meest geschikt waren. Op basis van deze analyse zijn 4 woningmarkt sites gekozen om mee te experimenteren.

Het experiment is vervolgens zo ontworpen dat het in beperkte tijd en met beperkte middelen binnen de geldende juridische kaders kon worden uitgevoerd, terwijl er wel voldoende resultaat uit zou komen om uitspraken te kunnen doen over de mogelijkheden. De volgende keuzes zijn gemaakt:

- *Frequentie*: De keuze moest worden gemaakt tussen dagelijkse of wekelijkse waarneming. Een wekelijkse waarneming vermindert de belasting op de sites en maakt de data processing eenvoudiger, maar daar staat tegenover dat je op die manier wel eens statistisch relevante bewegingen zou kunnen missen. Daar kunnen we echter pas iets over zeggen als we eerst hoogfrequent, dus dagelijks, gemeten hebben. Er is in dit experiment daarom gekozen voor een dagelijkse waarneming², mede omdat we de belasting op de sites op andere manieren laag konden houden (zie hieronder bij volume en variabelen).
- *Volume*: Uit praktisch oogpunt hebben we het experiment op dit punt sterk afgebakend. Eerste analyse had aangetoond dat het goed te doen was om de

¹ Inclusief de eventuele aanwezigheid van een instelling conform het robot exclusion protocol, waarmee een site de mate van gewenstheid van een internet crawler kan aangeven.

² Overigens is er gedurende de testperiode ook een kortlopend hoogfrequent onderzoekje uitgevoerd om de dynamiek van een site over de dag te meten en de dynamiekverschillen tussen werk en weekend dagen. Dat heeft inzicht geleverd die nuttig was voor de operationele kant van de zaak. Wij laten dat hier verder buiten beschouwing.

robots een geografische beperkt gebied te laten waarnemen, terwijl verwacht mocht worden dat het experiment statistische gezien nog voldoende resultaat zou opleveren. Dat moet dan natuurlijk wel voor alle robots dezelfde afbakening zijn om de data met elkaar te kunnen confronteren en de sites moesten geschikt zijn om die afbakening te kunnen uitvoeren. Aldus hebben we het experiment beperkt tot één provincie. Op basis van eerste analyse hebben we gekozen voor de provincie Utrecht. Dat leek een goede keuze wat betreft representativiteit, volume, doorlooptijd en processing.

- *Variabelen*: Ook hier hebben we het experiment afgebakend. Een site dieper onderzoeken levert meer informatie op, maar dat is vanwege belasting, performance etc. weer niet handig. Het experiment is daarom zo ontworpen dat alleen de variabelen op de overzichtspagina's van de sites werden waargenomen. Technisch gesproken is dat een optimum aan informatie tegen een minimum aan requests. Belangrijk was dat op die manier op iedere site de sleutel van woning objecten kon worden waargenomen (postcode, huisnummer, huisnummertoevoeging) voor koppeling aan andere bij het CBS bekende gegevens. Daarnaast konden op deze manier variabelen als URL, straat, woonplaats, aantal kamers, makelaar, woonoppervlakte, prijs, kk./von. waargenomen.
- *Periode*: Het experiment zou initieel een beperkte looptijd hebben van mei 2011 t/m dec 2011. Later is dat met een jaar verlengd (zie later).

2.2 De uitvoering

Het internet robots project heeft in het eerdere traject vóór 2011 meerdere vormen van dataverzameling van internet bronnen uitgetest. Het project heeft zelf internet robots gemaakt, maar heeft daarnaast ook goede ervaringen gehad met de inschakeling van in robot technologie gespecialiseerde externe partijen. In dit onderzoekstraject was er vooraf erg weinig bekend over de specificaties en was er veel behoefte aan flexibiliteit. Daarom is besloten om voortbordurend op de eerder opgedane ervaringen zelf internet robots te maken. De robots zijn gemaakt in de programmeertaal R³, waarbij zoveel mogelijk een standaardopzet is gekozen met een gemeenschappelijke software onderlaag (framework) dat generieke waarneem functionaliteit implementeert. Dit framework is domein neutraal en dus ook geschikt om sites met heel andere objecten dan de woningen uit te lezen⁴. De robots zijn allemaal gestart vóór mei 2011. Sommige sites bieden aanvullende functionaliteit aan zoals hoe lang een woning al te koop staat of welke woningen in een bepaalde periode verkocht zijn. Na overleg met het speerpunt woningmarkt is vanaf 12 mei

³ Eén robot is initieel in een andere tool gemaakt, maar later overgezet naar R. Op dit moment zijn alle robots gestandaardiseerd op R.

⁴ Dit framework wordt inmiddels ook gebruikt in een project op het gebied van prijswaarneming voor de statistiek.

2011 tevens een robot geïnstalleerd die van één van de sites maandelijks deze aanvullende informatie raadpleegt, namelijk de woningen die verkocht zijn in de provincie Utrecht. Eén site bleek per 1 juni 2011 geen wijzigingen meer te kennen, deze bleek uitgefaseerd.

Bij het ontwerp van de robots is gestreefd naar maximale robuustheid. Dat betekent dat de robots zo veel mogelijk bestand zijn tegen kleinere wijzigingen in de sites. Eén van de manieren om dat te doen is om zo weinig mogelijk verwerking van de informatie in de robot zelf te doen. Alleen het waarnemen en wegschrijven van informatiehoudende onderdelen gebeurt in de robot zelf. De interpretatie daarvan gebeurt in een postprocessing stap. Het voordeel van deze aanpak is dat eventuele tijdskritische bijstellingen aan een robot snel kunnen worden uitgevoerd waarbij de postprocessing later met terugwerkende kracht kan worden uitgevoerd.

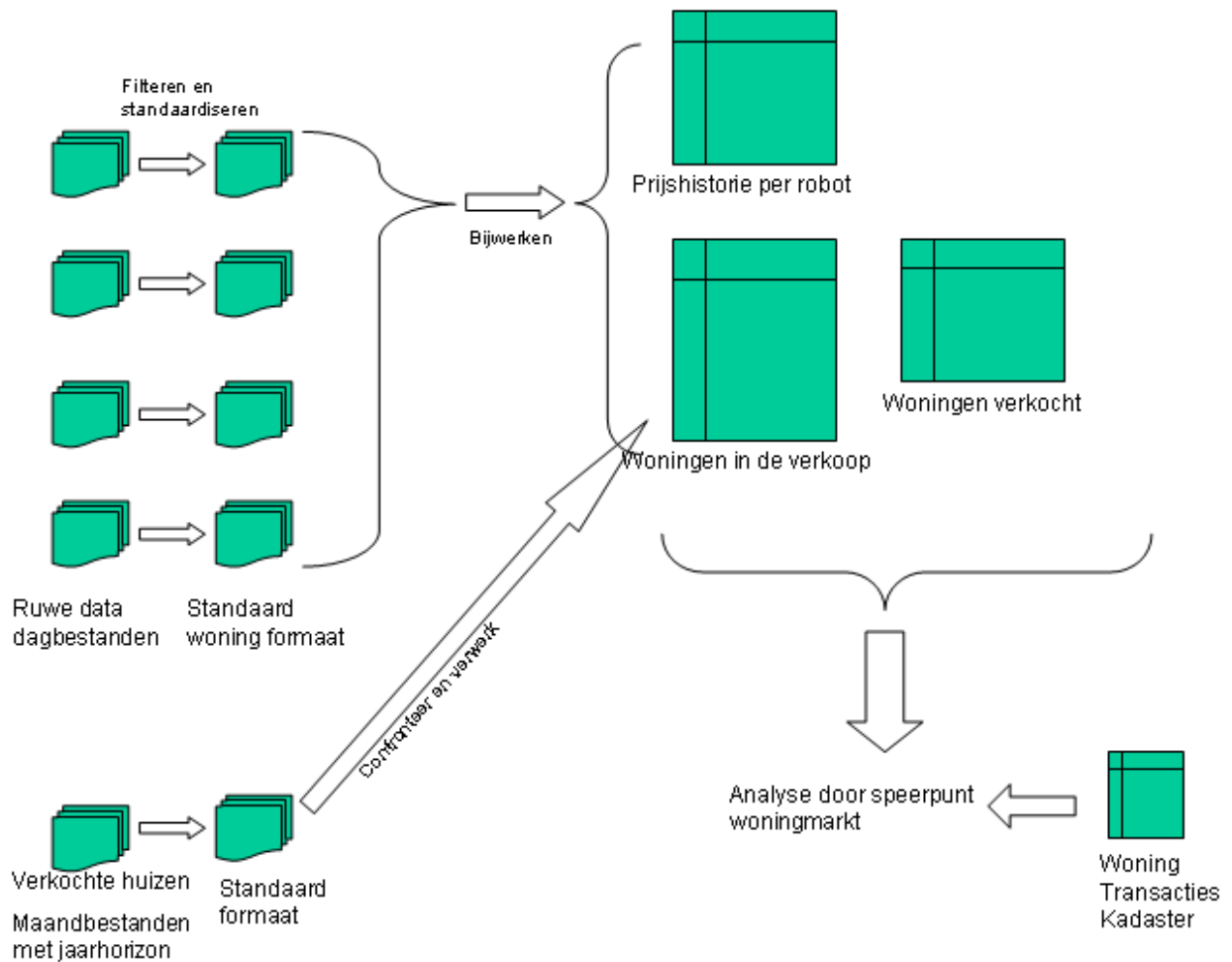
De robots draaiden dagelijks geheel automatisch vroeg in de ochtend. In het overgrote deel van de gevallen ging dit goed. Er zijn dagen dat er door onduidelijk oorzaak geen waarneming was of een halve waarneming. De output van de robots werd gedurende het project elke dag in de gaten gehouden via een automatisch mailbericht met informatie over de verzamelde data. Tijdens de looptijd van het project zijn enkele malen wijzigingen geweest aan de sites, waardoor de robots moesten worden bijgesteld. Meestal werd dit dezelfde dag nog ontdekt⁵ en gerepareerd. In sommige gevallen werd dit wel eerder ontdekt maar kon dit door afwezigheid pas later worden gerepareerd. De grootste onderbreking in de waarneming had een lengte van 13 dagen. De daadwerkelijke bijstelling aan de software nam slechts 2 uur in beslag. Voor alle robots geldt dat bijstellingen aan de robot zelf nooit meer dan een halve dag in beslag namen. De eventuele bijstellingen aan de postprocessing werden dan later uitgevoerd als de eventuele consequenties van de wijziging op de data helder waren.

De robots halen alleen die informatie op die voor het statistische experiment nodig is. Er worden dus geen foto's of filmpjes opgehaald. In totaal zijn er per dag in dit experiment meer dan 30.000 records binnengekomen. Een schatting van het totaal aantal verzamelde records over de twee jaar looptijd komt uit op ca 21 miljoen waarnemingen (records). Daarbij tekenen we aan dat hierbij vele dubbele waarnemingen zitten, afkomstig van verschillende sites over hetzelfde woningobject, of afkomstig van verschillende of dezelfde site maar gemeten op een ander tijdstip. Het is duidelijk dat om hier zinvolle statistische uitspraken over te kunnen doen de stroom gegevens moet worden ingedikt tot een voor de statistiek handzame vorm.

2.3 De verwerking van data voor analyse

In 2011 is een analyse proces ontwikkeld waarmee de robot data werd gefilterd, gestandaardiseerd, omgezet in een standaardformaat en werd ingelezen in een database. Dit proces is geschetst in figuur 1.

⁵ Een check van de omvang van het databestand zegt meestal al genoeg.



Figuur 1: Het analyse proces om robots data te verwerken

De robots produceren bestanden per dag (ruwe data in dagbestanden). Deze worden door robot-specifieke scripts (in de taal R) omgezet in gestandaardiseerde dagbestanden (standaard woning formaat) waar invalide waarnemingen verwijderd worden en postcodes aan vaste standaarden voldoen. Deze dagbestanden worden opgeslagen in een database, zodanig dat de woningen in de verkoop die bij verschillende robots zijn waargenomen aan elkaar worden gekoppeld en slechts eenmaal worden opgeslagen. Een eventuele onderbreking van de waarneming bij de ene site wordt op die manier deels opgevangen door de waarneming bij de andere site. Robot-specifieke variabelen (bijvoorbeeld woningindeling volgens één site) blijven opvraagbaar.

Identificatie van woningen gebeurt op basis van postcode, huisnummer en huisnummertoevoeging. Bij het koppelen van waarnemingen van verschillende sites die betrekking hebben op dezelfde woning moeten keuzen worden gemaakt. Als de kenmerken van een woning op verschillende sites niet hetzelfde is, moet een keuze worden gemaakt welke wordt overgenomen in het woningen in de verkoop bestand.

Per variabele is gekeken welke bron de voorkeur krijgt in geval van verschillen. De prijshistorie van een woning in de verkoop wordt per robot zodanig opgeslagen dat alleen prijsmutaties worden bijgehouden. De woningen die verkocht zijn maar waarvoor geen prijshistorie beschikbaar is worden in een separate tabel opgeslagen.

Aldus wordt de omvangrijke datastroom ingedikt tot een database met kenmerken van woningen in de verkoop en prijshistorie. Er zijn vier leveringen geweest van deze output aan het speerpunt woningmarkt met alle data die tot dan toe vergaard was. Het speerpunt woningmarkt heeft hierop analyse gedaan (zie volgende hoofdstuk), waarbij tevens bleek dat woningen toch nog dubbel voorkwamen. Dit werd veroorzaakt door minieme informatieverschillen in de waarnemingen van naar later bleek identieke objecten. Om deze op te lossen zou het koppelproces verder moeten worden verbeterd, iets wat wij in dit onderzoeksproject niet verder hebben geïmplementeerd.

Een van de conclusies van het onderzoek van het speerpunt woningmarkt was dat er behoefte was aan een langere reeks van onderzoeksdata (zie volgende hoofdstuk). De pilot is daarop verlengd naar geheel 2012. Tevens is op 6 september 2012 een nieuwe woningmarkt site toegevoegd met dezelfde opzet als de eerdere robots.

3. De bevindingen van het speerpunt woningmarkt

Het speerpunt woningmarkt heeft in 2011 statistische analyses uitgevoerd op de data om de haalbaarheid van het afleiden van statistische indicatoren te bekijken. De uitkomsten zijn beschreven in een eindrapport [5]. We citeren uit de conclusie:

De eerste resultaten van het onderzoek zijn veelbelovend, maar verder onderzoek wordt nog wel noodzakelijk geacht op de volgende gebieden:

- *Dubbele waarnemingen*
Seizoenspatronen
Dekkingsgraad sites in regio's
Koppeling met het Kadasterbestand

Daarnaast zou er nog onderzoek uitgevoerd kunnen worden over de volgende onderwerpen:

- *Aantal kamers (95% van de gevallen gevuld)*
- *Nieuwbouw*
- *Koppeling met de BAG voor de kenmerken woningtype en bouwjaar.*

Het werd duidelijk dat het speerpunt woningmarkt een langere reeks waarnemingen nodig heeft om de haalbaarheid van hun doelen nog beter te kunnen onderzoeken en daarnaast graag data zou willen hebben over heel Nederland. Om tegemoet te komen aan het eerste is de looptijd van het experiment verlengd. Voor het tweede (opschaling naar heel Nederland) is gekozen om in overleg te treden met de houders van de woningmarkt sites. Het internet robots project heeft geassisteerd bij deze gesprekken. Het voorlopige resultaat is dat vanaf 2013 één van de sites met toestemming wekelijkse in zijn geheel met een robot wordt waargenomen en dat een andere site wekelijks de data voor heel Nederland levert via een alternatieve koppeling. Voor de overigen zijn gesprekken nog gaande. Dit alles nog steeds als uitbreiding van de onderzoeksfase. Na afloop daarvan zal het CBS beslissen over eventuele opschaling naar productie.

4. De rol van internet robots voor de statistiek

Dit project heeft aangetoond dat het goed mogelijk is meerdere internetbronnen die hetzelfde domein beschrijven periodiek over langere tijd uit te lezen en de bronnen in te dikken tot een voor analyse bruikbare dataverzameling. Of dit voor de woningmarkt uiteindelijk ook de beste manier zal blijken te zijn om tot statistische productie te komen zal de komende tijd blijken. Zelfs als dat niet zo is hebben internet robots gefungeerd als katalysator voor het verkennen van nieuwe bronnen voor de statistiek. Dat zet ons aan het denken over de rol die internet en internet robots in bredere zin kunnen hebben voor de statistiek.

Uitzoomend komen we vanuit dit project op de volgende mogelijkheden:

- als middel om zicht te krijgen op mogelijke data bronnen voor een bepaald statistisch domein (“welke partijen zijn eigenlijk actief op de woningmarkt?”)
- als middel om de geschiktheid van bronnen voor statistiek te verkennen (“kunnen we eigenlijk statistiek maken uit de vraagprijzen van woningen?”)
- als middel om meer inzicht in kenmerken van een statistisch domein te verkennen (“hoe groot is de woningmarkt eigenlijk, tussen welke grenzen bewegen de vraagprijzen zich, hoe vaak wijzigen vraagprijzen gemiddeld en hoe dynamisch is het woningaanbod”)
- als instrument om meer inzicht te krijgen in de kwaliteit / volledigheid of representativiteit⁶ van anders verkregen data (“hoeveel mensen zetten er eigenlijk zelf hun huis te koop buiten alle officiële makelaars om?”)
- als middel om data te *verrijken* met aanvullende kenmerken (“we kennen de transactieprizen van woningen, maar hoe lang hebben deze objecten eigenlijk te koop gestaan”)
- als middel om relevante indicatoren op te halen die niet makkelijk uit andere bronnen te halen zijn (“welk soort woningen staan te koop in welke (soort) regio”, hoe lang staan woningen te koop (per regio)”)
- als katalysator voor het bedenken van nieuwe mogelijkheden (“kunnen we internet data ook gebruiken voor de berekening van huurprijzenindex voor expats?”)
- als middel om snel data voor onderzoeksdoeleinden beschikbaar te krijgen
- als waarneeminstrument voor productie van statistiek

⁶ Wij realiseren ons dat door internet robots verkregen data ook niet representatief hoeft te zijn. Wat we bij dit punt bedoelen is een indicentele aanvullende waarneming, bijvoorbeeld om te toetsen hoeveel woningen op een kleinere site waar mensen hun eigen huis te koop kunnen zetten ook in de reguliere makelaarssites zitten.

Deze lijst is vast niet uitputtend. We staan naar ons idee aan het begin van periode waarin internet bronnen op vele manieren bij de statistiek of statistisch onderzoek toepasbaar kunnen zijn. Dat wordt bevestigd door het groeiend aantal vragen dat ons bereikt om internet technieken (wij spreken hier wat algemener dan robots) in te zetten.⁷ De ervaringen van dit project kunnen dan als basis dienen.

⁷ Zo gaan wij binnenkort experimenteren met technieken voor het ondersteunen van bedrijfsstatistieken.

5. Juridische aspecten

Een belangrijk aandachtspunt bij het gebruik van internet robots is de juridische context waarin geopereerd wordt. In dit woningmarkt project is ervoor gekozen het experiment kleinschalig te houden in omvang. Tevens is voor maximale transparantie een melding in het meldingenregister gemaakt. Voor een eventuele opschaling van het experiment naar heel Nederland is door het speerpunt woningmarkt besloten in overleg te treden met de site eigenaren. Vanwege het aantal bronhouders is dat in dit geval heel goed mogelijk. In een gebied waar de informatie verspreid is over een groot aantal sites is dat echter niet meer praktisch haalbaar. In dat geval zou het goed zijn een universeel toepasbaar en transparant kader te hebben van waaruit geopereerd kan worden.

Ook hieraan is in dit project aandacht besteed. Vanuit het internet robots project is een voorstel gemaakt voor een strategie waarbij het CBS streeft naar een maximale transparantie van opereren en een minimale last voor internet bronhouders. Onderdeel van deze strategie is dat internet robots van het CBS zich duidelijk identificeren met een verwijzing naar publiekelijk beschikbare aanvullende informatie. Deze strategie kan als startpunt dienen voor toekomstige projecten op dit gebied.

6. Conclusie

Het internet robots project heeft in 2011 diverse robots gerealiseerd op woningmarkt sites om de bruikbaarheid van deze data voor de statistiek te kunnen onderzoeken. De robots namen één enkele provincie waar. Binnen de looptijd van het project zijn diverse wijzigingen aan de sites opgetreden. Deze konden steeds binnen een halve dag worden verwerkt in de robots. Ongeveer de helft van de tijd is gaan zitten in het verwerkingstraject van de data voor analyse voor het speerpunt woningmarkt.

Het speerpunt woningmarkt heeft geconcludeerd dat de eerste resultaten van het onderzoek veelbelovend waren, maar ook dat een langere reeks waarnemingen nodig is. Daarnaast was er behoefte aan data over heel Nederland. Ten slotte is gebleken dat analyse en verwerking een aanzienlijk deel van de resources kost. Om tegemoet te komen aan het eerste is de looptijd van het experiment verlengd. Bovendien is gedurende 2012 een nieuwe site toegevoegd. Voor het tweede is gekozen om in overleg te treden met de houders van de woningmarkt sites. Het internet robots project heeft geassisteerd bij deze gesprekken. Het voorlopige resultaat is dat vanaf 2013 één van de sites met toestemming wekelijks in zijn geheel met een robot wordt waargenomen en dat een andere site wekelijks de data voor heel Nederland levert via een alternatieve koppeling. Voor de overigen zijn gesprekken nog gaande. Dit alles nog steeds als uitbreiding van de onderzoeksfase. Na afloop daarvan zal het CBS beslissen over eventuele opschaling naar productie.

Dit project heeft aangetoond dat het goed mogelijk is meerdere internetbronnen die hetzelfde domein beschrijven periodiek over langere tijd uit te lezen en de bronnen in te dikken tot een voor analyse bruikbare dataverzameling. Op deze manier hebben internet robots gefungeerd als katalysator voor het verkennen van nieuwe bronnen voor de statistiek. Wij denken dat internet en internet technieken breder kunnen worden ingezet voor de statistiek, bijvoorbeeld als middel om zicht te krijgen op mogelijke data bronnen voor een bepaald statistisch domein, als middel om de geschiktheid van bronnen voor statistiek te verkennen, als middel om meer inzicht te krijgen in de kenmerken van een statistisch domein, als instrument om meer inzicht te krijgen in de kwaliteit / volledigheid of representativiteit van anders verkregen data, als middel om data te verrijken met aanvullende kenmerken, als katalysator voor het bedenken van nieuwe mogelijkheden of als waarneeminstrument voor de productie van statistiek.

We staan naar ons idee aan het begin van periode waarin internet, internet technieken en internet bronnen op vele manieren bij de statistiek of statistisch onderzoek toepasbaar kunnen worden gemaakt. De ervaringen van dit project en de ontwikkelde generieke waarneem software (“het framework”) kunnen daarbij als basis dienen.

7. Referenties

- [1] *Automated Data Collection from Web Sources for Official Statistics: First Experiences*, Rutger. Hoekstra, Olav ten Bosch and Frank Hartevelde, June 30 2010, (to be published in the Journal of the International Association for official Statistics (IAOS))
- [2] *New data sources for statistics: Experiences at Statistics Netherlands*, P. Daas, M. Ros, C. de Blois, R. Hoekstra, O. ten Bosch en Yinyi Ma, NTTS conference, 22-24 Feb 2011, Brussels
- [3] *Internetrobots woningmarkt – doelen, variabelen en sites*, Heymerik van der Grient, Gerrit Zijlmans, 18 oktober 2010
- [4] *Wenselijkheid en mogelijkheid van spanningsindicator woningmarkt (versie 1)*, Heymerik van der Grient, 1 september 2010
- [5] *Eindrapport Woningen in verkoop deelproject spanningsindicator*, Karine Tanis en Christian Dijkstra, 14 november 2011