VOGIN-IP-lezing 2022
11 mei - OBA Amsterdam

Het VOGIN-IP-team heet je van harte welkom bij de 10de editie van de VOGIN-IP-lezing, het jaarlijkse evenement waar vindbaarheid van informatie centraal staat. We wensen je veel plezier met lezingen en workshops, en met je contacten met vakgenoten.

De VOGIN-IP-lezing is een gezamenlijk initiatief van
Stichting VOGIN en vakblad IP

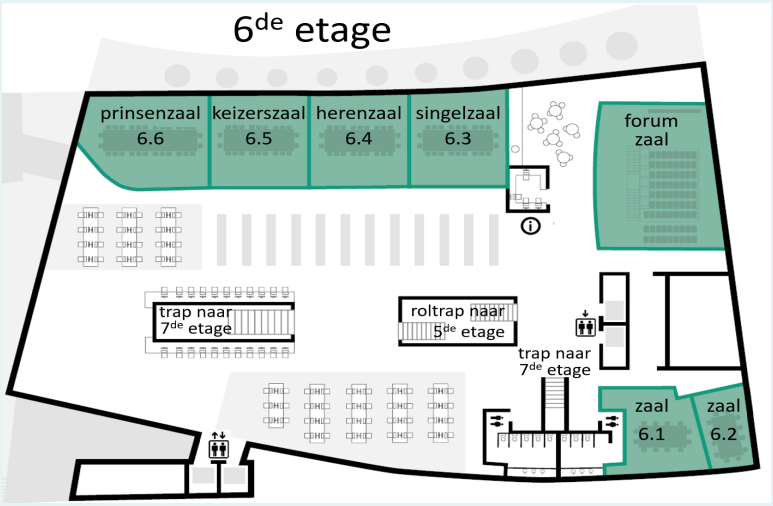
Programma VOGIN-IP-lezing 11 mei 2022

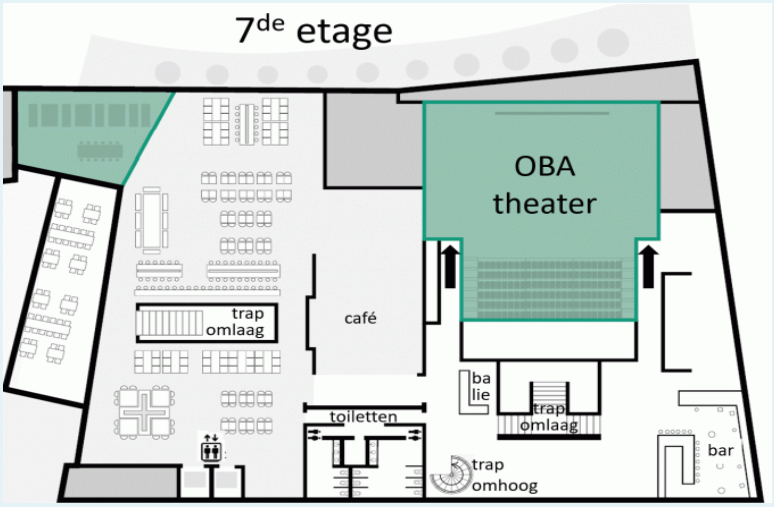
Dagvoorzitter:	Peter van Gorsel	
9.00 – 9.45	Ontvangst en registratie	
9.45– 10.30	*Plenair programma in Theaterzaal*	
9.45 – 9.50	Opening	
9.50 – 10.30	**Keynote: Andrew Yates** **Improving search with neural ranking methods**	
10.30 – 10.55	Koffiepauze	
10.55 – 12.55	*Keuze tussen lezingentrack (in Theaterzaal) en workshops (zie volgende bladzijden)*	
10.55 – 11.35	Brecht Castel - Hoe online onderzoek helpt in de strijd tegen desinformatie	workshops ↓
11.35 – 12.15	Frank van Harmelen - Van de droom van het Semantic Web naar de realiteit van Linked Open Data	
12.15 – 12.55	Enno Meijers - Termennetwerk – een katalysator voor Verbonden Erfgoed	
12.55 – 13.50	Lunch en netwerken	
13.50 – 15.50	*Keuze tussen lezingentrack (in Theaterzaal) en workshops (zie volgende bladzijden)*	
13.50 – 14.30	Cynthia Liem - Als zoeken te fanatiek wordt: een digitale analyse van het toeslagenschandaal	workshops ↓
14.30 – 15.10	Ivo Zandhuis & Merel Geerlings - Records in Contexts – nieuwe metadatastandaard Stadsarchief Amsterdam	
15.10 – 15.50	Paul Groth - Minimum viable data reuse	
15.50– 16.15	Theepauze	
16.15 – 17.05	*Plenair programma in Theaterzaal*	
16.15 – 17.00	**Keynote: Bregtje van der Haak & Rana Klein** **Archief van de toekomst**	
17.00 – 17.05	Afsluiting	
17.05 – 18.30	Borrel, hapjes en netwerken	

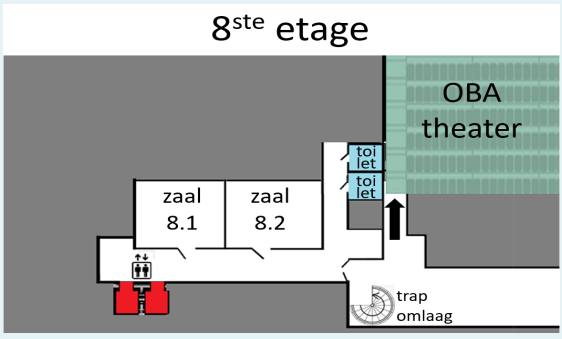
Workshops

docenten	*titel*	*tijden*	*locatie **
Joyce van Aalten	Aan de slag met knowledge graphs	10.55-12.55 13.50-15.50	zaal 8.2
Wouter Beek Thomas de Groot	Werken met Liinked Data met de TriplyDB software	10.55-12.55	zaal 6.2
Jeroen Bosman Bianca Kramer	Systematisch zoeken op internet; kan dat?	10.55-12.55 13.50-15.50	zaal 6.4 Herenzaal
Guus van den Brekel Robin Ottjes	Hoe kom ik nu aan de full-text?	13.50-15.50	zaal 6.6 Prinsenzaal
Erik Elgersma	Een startende éénpitter in informatie-land	13.50-15.50	zaal 6.2
Sandra Fauconnier	Bewerken van datasets met OpenRefine	10.55-12.55 13.50-15.50	zaal 6.1
Kristina Hettne	Een gereedschapskist voor digitale vaardigheden	10.55-12.55	zaal 6.6 Prinsenzaal
Max Kemman	Grote hoeveelheden tekst analyseren als data	10.55-12.55	zaal 8.1
Hanno Lans	Werken met Wikidata	13.50-15.50	zaal 8.1
Leon Pauw (i.p.v. Alexander Pleijter)	Zo word je factchecker	10.55-12.55 13.50-15.50	zaal 6.5 Keizerszaal
Ewoud Sanders	Slimmer zoeken in Delpher	10.55-12.55 13.50-15.50	zaal 6.3 Singelzaal

* Zie volgende pagina voor plattegronden met zaallocaties


6de etage
prinsenzaal
6.6
keizerszaal
6.5
herenzaal
6.4
singelzaal
6.3
forum
zaal
trap naar
7de etage
roltrap naar
5de etage
trap naar
7de etage
zaal
6.1
zaal
6.2


7de etage
OBA
theater
trap
omlaag
café
toiletten
ba
lie
trap
omlaag
trap
omhoog
bar


8ste etage
OBA
theater
toi
let
toi
let
zaal
8.1
zaal
8.2
trap
omlaag

Lezingen *[meer tekst op https://vogin-ip-lezing.net/lezingen-2022/]*

Andrew Yates

Improving search with neural ranking methods

Text ranking, the task of ordering pieces of text based on their relevance to a query, is a common task that appears in many search scenarios. Web search is a common example. At its core, text ranking relies on a ranking method to produce good estimates of how well different pieces of text (e.g., Web pages) match a given query, though many other signals may ultimately influence a ranking. Traditionally, such ranking methods relied on term statistics to balance the discriminativeness of a query term with the term's importance in a piece of text. In recent years, however, deep learning has begun to reshape how text ranking is performed by improving textual understanding. These neural ranking methods have substantially increased ranking quality while breaking the rules that successful statistical ranking methods follow. In this talk, I will take a deep dive into these methods to explore why they work so well, where they fall short, how they can already be leveraged to improve search, and what the future might bring.

Brecht Castel

Hoe online onderzoek helpt in de strijd tegen desinformatie

We delen dagelijks foto's en video's op sociale media, vaak zonder nadenken of de bron te checken. Zo kan desinformatie zich als een lopend vuurtje verspreiden. Hoe hou je dat tegen? Door met gratis online tools snel en efficiënt de juiste context van beelden te achterhalen. Dan deel je pas als iets klopt. In deze lezing krijg je enkele praktische tips & tricks aangereikt om zelf te factchecken. Die zoektechnieken kunnen trouwens ook helpen op totaal andere terreinen. Iedereen kan zijn voordeel doen met wat OSINT: Open Source INTelligence, het gebruiken van publieke online bronnen voor onderzoek.

Frank van Harmelen

Van de droom van het Semantic Web naar de realiteit van Linked Open Data

Al twee decennia wordt er gesproken over (en gewerkt aan) een "semantic web": een verrijking van het world wide web waarop niet alleen mensen informatie uitwisselen, maar waarop ook computers op basis van formele representaties informatie en kennis kunnen vinden, koppelen en uitwisselen. Na 20 jaar heeft dat "semantic web" nog niet bepaald de huiskamers van consumenten en de leeszalen van bibliotheken bereikt. Moeten we spreken van een mislukking? Deze lezing vertelt hoe de droom van het semantic web wel degelijk dagelijkse praktijk geworden is in de vorm van "linked open data", en hoe zulke linked open data de moderne grondslag is voor informatie uitwisseling tussen wetenschappers onderling, tussen bedrijven en consumenten, en tussen overheden en burgers.

Enno Meijers

Termennetwerk – een katalysator voor Verbonden Erfgoed

Het Netwerk Digitaal Erfgoed (NDE) werkt, als samenwerkingsverband van Nederlandse erfgoedinstellingen, aan het beter vindbaar maken van erfgoedinformatie. Naar schatting ruim 2000 erfgoedinstellingen in Nederland publiceren digitale informatie over hun collecties. Het vinden van de beschikbare informatie over meerdere instellingen heen is echter voor veel gebruikers een lastige opgave. Om die reden werkt NDE aan diensten die het samenbrengen van verspreide erfgoedinformatie eenvoudiger maakt. Een van deze diensten is het Termennetwerk. Het Termennetwerk helpt om bij het beschrijven van erfgoedobjecten snel de juiste gestandaardiseerde definitie te vinden voor bijvoorbeeld personen of onderwerpen. Door gebruik van deze gestandaardiseerde termen kan informatie beter met elkaar verbonden worden. Zeker wanneer de informatie ook nog als Linked Data beschikbaar gemaakt wordt. In tegenstelling tot bestaande vergelijkbare oplossingen zoekt het Termennetwerk realtime in meerdere terminologiebronnen. De zoekresultaten worden geharmoniseerd en teruggegeven via een standaard beschrijvingsformaat (SKOS) en zoekinterface (GraphQL). Ook de koppeling met de terminologiebronnen verloopt via een standaard zoektaal (SPARQL) waardoor elke bron, die dit Linked Data protocol ondersteunt, koppelbaar is. In de presentatie wordt uitgebreid ingegaan op de functionaliteit van het Termennetwerk en de mogelijkheden om deze open source software ook voor andere informatiedomeinen in te zetten..

Ivo Zandhuis

Merel Geerlings

Records in Contexts – nieuwe metadatastandaard Stadsarchief Amsterdam

De nieuwe beschrijvingsstandaard voor archieven heet Records in Contexts, en het Stadsarchief Amsterdam is een van de eerste organisaties die de nieuwe standaard van de International Council on Archives implementeert. Het Stadsarchief was direct enthousiast over de nieuwe standaard, omdat deze weet om te gaan met een van de grootste uitdagingen waar het archief momenteel voor staat: het beschrijven van digitaal geboren en gedigitaliseerd archief. Ook is Records in Contexts gebaseerd op Linked Data, wat voor de gebruiker een heel scala aan nieuwe mogelijkheden biedt om informatie te zoeken en te vinden. In deze presentatie meer over de standaard zelf, welke mogelijkheden het biedt, waarom het Stadsarchief erop inzet en wat dat deze keuze betekent voor de gebruiker.

Cynthia Liem

Als zoeken te fanatiek wordt: een digitale analyse van het toeslagenschandaal

In het veelbesproken toeslagenschandaal speelden digitale en algoritmische componenten een belangrijke rol. Hoewel er in het toeslagenschandaal sprake was van een toepassing die ver af lijkt te staan van de informatieprofessional (automatische inschatting van onrechtmatigheid/fraude), hebben de gebruikte digitale en algoritmische componenten wel degelijk veel overeenkomsten met componenten waar een informatieprofessional mee te maken zou krijgen in zoek- en aanbevelingsscenario's. In mijn presentatie sta ik hierbij stil, en deel ik mijn bredere ervaringen rond de voorbereiding van het Trouw-artikel dat de algoritmische kant van het toeslagenschandaal belichtte. Om de discussie van 'wat een algoritme exact doet' goed te kunnen voeren, is het minstens zo belangrijk om bredere digitale geletterheid en een meer systemische blik te bevorderen, uitdrukkelijk ook bij mensen die zich van nature als 'niet-technisch' identificeren.

Paul Groth

Minimum viable data reuse

There are a myriad of recommendations, advice, and guidelines about what data providers should do to facilitate data reuse. It can be overwhelming. Based on recent empirical work (analyzing data reuse proxies at scale, understanding data sensemaking and looking at how researchers search for data), Paul Groth will talk about what practices are a good place to start for helping others to reuse your data..

Bregtje v.d. Haak

Rana Klein

Archief van de toekomst

Samen met ontwerpbureau Sudox en het Nederlands Instituut voor Beeld en Geluid ontwikkelen VPRO Tegenlicht en VPRO Medialab het 'Archief van de Toekomst'. Dit online archief gebruikt kunstmatige intelligentie om meer dan 500 uitzendingen van Tegenlicht gemaakt sinds 2002 op nieuwe manieren beschikbaar en doorzoekbaar te maken. Met gebruik van algoritmes voor beeld-herkenning en spraak- en tekstanalyse zijn alle uitzendingen omgezet in een dataset. Daarin kunnen gebruikers zelf zoeken naar specifieke fragmenten, citaten en zelfs losse shots. De op tijdlijn en thema's gebaseerde website is bedoeld voor alle geïnteresseerden, zoals scholieren, docenten, onderzoekers, studenten, kunstenaars en ontwerpers. Zij kunnen grasduinen in fragmenten en dwars-verbanden ontdekken om tot nieuwe visies op de toekomst te komen. Archief van de toekomst is gebaseerd op Open Standaarden en Open Source technologie.

Workshops *[zie ook: https://vogin-ip-lezing.net/workshops-2022/]*

Joyce van Aalten

Aan de slag met knowledge graphs

We kennen de knowledge graph inmiddels al een tijd, en komen de knowledge graph in steeds meer hoedanigheden tegen. Uiteraard is daar de Google Knowledge Graph en de Facebook Graph. En we horen over knowledge graphs binnen websites zoals Zalando of AirBNB. Maar ook binnen organisaties en specifieke organisatiedomeinen speelt een (enterprise) knowledge graph steeds vaker een belangrijke rol of wordt ermee geëxperimenteerd. Maar hoe ga je nu zelf aan de slag met je eigen knowledge graph? Deze workshop begint met een korte introductie: wat zijn knowledge graphs, hoe zijn ze opgebouwd, hoe worden ze gebruikt en waarin worden ze opgeslagen. En wat hebben kreten als RDF, SKOS, triples, ontologieën en Linked Open Data hiermee te maken? Maar deze workshop is vooral: zelf aan de slag gaan! Een knowledge graph bouw je uiteraard niet in een uurtje. Maar we gaan wel samen de eerste stappen zetten: bepalen hoe je een knowledge graph opzet en hoe je die kunt toepassen. Zo maak je kennis met de belangrijkste basisprincipes, uitgangspunten en tools, die je na de workshop zelf kunt blijven toepassen.

Wouter Beek

Thomas de Groot

Werken met Linked Data met de TriplyDB software

Organisaties wisselen steeds meer verschillende data met elkaar uit. Hierdoor wordt het alsmaar lastiger om data efficiënt te combineren, en ontstaan steeds meer kopieën van dezelfde gegevens. Linked data is een nieuwe manier om verschillende databronnen met elkaar te verbinden. Het is bovendien gebaseerd op open standaarden. Triply ontwikkelt met TriplyDB het eerste SaaS-georiënteerde database product waarmee op een integrale manier met linked data gewerkt kan worden.
In deze workshop laten Wouter en Thomas middels concrete voorbeelden zien wat de voordelen van linked data zijn. Deelnemers krijgen de mogelijkheid om zelf data als linked data in TriplyDB te uploaden en gebruiken. Hiermee worden aanknopingspunten geboden over hoe jij linked data binnen jouw organisatie kunt inzetten om jouw werk makkelijker te maken.

Jeroen Bosman

Systematisch zoeken op internet kan niet …. of wel?

Wie zoekt naar informatie wil dat graag goed doen, met een hoge “recall & precision”. Ook wil je graag controle over je zoekproces en dat liefst later kunnen reproduceren. In extreme mate gelden die eisen bij zoeken naar wetenschappelijke literatuur voor zogenaamde ‘systematic reviews’. De vraag is of je diezelfde systematische zoekaanpak ook kunt toepassen op serieuze internetzoekacties om bij een goed omschreven onderzoeksvraag nieuws, rapporten, websites, blogs en meer te kunnen vinden. In

Bianca Kramer

deze workshop wordt u, deels via demonstraties, deels via eigen zoekacties en deels via discussie, langs alle stappen geleid die samen aangeduid kunnen worden als 'systematic search applied to the web'. Het wordt een spannende tocht want vergeleken met goed georganiseerde wetenschappelijke literatuur stelt het internet je voor veel meer valkuilen en verrassingen. Toch zal blijken dat er met de nodige creativiteit meer mogelijk is dan u misschien denkt.

Guus v/d Brekel

Robin Ottjes

Hoe kom ik nu aan de full-text?"

Thuiswerken en off-campus toegang tot wetenschappelijke informatie en bibliotheekdiensten is crucialer dan ooit door Corona. Daarnaast is toegang tot tijdschriften erg prijzig en veel instituten hebben niet de middelen om een licentie op alles te nemen. Vooral organisaties buiten de academische wereld worstelen met toegang tot onderzoek. Organisaties met toegang moeten bovendien de vraag stellen, wat te doen als ze plotseling de toegang tot een reeks tijdschriften verliezen omdat het abonnement is geannuleerd? Als de bibliotheek een artikel niet kan leveren, weten we dat onze gebruikers andere manieren zullen gebruiken om de pdf te krijgen. Echter, deze andere manieren bespreken we bijna nooit. Hoeveel verschillende manieren zijn er precies? Hoe werken ze? Hoe kan ik het groeiende aantal open access-artikelen -die her en der verspreid zijn over het web- op de meest efficiënte manier vinden? Deelnemers aan de workshop leren over alle mogelijke manieren en hulpmiddelen en krijgen tips om de full-tekst van wetenschappelijke publicaties te vinden. We bespreken "best practices" van diensten en bibliotheektools (Linkresolvers, Discovery, LeanLibrary, LibX, Apps, etc.) en testen vooral alle alternatieve tools zoals EndNote Click (voorheen Kopernio), Unpaywall, CORE Discovery, OpenAccess-Button, LibKey Nomad, Google Scholar-Button en andere extensies of bookmarklets.

Erik Elgersma

Een startende éénpitter in informatie-land

In de workshop evalueert Erik Elgersma na de eerste twee jaar wat wel en niet werkte in zijn informatie-éénpitter rol. Ter inspiratie gebruikt hij actuele gegevens van zijn eigen bedrijf als case study. De evaluatie bestaat uit vragen die relevant zijn voor elke informatie professional:

Wie zijn je klanten en hoe goed ken je ze?
Wie is eigenlijk je concurrent en hoe kun je daarvan winnen?
Hoe zorg je dat je beoogde klanten jou kennen?
Wat werkt? Wat werkt niet?
Hoe vind je uit waarmee je ze het beste kan bedienen?
Hoe innoveer je je diensten om ze nog beter te kunnen bedienen?

Na afloop van de workshop heeft de deelnemer vragen en antwoorden ter reflectie. De workshop kan bijdragen aan het nog beter kunnen invullen van een rol als informatie-professional.

Sandra Fauconnier

Bewerken van datasets met OpenRefine

OpenRefine is gratis open source software om datasets te bewerken en op te schonen. OpenRefine is krachtig genoeg om grote datasets (tot honderdduizenden records) te manipuleren, te transformeren tussen verschillende formaten, en te koppelen (reconcilen) met externe databronnen en kennisbanken, waaronder Wikidata. Wereldwijd wordt OpenRefine gebruikt door bibliotheek-medewerkers, onderzoekers, datajournalisten en in vele projecten rond Linked Open Data; het is ook een populaire tool in de Wikimediagemeenschap. Programmeerervaring is niet nodig. Bovendien is OpenRefine bijzonder privacyvriendelijk: de software draait als een server op je eigen machine en de gebruikersinterface bevindt zich in je browser. Je bewerkt je data lokaal, en deelt alleen datasets met anderen als je daar zelf voor kiest.
In deze workshop leer je alle basishandelingen van de software: data analyseren, filteren, transformeren, clusteren, en 'reconcilen' met Wikidata. Na de workshop ben je zelfstandig in staat om met OpenRefine aan de slag te gaan..

Kristine Hettne

Een gereedschapskist voor digitale vaardigheden

Digitale vaardigheden ontwikkelen: hoe doe je dat? Welke tools heb je nodig en wanneer gebruik je die? Van dagelijkse databewerking en het maken van mooie figuren voor rapporten tot het goed kunnen communiceren met onderzoekers en hen ondersteunen. Het zijn allemaal voorbeelden waar digitale vaardigheden van pas komen. In de workshop gaan we samen ontdekken hoe tools zoals spreadsheetprogramma's, OpenRefine, R en RMarkdown je in je werk kunnen ondersteunen en wanneer je het beste welke tool kan gebruiken. Om optimaal gebruik te kunnen maken van digitale tools zijn er basale digitale vaardigheden nodig. We kijken samen naar de lessen van de "Carpentries" als een voorbeeld van een leeromgeving waar je ook zelf mee aan de slag kan. Het doel van de workshop is om ervaringen uit te wisselen en nieuwe vaardigheden op te doen om met een goede basis verder zelf te kunnen leren.

Max Kemman

Grote hoeveelheden tekst analyseren als data

Er wordt steeds meer tekst geproduceerd. Gelukkig komen er ook steeds meer tools beschikbaar om grip te houden op de grote hoeveelheid informatie die beschikbaar komt. In deze workshop leren we een aantal tools die helpen om overzichten te creëren van (grote) datasets van teksten. Zo is het mogelijk om overzichten te maken van boeken, wetenschappelijke papers, kranten, tweets of e-mails en te ontdekken hoe verschillende thema's zich door de tijd ontwikkelen Tijdens de workshop gebruiken we hiervoor Voyant Tools, een online omgeving waarin verschillende van dit soort overzichten gemaakt kunnen worden zoals woordenwolken, topic models, woordentrends en afwijkende woorden per uitsnede. Om te oefenen kan je werken met een collectie teksten die ik klaar zet, maar als je zelf een dataset van teksten hebt is het nog mooier.

Hanno Lans

Werken met Wikidata

Wikidata is een vrij bewerkbare linked open dataset waaraan wereldwijd informatie wordt toegevoegd. Dit heeft voor instellingen als voordeel dat de soms summiere informatie die in de eigen data aanwezig is via dit kanaal verrijkt wordt. In de workshop wordt vooral gekeken naar de mogelijkheden die Wikidata biedt om de datakwaliteit van datasets van erfgoedinstellingen te verhogen. Dit geeft een instelling met betrekkelijk weinig inspanning toegang tot gegevens als man/vrouwverdeling, auteursrechtvertegenwoordigers en welke andere instellingen werken van een persoon hebben. Ook is de informatie als linked open data beschikbaar. Voor Google, Apple en Amazon is dat een welkome bron voor het aanbieden van informatie in verschillende applicaties. Doordat Wikidata is opgezet als een referentiedatabase, betekent dit dat de toegeleverde informatie en de bronnen daarvan beschikbaar komen in zoeksystemen zoals Google. Aan de hand van de collecties van Paleis het Loo en Wereldculturen kijken we hoe op gestructureerde wijze data toegevoegd, gemonitord en teruggehaald kan worden. Ook gaan we praktisch in op de werkwijze binnen Wikidata, op aandachtspunten waarmee rekening moet worden gehouden bij gebruik van dergelijke dynamische data en op handige hulpmiddelen zoals OpenRefine om zelf data te kunnen bewerken.

Leon Pauw

(vervanger voor Alexander Pleijter)

Zo word je factchecker

Fake news! Het is inmiddels een gevleugelde term. Als een journalist een fout maakt of een verzonnen bericht rumoer veroorzaakt, dan valt al gauw de term ‘nepnieuws’. Nieuwscheckers houdt zich sinds 2009 bezig met het checken van discutabele nieuwsberichten. Berichten van gerenommeerde media, maar ook van vage clickbaitwebsites. Sinds 2021 maakt Nieuwscheckers deel uit van een Nederlands-Vlaams consortium van factcheckers, nieuwsmedia en onderzoekers dat zich bezighoudt met de bestrijding van desinformatie. In deze workshop gaat het over de diverse kanten van het factchecken: hoe controleer je of nieuws klopt? Welke tools gebruik je daarbij? Hoe ga je te werk? Aan de hand van diverse voorbeelden en opdrachten krijg je een kijkje in de keuken van de factchecker.

Ewoud Sanders

Slimmer zoeken in Delpher

Met ruim 100 miljoen gedigitaliseerde pagina’s uit Nederlandse kranten, boeken en tijdschriften is Delpher.nl een ware goudmijn. Waar je meer in vindt als je verschillende zoektechnieken toepast. In deze workshop, waarbij de deelnemers zelf gaan zoeken, laat Ewoud Sanders zien hoe je het meeste uit Delpher kunt halen. Als auteur van verschillende boekjes over zoeken op internet en in Delpher in het bijzonder, is Ewoud je aangewezen gids hierbij.

#voginip http://vogin-ip-lezing.net
@VoginAcademie @VOGIN_IP
https://www.facebook.com/voginip

netwerk:	OBA Wireless
naam:	congres2021
wachtwoord:	vanhartewelkom

Met speciale dank aan sponsoren en partners

LexisNexis®
https://www.lexisnexis.nl/dutch/home.page

EBSCO
https://www.ebsco.com/e/nl-nl

ProQuest
https://about.proquest.com/en/

INGRESSUS
http://www.ingressus.nl/

GO Stichting GO Fonds
meedoen in de informatiesamenleving
https://www.gofonds.nl/pages/1/home.html

knvi
https://www.knvi.nl/

GO opleidingen | School voor informatie
https://www.goopleidingen.nl/

VISION4EVENTS