

# Zoekvoorbeelden Locatieserver

In dit document wordt een aantal voorbeelden gegeven van zoekvoorbeelden van Locatieserver.

URL's: \* PDOK-viewer: <http://pdokviewer.pdok.nl/> \* Locatieserver suggest endpoint: <https://geodata.nationaalgeoregister.nl/locatieserver/v3/suggest> (geeft de eerste 10 resultaten terug van alle records) \* Locatieserver lookup endpoint: <https://geodata.nationaalgeoregister.nl/locatieserver/v3/lookup> (geeft een foutmelding terug, omdat deze een ID-parameter verwacht) \* Locatieserver free endpoint (vrije geocodeerservice): <https://geodata.nationaalgeoregister.nl/locatieserver/v3/free> (geeft de eerste 10 resultaten terug van alle records)

Zie hieronder voor meer uitleg over hoe de endpoints aangeroepen moeten worden.

## Achtergrond

Locatieserver maakt gebruik van [Apache Solr](#). Dit is een zoekmachine bedoeld om in grote hoeveelheden verschillende soorten data te kunnen zoeken. Het is in principe bedoeld voor allerlei soorten ongestructureerde informatie (NoSQL-database), maar het kan ook worden gebruikt om meer gestructureerde informatie, zoals geodata, te ontsluiten. Het kent zelfs functionaliteit om ruimtelijke vragen te stellen, bijv. het sorteren van zoekresultaten op afstand van een meegegeven coördinaat. In vergelijking met een geo-database is dit beperkt, maar voor een eerste analyse op de informatie is het voldoende. Dit past binnen de scope van een geocodeerservice.

Voordat er met Solr gezocht kan worden, moet er eerst geïndexeerd worden. Dit is het proces om data te leveren aan de search engine. Bij het indexeren worden verschillende databronnen ingelezen, waarna Solr een zoekindex aanmaakt. Bij het indexeren kan tevens informatie aan elkaar worden gekoppeld en verrijkt, bijv. om aan objecten een gemeente en woonplaats toe te kennen.

De Locatieserver-service heeft de volgende drie endpoints: \* Suggest: hiermee kunnen suggesties voor geo-objecten worden opgevraagd o.b.v. een opgegeven zoekterm; \* Lookup: hiermee kan detailinformatie over een gevonden geo-object worden opgevraagd; \* Free: hiermee kan "vrij" worden gezocht. Dit is de "vrije geocodeerservice".

Bij het suggest-endpoint wordt gebruik gemaakt van de weergavenaam van een object. Dit is een tekst aan de hand waarvan het object kan worden herkend, bijv. adres+postcode+woonplaats. Bij het free-endpoint kan op alle velden van het object worden gezocht. Deze informatie wordt bij het indexeren in het zoekveld "tekst" geplaatst.

De PDOK-viewer maakt alleen gebruik van het suggest- en lookup-endpoint. Deze werken als een tweetrapsraket. Met het suggest-endpoint wordt weliswaar een beperkt aantal velden teruggegeven, maar dit biedt voldoende informatie om de gebruiker een keuze te laten maken. Hierna kan alle beschikbare informatie via het lookup-endpoint worden opgevraagd. De oude geocodeerservice maakt gebruik van één zoekvraag. In de response wordt alle beschikbare informatie over de geo-objecten teruggegeven. Bij de oude service is deze hoeveelheid beperkt,

maar bij de locatieserver kan veel meer informatie worden ontsloten.

De suggest-service is speciaal ontworpen om grote hoeveelheden zoekvragen aan te kunnen. Vandaar dat er maar een paar velden worden teruggegeven. Verder wordt de data waarop met de suggest-service gezocht kan worden op een speciale manier geïndexeerd, zodat deze geschikt wordt voor autocomplete-functionaliteit. Deze werkwijze heet "edge ngram". Iedere zoekterm wordt meerdere keren in de database opgeslagen, waarbij steeds één letter meer wordt toegevoegd. De naam "Hofstraat" wordt op de volgende manieren opgeslagen: "H", "Ho", "Hof", "Hofs", "Hofst", "Hofstr", "Hofstra", "Hofstraa" en "Hofstraat".

De vrije geocodeerservice wordt niet via de PDOK-viewer ontsloten, maar wordt alleen als een aparte service beschikbaar gesteld. Het verschil met de suggest-service is dat er in alle velden wordt gezocht. Zolang alleen de BAG via Locatieserver wordt ontsloten, is de toegevoegde waarde hiervan niet heel groot, omdat de meest relevante informatie via de weergavenaam wordt ontsloten. De weergavenaam is de tekst die in de PDOK-viewer wordt getoond wanneer een geo-object is gevonden. Wel kan bij de vrije geocodeerservice alle informatie over de gevonden objecten in één keer opgevraagd worden.

## Suggest service

De suggest service wordt ontsloten via het suggest-endpoint.

### Voorbeeld:

```
https://geodata.nationaalgeoregister.nl/locatieserver/v3/suggest?wt=xml&q=hofstraat%20110
```

Betekenis parameters: \* q: hierin kan de zoekstring worden opgegeven. Wanneer zoektermen bij elkaar horen, bijv. bij de woonplaats De Bilt, dan kunnen ze tussen quotes worden geplaatst, bijv. q="De Bilt". \* wt: hiermee kan het outputformaat worden opgegeven. Standaard is dit json, maar de PDOK-viewer maakt gebruik van XML. \* indent=true (niet getoond): hiermee kan de teruggegeven JSON-output worden geïndenteerd, zodat het beter leesbaar is, standaard false. \* rows=\ (niet getoond): het maximaal aantal terug te geven zoekresultaten, standaard 10. \* start=\ (niet getoond): de index van het eerste terug te geven zoekresultaat, standaard 0. De parameters rows en start kunnen worden gebruikt voor paginering. \* fq: hiermee kan een filter query worden opgegeven (bijv: fq=type:adres), of de standaard filter query worden opgeheven (d.m.v. fq=\*:\*).

Dit levert een XML-document op met daarin de zoekresultaten. Deze bevraging levert de volgende resultaten op:

ID	Type	Weergavenaam
adr-cd8ae950d6870241e8f3a2208e6fd086	adres	Hofstraat 110, 7311KZ Apeldoorn
adr-769ce7b94efc5bcd615fe42850aa9189	adres	Hofstraat 110, 8261BP Kampen
adr-b34b6d8d843065237656e006cc5a0042	adres	Hofstraat 110, 1779CG Den Oever

Tevens worden twee andere adressen in dezelfde straat in Kampen teruggegeven met huisnummer

'11-01' en '11-02'. Dit heeft er mee te maken dat bij het indexeren huisnummers, -letters en -toevoegingen op verschillende manieren worden gecombineerd, waardoor deze twee adressen ook als match worden gezien. Wel is hiervan de score lager. In een toekomstige versie zal dit worden verwijderd.

De structuur van het teruggegeven XML-document is als volgt:

```
<?xml version="1.0" encoding="UTF-8"?>
<response>
  <result name="response" numFound="5" start="0" maxScore="16.553986">
    <doc>
      <str name="type">adres</str>
      <str name="weergavenaam">Hofstraat 110, 7311KZ Apeldoorn</str>
      <str name="id">adr-cd8ae950d6870241e8f3a2208e6fd086</str>
      <float name="score">16.553986</float>
    </doc>
    ...
  </result>
  <lst name="highlighting">
    <lst name="adr-cd8ae950d6870241e8f3a2208e6fd086">
      <arr name="suggest">
        <str>&lt;b&gt;Hofstraat&lt;/b&gt; &lt;b&gt;110&lt;/b&gt;, 7311KZ
Apeldoorn</str>
      </arr>
    </lst>
    ...
  </lst>
  <lst name="spellcheck">
    <!-- Niet gebruikt -->
  </lst>
  <lst name="collations">
    <!-- Niet gebruikt -->
  </lst>
</response>
```

Betekenis XML-elementen: \* In response/result wordt in het attribuut "numFound" het totale aantal zoekresultaten gegeven, alsmede de gebruikte start-index in het attribuut "start". \* Bovengenoemd element bevat een lijst met "documenten" met daarin de zoekresultaten. Dit zijn gewoon records. Alleen de velden id, type, weergavenaam en score worden teruggegeven. Merk op dat de veldvolgorde kan wijzigen. \* In response/lst[name='highlighting'] wordt de highlighting-informatie teruggegeven. Aangezien de weergavenaam is gebruikt om te zoeken, wordt de weergavenaam teruggegeven met daarin HTML-opmaak (vanwege de weergave als XML zijn sommige karakters omgezet). Hiermee wordt gemarkeerd met welke zoektermen een match is. In dit geval alle componenten van de straatnaam en het huisnummer, maar niet de postcode of woonplaats.

De spellcheck en collations worden niet gebruikt. Aan de inrichting hiervan is nog geen aandacht besteed. Deze worden gevuld wanneer er weinig resultaten gevonden zijn. Een collation is een alternatieve query met de spellcheck resultaten.

Voorbeeld met synoniemen:

<https://geodata.nationaalgeoregister.nl/locatieserver/v3/suggest?wt=xml&q=ln%20v%20west%20701>

Deze bevraging levert de volgende resultaten op:

ID	Type	Weergavenaam
adr-b78fd594f712007b916132b5c5e0480d	adres	Laan van Westenenk 701, 7334DP Apeldoorn
adr-115c1d5b090dae4a0402ed5747ea3506	adres	Laan van Wateringse Veld 701, 2548BN 's-Gravenhage
adr-f3171a929f8b3b1e093f4e13141b3e68	adres	Van Woustraat 70-1, 1073LN Amsterdam

Hierbij vallen een aantal zaken op: \* Er is gezocht op "ln", maar er worden resultaten met "laan" teruggegeven. \* Er is gezocht op "west", maar er wordt een resultaat met "w" teruggegeven. \* Tevens wordt er een postcode met 'ln' teruggegeven.

De eerste twee zijn het gevolg van het toepassen van synoniemen. "Ln" is een synoniem van "laan" en "w" van "west". In het eerste geval is dit gewenst en in het tweede geval is dit onbedoeld. Aangezien de teruggegeven resultaten suggesties zijn, is dit acceptabel. Een generieke service als Locatieserver weet immers niet met welk doel de zoekvraag is gesteld. In andere gevallen, bijv. wanneer er gebruik wordt gemaakt van PTT-afkortingen, is het wel gewenst dat "west" en "w" uitwisselbaar zijn, maar alleen in die richting. Het is mogelijk om Solr op deze manier in te richten, maar in het geval van straatnamen (en in mindere mate plaatsnamen) hebben bepaalde ontwerpkeuzes veel impact op het uiteindelijke resultaat. Het ontwikkelteam staat open voor suggesties van de klankbordgroep om Locatieserver beter te tunen.

Voorbeeld met filtering op type en sortering op afstand:

<https://geodata.nationaalgeoregister.nl/locatieserver/v3/suggest?wt=xml&q=type:woonplaats&lat=52.21179&lon=5.96247>

Deze bevraging levert 2501 resultaten op, namelijk alle woonplaatsen. De eerste 5 resultaten zijn:

ID	Type	Weergavenaam
wpl-0bec2fbac4d93b0212919b83b1e08f00	woonplaats	Apeldoorn, Apeldoorn, Gelderland
wpl-11b82956b26d7990d024e13fa4dcc44a	woonplaats	Wenum Wiesel, Apeldoorn, Gelderland
wpl-e11d68ec27e9e95c5f4cf1c1c661ad89	woonplaats	Beemte Broekland, Apeldoorn, Gelderland
wpl-8e17cf28e36ef51e263cff6dc04c468b	woonplaats	Ugchelen, Apeldoorn, Gelderland
wpl-7cdcb117c093169f389092639dbafa93	woonplaats	Teuge, Voorst, Gelderland

De meegegeven coördinaten geven De Grift in Apeldoorn aan, vandaar dat de woonplaats Apeldoorn zelf als eerste wordt gevonden. Merk op dat de woonplaatsen in de database worden weergegeven als centroiden, dus het eerste resultaat hoeft niet de woonplaats te zijn waar de coördinaten zich in bevinden, maar is de dichtstbijzijnde woonplaatscentroïde.

## Lookup service

De lookup service wordt ontsloten via het lookup-endpoint.

### Voorbeeld:

<https://geodata.nationaalgeoregister.nl/locatieserver/v3/lookup?wt=xml&id=adr-cd8ae950d6870241e8f3a2208e6fd086>

Betekenis parameters: \* id: hiermee kan het ID van het te zoeken object worden opgegeven, wat wordt teruggegeven door de suggest service. \* wt: outputformaat, zie suggest service.

Het teruggegeven XML-document is als volgt:

```
<?xml version="1.0" encoding="UTF-8"?>
<response>
  <result name="response" numFound="1" start="0">
    <doc>
      <str name="bron">BAG</str>
      <str name="woonplaatscode">3560</str>
      <str name="type">adres</str>
      <str name="woonplaatsnaam">Apeldoorn</str>
      <str name="huis_nlt">110</str>
      <str name="openbareruimtetype">Weg</str>
      <str name="gemeentecode">0200</str>
      <str name="weergavenaam">Hofstraat 110, 7311KZ Apeldoorn</str>
      <str name="straatnaam_verkort">Hofstr</str>
      <str name="id">adr-cd8ae950d6870241e8f3a2208e6fd086</str>
      <str name="gemeentenaam">Apeldoorn</str>
      <str name="identificatie">0200010000090244-0200200000007079</str>
      <str name="openbareruimte_id">0200300022471548</str>
      <str name="provinciecode">PV25</str>
      <str name="postcode">7311KZ</str>
      <str name="provincienaam">Gelderland</str>
      <str name="centroide_ll">POINT(5.96244217 52.21173368)</str>
      <str name="nummeraanduiding_id">0200200000007079</str>
      <str name="adresseerbaarobject_id">0200010000090244</str>
      <int name="huisnummer">110</int>
      <str name="provincieafkorting">GD</str>
      <str name="centroide_rd">POINT(194315.783 469449.074)</str>
      <str name="straatnaam">Hofstraat</str>
      <arr name="gekoppeld_perceel">
        <str>APD01-U-9473</str>
        <str>APD01-U-9474</str>
        <str>APD01-U-9475</str>
        <str>APD01-U-9477</str>
        <str>APD01-U-9478</str>
        <str>APD01-U-9479</str>
        <str>APD01-U-9480</str>
        <str>APD01-U-9481</str>
      </arr>
    </doc>
  </result>
</response>
```

Ook hier wordt een response-document teruggegeven, met daarin de resultaten. Het aantal teruggegeven documenten is in principe altijd nul of één. Indien er meer documenten gevonden worden, duidt dit op een niet-uniek ID in de index. Dit is met de huidige configuratie niet mogelijk. (Documenten met dubbele ID's worden bij het indexeren overschreven.)

Opmerkingen: \* Alleen de attributen waarvoor een waarde bekend is, worden teruggegeven. Hier ontbreekt bijv. het attribuut "toevoeging". \* De sortering van de attributen is willekeurig. \* De geometrievelden (geometrie\_rd en geometrie\_ll) worden niet standaard teruggegeven, omdat deze soms veel data kunnen bevatten (nu woonplaatsen en gemeenten, later ook bij wegen). Ze kunnen worden opgevraagd door de queryparameter fl=\* aan het request mee te geven.

## Vrije geocodeerservice

De vrije geocodeerservice wordt ontsloten via het free-endpoint. Deze wordt in dit document niet verder behandeld.

## Vulling

Op dit moment bevat de testomgeving van Locatieserver de volgende data:

Type	ID-prefix	Geometrie	Standaard beschikbaar in suggest en free
gemeente	gem	vlak	ja
woonplaats	wpl	vlak	ja
weg (BAG openbare ruimte)	weg	punt (centroïde van de omvattende adressen)	ja
postcode	pcd	punt (centroïde van de omvattende adressen)	ja
adres	adr	punt	ja
perceel	pcl	vlak	nee
hectometerpaal	hmp	punt	nee
waterschapsgrens	wsg	vlak	nee
provincie	pro	vlak	nee
wijk	wij	vlak	nee
buurt	buu	vlak	nee

Om percelen op te vragen met de suggest- en free-services dient de query parameter fq=\* : \* te worden toegevoegd.

Bij het indexeren wordt wel rekening gehouden met de volgende zaken: \* Hoofdletter-ongevoelig; \* Diakrieten-ongevoelig; \* Uitwisselbaarheid "ij" vs. "y"; \* Afkortingen / synoniemen: \* Woonplaatsen; \* Straatnaamsuffixes (-straat/-weg/-laan/...); \* Huislettertoevoegingen (bis/huis/zwart/rood/...) \* Stopwoorden (voorzetsels/lidwoorden); \* Rangtelwoorden (eerste/tweede/...); \* Windrichtingen / zijden (noord/noordelijke/noordzijde/...); \* Titels (burgemeester/dokter/meester/kardinaal/...); \* Veelvoorkomende bijvoegelijke naamwoorden (korte/gedempte/oude/...); \* Overig (kasteel). \* Leestekens-ongevoelig; \* Spaties: ieder woord in de databron of zoekvraag wordt als een aparte

zoekterm gezien.

## Sortering resultaten

De sortering van de resultaten van de suggest- en vrije geocodeerservices wordt op de volgende manier bepaald:

1. Scoreberekening Solr;
2. Sortering huisnummers (1, 1A, 1B, 2, 2A, 3, ...) of sortering secties/perceelnummers (... , Y, Z, AA, AB, ...);
3. Alfabetische sortering.

De scoreberekening van Solr is een wiskundige berekening, welke van meerdere factoren afhangt:

\* De mate waarin de zoektermen overeenkomen met de zoekresultaten; \* Hierbij wordt de voorkeur gegeven aan korte namen, omdat dan een groter deel overeenkomt; \* Wanneer er meerdere zoektermen opgegeven zijn (gescheiden door spaties), wordt de score hoger naarmate er meer matches zijn. De woonplaats De Bilt zal dus hoger scoren dan Bilthoven of De Lier; \* Tevens wordt er rekening gehouden met boosting parameters. \* Bij het meegeven van lat/lon wordt ook de afstand meegenomen.