

BIG DATA



POSTACADEMISCHE OPLEIDING
26 FEBRUARI 2015 – 4 JUNI 2015

'Big Data' kunnen omschreven worden als gegevenscollecties die niet efficiënt met traditionele gegevensbeheer en -verwerkingsmethodes kunnen worden behandeld. Bepalende factoren daarbij zijn de grotere datavolumes, de grotere snelheden waarmee de data worden aangeboden en de grotere variëteit aan dataformaten. De tendens naar 'Big Data' wordt gevoed door de almaar groeiende beschikbaarheid van digitale informatie uit nieuwsbronnen, multimedia, sensors, ... en gaat gepaard met nieuwe uitdagingen om deze data efficiënt te kunnen verzamelen, opslaan, beheren, analyseren en presenteren.

Het inzetten van geavanceerde technologieën die specifiek zijn afgestemd op het verwerken van zeer grote hoeveelheden data, kan bedrijven helpen om beter tegemoet te komen aan de steeds groter wordende informatienoden die vaak vereist zijn om gegevensanalyse nog beter te kunnen onderbouwen. Een beter inzicht in de beschikbare data en een optimale exploitatie ervan levert de beste garantie om met meer kennis van zaken belangrijke beslissingen te onderbouwen en daar dan ook een concurrentieel voordeel mee te behalen.

Doelpubliek

U krijgt inzicht in de problematiek die gepaard gaat met 'Big Data' en in de beschikbare ICT-oplossingen die momenteel voorhanden zijn. Er wordt aangetoond hoe de aangereikte oplossingen werken, wat hun beperkingen en voordelen zijn en waar en wanneer ze het beste kunnen worden ingezet.

De opleiding is dusdanig opgevat dat deze toegankelijk is voor iedereen die ietwat vertrouwd is met informatica.

Er wordt gewerkt rond drie thema's: gegevensbeheer, gegevensanalyse en visualisatie. Theoretisch geïnspireerde hoorcolleges worden afgewisseld door praktische getuigenissen over 'Big Data'-oplossingen uit diverse sectoren zoals de medische en de financiële sector, de media, de overheid en de bibliotheekwereld.

Getuigschrift

U ontvangt een getuigschrift, indien u deelneemt aan de volledige opleiding en slaagt voor het bijbehorende examen.

Wetenschappelijke Coördinatie

Prof. dr. Guy De Tré, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent

Lesgevers

- Vivek Bajaj, IBM
- Antoon Bronselaer, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- Hans Constandt, Ontoforce
- Thomas Demeester, Vakgroep Informatietechnologie, Universiteit Gent
- Wesley De Neve, Vakgroep Elektronica en Informatiesystemen, Universiteit Gent
- Guy De Tré, Vakgroep Telecommunicatie en Informatieverwerking, Universiteit Gent
- Filip De Turck, Vakgroep Informatietechnologie, Universiteit Gent
- Wim De Wispelaere, Amplidata
- Erik Duval, Departement Computerwetenschappen, KU Leuven
- Jan Fostier, Vakgroep Informatietechnologie, Universiteit Gent
- Alain Houf, Intersystems
- Joris Klerckx, Departement Computerwetenschappen, KU Leuven
- Peter Lambert, Vakgroep Elektronica en Informatiesystemen, Universiteit Gent
- Femke Ongenaes, Vakgroep Informatietechnologie, Universiteit Gent
- Dirk Van den Poel, Vakgroep Marketing, Universiteit Gent
- Inge Van Nieuwerburgh, Directie Onderzoeksangelegenheden, Afdeling Universiteitsbibliotheek, Universiteit Gent
- Nico Verplancke, Vlaams Instituut voor Archivering

1. Gegevensbeheer

Inleiding en NoSQL

In de introductie wordt aandacht besteed aan de oorsprong van de term 'Big Data', de interpretatie en het belang ervan, de problematiek die er dikwijls mee gepaard gaat en de kritiek op 'Big Data'.

Daarna komen de verschillende vormen en karakteristieken aan bod. Bijzondere aandacht gaat naar ongestructureerde en semi-gestructureerde gegevens en de aparte informatie-beheersnaden die daarmee gepaard gaan. Vervolgens wordt dieper ingegaan op de belangrijkste NoSQL toegangstechnieken ('Not only' SQL) zoals 'Key-/Value-stores', 'Document Databases' en 'Column-Oriented Databases'.

Lesgever: Guy De Tré

Datum: 26 februari 2015

Datakwaliteit en information retrieval

Er wordt vertrokken vanuit de vaststelling dat modaliteiten voor de garanties van datakwaliteit bij NoSQL opslagsystemen van ondergeschikt belang zijn ten voordele van een betere throughput. De kwaliteitscontrole van data wordt dan ook verschoven naar het niveau van applicaties die gebruik maken van data. Er wordt toegelicht hoe men in een applicatie een dergelijk controlemechanisme kan inbouwen op een efficiënte manier. Nauw aansluitend daarbij worden technieken voor dataprofielering toegelicht. Er wordt een overzicht gegeven van een aantal specifieke kwaliteitsproblemen en diens oplossingen.

Daarna is er een inleiding op 'Information Retrieval' (IR), de technologie die toelaat om grote digitale collecties van ongestructureerde teksten efficiënt te doorzoeken. De bedoeling is eerder om een overzicht te geven van de basisconcepten zoals indexering en retrieval modellen, dan om hier heel technisch op in te gaan.

Lesgevers: Antoon Bronselaer en Thomas Demeester

Datum: 5 maart 2015

Web search en documentclassificatie

Er wordt dieper ingegaan op 'Information Retrieval' binnen de context van het World Wide Web. Specifieke uitdagingen en mogelijkheden rond web search (o.a. het gebruik van hyperlinks voor nieuwe zoekmodellen) en een aantal gerelateerde onderwerpen ('crawlers' en 'web spamming') komen aan bod.

Na de onderdelen rond het doorzoeken van grote tekst-gebaseerde collecties, worden de mogelijkheden bekeken om in dergelijke collecties structuur aan te brengen door de documenten op specifieke manieren te gaan groeperen. Enkele basisconcepten rond classificatie en clustering worden aangebracht, waarna een paar belangrijke algoritmen worden besproken. Vervolgens worden enkele applicaties toegelicht.

Lesgever: Thomas Demeester

Datum: 12 maart 2015

Linked Open Data

De evolutie van het World Wide Web tot een globaal en wereldwijd platform voor de meest uiteenlopende digitale diensten, heeft er toe geleid dat een goede representatie van data op het web steeds belangrijker wordt.

Tijdens deze lesavond worden verschillende technologieën behandeld met het oog op het machine-leesbaar maken van data en informatie op het web (cf. semantisch web). Het koppelen van verschillende databronnen (Linked Data) zal daar een belangrijke rol in spelen. Daarnaast worden ook de principes van Open Data besproken, met als doel het uniform beschikbaar stellen van data via het web.

Linked Open Data leidt aldus tot een aantal specifieke uitdagingen op het vlak van web-gebaseerde data-analyse en interpretatie.

Lesgever: Peter Lambert

Datum: 19 maart 2015

Getuigenissen uit de praktijk, deel I

De eerste spreker is Wim De Wispelaere. Hij is Chief Technology Officer en medeoprichter van Amplidata, een bedrijf dat schaalbare opslagoplossingen biedt voor 'Big Data'. Naast een technologisch overzicht is er aandacht voor de toepassingen in de life sciences, de olie- en de gassector.

De tweede spreker is Nico Verplancke, directeur van het VIAA (Vlaams Instituut voor Archivering). Eind 2012 besliste de Vlaamse regering om een Vlaams Instituut voor Archivering op te richten. iMinds kreeg de opdracht om deze opstart te realiseren. Het VIAA moet als digitaal platform diensten aanbieden om erfgoed, kennis en geschiedenis op een verstandige manier te bewaren en te ontsluiten voor de huidige en toekomstige generaties. Deze lezing laat zien waar het VIAA staat en wat de concrete toekomstplannen zijn.

De derde spreker is Hans Constandt, CEO en medeoprichter van Ontoforce. Ontoforce is een Belgisch bedrijf dat oplossingen aanbiedt voor 'information flow' en management. In deze lezing wordt een gebruiksvriendelijk data zoekplatform voorgesteld dat gedreven is door semantische technologie en in staat is om unieke intelligente links tussen gegevens te bouwen. Tevens wordt een toepassing voor de farmaceutische industrie toegelicht.

Datum: 26 maart 2015

2. Gegevensanalyse

Het MapReduce programmeermodel

MapReduce is een generiek raamwerk dat bijzonder geschikt is om op eenvoudige wijze analyses uit te voeren op enorme hoeveelheden data, gebruik makend van een parallel computersysteem. De basisconcepten rond 'Mappers' en 'Reducers' komen aan bod, alsook enkele veelgebruikte ontwerp patronen. Deze laatste worden gestaafd aan de hand van eenvoudig te begrijpen voorbeelden. De Hadoop-implementatie van MapReduce wordt besproken, alsook het verwante Hadoop Distributed File System (HDFS). De doelstelling is dat deelnemers na deze sessie in staat zijn Hadoop MapReduce toe te passen op eigen problemen.

Voorkennis van de programmeertaal Java is aan te raden, maar niet noodzakelijk.

Lesgever: Jan Fostier

Datum: 2 april 2015

Gedistribueerde gegevensverwerking

Tijdens deze lesavond worden ontologieën geïntroduceerd als veelgebruikte manier om kennis te representeren in gedistribueerde applicaties. Ook alternatieve manieren voor kennisrepresentatie worden behandeld. Vervolgens komen 'reasoners' aan bod die op basis van ontologieën en geobserveerde data in staat zijn om nieuwe kennis af te leiden. De combinatie van 'reasoners' en 'rule engines' wordt toegelicht. Schaalbaarheid van 'reasoners' is zeer belangrijk in de praktijk en verschillende technieken worden aangeleerd om deze schaalbaarheid te verhogen. Naast 'deterministische reasoning' wordt ook 'probabilistische reasoning' behandeld, wat zeer nuttig is als er onzekerheid in de kennisbanken in rekening gebracht wordt.

Recent wint ook 'stream reasoning' meer aan belang. Dit is een techniek om efficiënte 'reasoning' te implementeren voor gigantische stromen informatie, zoals bijvoorbeeld in 'Internet-of-Things'-omgevingen. De onderliggende technieken van 'stream reasoning' worden in detail behandeld. Verder worden er tal van voorbeelden uit de praktijk gegeven en worden er ook live demo's getoond.

Lesgevers: Filip De Turck en Femke Ongenae

Datum: 23 april 2015

Deep learning

Kunstmatige neurale netwerken zijn in staat om het menselijk leerproces na te bootsen door het veranderen van de sterkte van gesimuleerde neurale verbindingen. Deze eigenschap zorgt ervoor dat kunstmatige neurale netwerken uiterst effectief zijn in het automatisch terugvinden van patronen in grote hoeveelheden data (deep learning). Dit heeft onlangs geleid tot een aantal doorbraken op het vlak van taalverwerking en audiovisuele analyse.

Bijzondere aandacht zal besteed worden aan het gebruik van meerlagige neurale netwerkarchitecturen en de technieken die de inzet van deze netwerkarchitecturen praktisch haalbaar hebben gemaakt. Vervolgens wordt er stilgestaan bij verschillende toepassingen op het vlak van taalverwerking en audiovisuele analyse en wordt er een overzicht gegeven van toekomstige uitdagingen op het vlak van onderzoek en ontwikkeling in het domein van deep learning.

Lesgever: Wesley De Neve

Datum: 30 april 2015

Big Data Analytics

Volgende onderdelen worden behandeld:

- Inleiding tot 'Analytics' ('Descriptive', 'predictive' en 'prescriptive analytics').
- De Spark (Streaming) software stack (Berkeley Data Analytics Stack (BDAS)).
- Gebruik van MLlib (machine learning library) binnen Spark voor 'Analytics'.
- Bespreking van enkele gebruikscases van de Spark software stack voor 'Analytics' (recommender systemen).
- Hands-on demo van de UGent-implementatie van de Berkeley open source software stack.

Lesgever: Dirk Van den Poel

Datum: 7 mei 2015

Getuigenissen uit de praktijk, deel II

De eerste spreker is Vivek Bajaj, Director, Global Financial Services - Big Data & Analytics bij IBM. Hij zal de IBM oplossing voor 'Big Data Analytics' toelichten en illustreren.

In de tweede lezing zal Alain Houf, sales engineer bij Intersystems, de NoSQL oplossing van Intersystems toelichten en onder andere illustreren aan de hand van een case-study over de Gaia-missie van de Europese Ruimtevaartorganisatie ESA die de melkweg nader in kaart moet brengen.

De derde spreker is Inge Van Nieuwerburgh, medewerkster van de Gentse Universiteitsbibliotheek. Ze gaat dieper in op recente trends in digitale bibliotheken en het 'open archives' initiatief.

Datum: 21 mei 2015

3. Visualisatie

Een uitgelezen manier om mensen te helpen om Big Data te exploreren en te begrijpen, is het visualiseren van de data: we zijn immers vaak erg goed in staat om patronen, tendensen, uitschieters, ... te begrijpen met behulp van visualisaties. Mede aan de hand van een groot aantal concrete voorbeelden wordt uiteengezet hoe een goede interactieve visualisatie kan worden opgebouwd, wat de typische misvattingen zijn, hoe visualisaties kunnen misbruikt worden, ... Er wordt ook een overzicht gepresenteerd van een aantal typische technieken en hulpmiddelen voor interactieve informatie-visualisatie.

Lesgevers: Erik Duval en Joris Klerkx

Data: 28 mei 2015 en 4 juni 2015

MEER INFO & INSCHRIJVEN
www.ivpv.ugent.be/bigdata

Prijs

Deelnameprijs omvat lesgeld, hand-outs, frisdranken, koffie en broodjes.

Betaling geschiedt na ontvangst van de factuur.

Alle facturen zijn contant betaalbaar dertig dagen na dagtekening.

Alle vermelde bedragen zijn vrij van BTW.

Voor iedere module kan er afzonderlijk ingeschreven worden.

Module 1: Gegevensbeheer	€ 750
Module 2: Gegevensanalyse	€ 750
Module 3: Visualisatie	€ 300
Volledige opleiding	€ 1.440

Korting

- ▶ Indien minstens één deelnemer van een bedrijf inschrijft voor de volledige opleiding wordt voor alle bijkomende gelijktijdige inschrijvingen van hetzelfde bedrijf, per module of volledig pakket, een korting van 20% verleend. Facturatie geschiedt dan d.m.v. een gezamenlijke factuur.
- ▶ 15% korting op de in de tabel vermelde prijzen voor Agoria leden.
- ▶ 10% korting op de in de tabel vermelde prijzen voor leden van AIG (Alumnivereniging van Ingenieurs afgestudeerd aan de Universiteit Gent) en van VBIG (Verbond Afgestudeerde Bio-ingenieurs van de Universiteit Gent).
- ▶ Aangepaste prijzen voor personeel van UGent en geassocieerde hogescholen.
- ▶ Kortingen zijn niet cumuleerbaar.

KMO-portefeuille

Universiteit Gent aanvaardt betalingen via de KMO-portefeuille (www.kmo-portefeuille.be; gebruik autorisatiecode DV.0103194).

Annulering

Raadpleeg onze annulatievoorwaarden op www.ivpv.ugent.be/annulatievoorwaarden

Tijdstip en locatie

- ▶ De lessen worden gegeven van **18u tot 21u30**, in 2 delen, gescheiden door een broodjesmaaltijd en vinden plaats aan de **Universiteit Gent, Instituut voor Permanente Vorming, IVPV leszaal, Technologiepark 904, 9052 Zwijnaarde**.
- ▶ Data onder voorbehoud van wijzigingen om onvoorziene omstandigheden.

MEER INFO & INSCHRIJVEN

www.ivpv.ugent.be/bigdata



ORGANISATIE

Universiteit Gent
 Instituut voor Permanente Vorming (IVPV)
 Technologiepark 904, 9052 Zwijnaarde
 Tel: +32 9 264 55 82, Fax: +32 9 264 56 05
 E-mail: ivpv@UGent.be