

# Organisatie Bachelorreferaat (BIT: 192399779; TI: 192199259)

Versie januari 2014

## 1 Inleiding en leerdoelen

Wat zijn de ontwikkelingen in het vakgebied? Wat is de bijdrage van onderzoekers aan deze ontwikkelingen en hoe doe je onderzoek?

Karakteristiek voor een wetenschappelijke opleiding is de relatie tussen onderwijs en onderzoek. In het studieonderdeel Bachelorreferaat van de opleidingen Bedrijfsinformatietechnologie en Technische Informatica is deze relatie expliciet aanwezig. Dit studieonderdeel richt zich op een actieve kennismaking met wetenschappelijk onderzoek en natuurlijk ook op uitdagende ontwikkelingen in het vakgebied.

Ieder semester worden thema's uit het vakgebied en uit het onderzoek van leerstoelen, of leerstoelclusters, aangeboden voor onderzoek. Deelnemende studenten schrijven een voorstel voor het uitvoeren van onderzoek. Als dat voorstel wordt geaccepteerd, voert de student het onderzoek uit en schrijft daar een paper over dat in overeenstemming is met de eisen van wetenschappelijk schrijven. Als die paper wordt geaccepteerd, wordt deze gepresenteerd tijdens de Twente Student Conference on IT en gepubliceerd op de site van die conferentie.

Informatie over de aangeboden thema's (tracks) en resultaten (papers) van eerdere keren dat het vak heeft gedraaid zijn te vinden via <http://referaat.cs.utwente.nl>

Het vak heeft de volgende leerdoelen:

- Onderzoek van beperkte omvang uitvoeren en daar een paper over schrijven.  
Dit leerdoel heeft subdoelen:
  - o Wetenschappelijke literatuur zoeken, lezen en beoordelen op kwaliteit en relevantie voor het eigen uit te voeren onderzoek.
  - o Een beargumenteerde keuze maken voor een onderzoeksmethode.
  - o Een research proposal schrijven volgens een gegeven template.
  - o Een paper volgens wetenschappelijke maatstaven schrijven en daarbij gebruik te maken van een gegeven template.
- Een review schrijven van research proposals van collega-studenten.
- Een review schrijven van papers van collega-studenten.
- Een presentatie over het uitgevoerde onderzoek geven.

Verder draagt deelname aan het vak bij aan:

- Het ontwikkelen van een onderzoekende houding.
- Het ontwikkelen van kritische vermogens en instelling tegenover eigen en andermans werk.

## 2 Organisatie

Voorafgaand aan het semester, kiest elke student een track uit de aangeboden tracks. Elke track heeft een trackcoördinator. Daarnaast kunnen nog andere docenten/onderzoekers aan de begeleiding binnen een track deelnemen; zie Blackboard voor de tracks en de bemensing ervan. Op Blackboard wordt een planning van de activiteiten in het semester gepubliceerd (zie 'Activiteitschema').

Het Bachelorreferaat wordt individueel uitgevoerd. Het is mogelijk dat verschillende studenten aan hetzelfde onderwerp werken, maar ook in dat geval heeft elke student een eigen paper en een eigen presentatie.

De voorkenniseisen zijn:

BIT: P + 60 EC's

TI: P + 60 EC's (exclusief minorvakken), Technical writing

De controledatum is enkele weken voor de start van het semester.

De coördinatoren zijn Herman Koppelman (kamer: Zilverling 2033; [h.koppelman@utwente.nl](mailto:h.koppelman@utwente.nl)) en Céline Heijnen (Ravelijn 3260, [c.m.m.heijnen@utwente.nl](mailto:c.m.m.heijnen@utwente.nl)).

### 3 Tracks

De activiteiten van studenten vinden plaats binnen de groep die gekozen heeft voor dezelfde track. Deze groep functioneert als onderzoeksgroep en bespreekt zowel literatuur als het proces en de resultaten van individuele onderzoekactiviteiten. Elke student heeft een vaste begeleider vanuit de track. Elke track houdt regelmatig groepsbijeenkomsten (in principe elke maandagmiddag), waaraan alle studenten van een track deelnemen en ook de begeleider(s). In deze bijeenkomsten worden de activiteiten en producten van de studenten besproken.

### 4 In te leveren producten en onderwijsmateriaal

In deze paragraaf wordt besproken welke producten ingeleverd moeten worden. Naast de genoemde producten kunnen in tracks afspraken worden gemaakt voor het inleveren van aanvullende producten.

Het belangrijkste onderwijsmateriaal is literatuur die studenten zelf zoeken. Die literatuur verschilt uiteraard per student. Naast deze literatuur is er ondersteunend materiaal, dat structuur kan geven aan het schrijven van proposal, paper, review en presentatie. In de volgende paragrafen worden de in te leveren producten besproken, en wordt per product gewezen op onderwijsmateriaal dat ter beschikking is.

#### Ondersteunend materiaal

- *Seven steps to graduation*, Bijlage G, een overzicht van het gehele proces, van Mariëlle Stoelinga.

#### 4.1 Proposal

In het eerste blok wordt het thema en de bijbehorende literatuur verkend. Dit resulteert in een voorstel ('proposal') voor het uit te voeren onderzoek. Dit voorstel wordt geschreven volgens een vaste Word of LaTeX template, die te vinden is op Blackboard. De omvang van een proposal is in principe ongeveer 3 pagina's volgens de gegeven template.

Voorafgaand aan de definitieve proposal wordt er een concept proposal geschreven die wordt gepeerreviewed (zie volgend punt).

De proposal en later ook de paper worden bij voorkeur geschreven in het Engels. Het doel is de studenten enige ervaring te laten opdoen in het gebruik van Engels in een wetenschappelijke context. Indien de begeleider het Nederlands niet in voldoende mate beheerst, is het gebruik van Engels verplicht.

#### Ondersteunend materiaal

- *Efficient reading of papers*, van Michael J. Hanson  
<http://www.cs.columbia.edu/~hgs/netbib/efficientReading.pdf>  
In veel gevallen is er een overvloed aan papers beschikbaar. Het is dan geen doen om alle papers te lezen. Je moet selecteren welke papers je wel of niet serieus wilt bekijken. Het is dan handig als je snel een indruk kunt krijgen van een paper, zonder die van a tot z te hoeven lezen. Dit document geeft daar aanwijzingen voor.
- *Onderzoeksvoorstel: opzet en review*: Bijlage A.  
Geeft de structuur van een proposal.
- *Word/LaTeX template*: zie Blackboard.  
Template die gebruikt dient te worden voor proposal en paper.
- Voorbeeld van een goed proposal: Bijlage F

#### 4.2 Peer review proposal

Elke student schrijft een review van een of meer proposals van medestudenten. Het nut hiervan is dat de peer reviewer leert kritisch te kijken naar het werk van anderen. Daarnaast kan de gereviewde student de reviews gebruiken om de proposal te verbeteren.

#### Ondersteunend materiaal

- *Onderzoeksvoorstel: opzet en review*: Bijlage A.  
Geeft aanwijzingen voor het reviewen van een proposal

### 4.3 Paper

In het tweede blok voeren studenten zelfstandig een beknopt onderzoek uit dat resulteert in een conferentiepapier. Voor de paper wordt dezelfde Word of LaTeX template gebruikt als voor de proposal. De omvang van een paper is idealiter 8 pagina's. Papers zijn **maximaal 10 pagina's** (inclusief bijlages); papers die langer zijn worden niet in de proceedings opgenomen. Er geldt geen minimum voor het aantal pagina's van een paper. De begeleider bepaalt of de paper lang genoeg is om te worden geaccepteerd.

Voordat de definitieve paper wordt opgeleverd, zullen er diverse drafts zijn geschreven en besproken in de track. Een van die drafts wordt gepeerreviewed (zie volgend punt).

De paper wordt bij voorkeur in het Engels geschreven.

Voor het inleveren van de definitieve paper geldt een harde deadline. De tijd tussen deze deadline en de conferentie is nodig voor de organisatie van de conferentie.

#### Ondersteunend materiaal

- *How to write a paper*: slides gelijknamig hoorcollege, zie Blackboard.
- *Structure of a paper*: zie bijlage B.
- *Checklist paper*: bijlage E.
- *Paper requirements*, zie site conferentie, <http://referaat.cs.utwente.nl/page/5/paper-requirements/>  
Geeft aanwijzingen voor het inleveren van camera ready paper.
- *English Styleguide*,  
[http://www.utwente.nl/mc/en/service/websites\\_socialmedia/standards-agreements/engelse-taal/English\\_Style\\_Guide/](http://www.utwente.nl/mc/en/service/websites_socialmedia/standards-agreements/engelse-taal/English_Style_Guide/)  
Enige aanwijzingen voor schrijven in het Engels.

### 4.4 Peer review paper

Elke student schrijft een peer review van een of meer proposals van medestudenten. Ook de peer review van de paper heeft een tweeledig doel. De peer reviewer leert kritisch te kijken naar het werk van anderen en de gereviewde student kan de reviews gebruiken om de paper te verbeteren.

#### Ondersteunend materiaal

- *Checklist peer review paper*: Bijlage C.

### 4.5 Presentatie

De slides van de presentatie dienen in het Engels te worden gesteld. De presentatie zelf is bij voorkeur in het Engels. Het doel is weer studenten te laten oefenen met het presenteren in het Engels. Het is toegestaan, na overleg met de begeleider, de presentatie in het Nederlands te geven, tenzij de begeleider het Nederlands niet in voldoende mate beheerst. Eventueel kan voor het maken van de presentatie gebruik worden gemaakt van een van de presentatiesjablonen van de UT: zie <http://www.utwente.nl/huisstijl/downloads>

## 5 Hoorcolleges

Er wordt een aantal hoorcolleges aangeboden ter voorbereiding op belangrijke activiteiten.

- Een inleidend hoorcollege, waarin de organisatie wordt besproken.
- Een hoorcollege van een informatiespecialist over het zoeken en verwerken van informatie. Belangrijke onderwerpen zijn: welke informatiebronnen zijn er, welke zoeksystemen zijn er, hoe zoek je systematisch en efficiënt relevante literatuur, hoe beoordeel je de kwaliteit van gevonden literatuur, hoe refereer je aan literatuur.
- Twee hoorcolleges over het opzetten van onderzoek, met onderwerpen als: welke soorten onderzoek zijn er, wat zijn goede onderzoeksvragen, welke onderzoeksmethodes zijn er.
- Een hoorcollege over het schrijven van een paper, met onderwerpen als: welke soorten papers zijn er, uit welke onderdelen bestaat een paper en wat moet de inhoud daarvan zijn, hoe formuleer je onderzoeksvragen, hoe kun je de kans vergroten dat een paper wordt geaccepteerd.
- Een hoorcollege over het geven van een presentatie.

De slides van de (meeste) hoorcolleges zijn beschikbaar via Blackboard.

## 6 Beoordeling

De belangrijkste producten die worden beoordeeld zijn de proposal en de paper.

Voor de proposal wordt een go/no go besluit genomen. 'No go' betekent: stoppen met het onderzoek en in een volgend semester opnieuw beginnen, met een ander onderzoek. Een proposal dat tegen een 'go' aanzit, maar dat net niet krijgt, komt in aanmerking voor een snelle 'herkansing': de betreffende student krijgt ongeveer 1 week de tijd om de proposal te verbeteren.

Een paper wordt al dan niet geaccepteerd voor de conferentie. Acceptatie betekent dat het eindcijfer hoger dan 5 is en dat de paper wordt gepresenteerd en op de site van de conferentie komt te staan. Als een paper niet wordt geaccepteerd, moet de betreffende student in een volgend semester opnieuw starten. In principe kiest de student dan een ander onderwerp, maar als de begeleider daarmee instemt, kan worden afgesproken dat een student verder gaat met hetzelfde onderzoek.

Een paper kan ook voorwaardelijk worden geaccepteerd, als de kwaliteit net niet voldoende is. De student krijgt dan de mogelijkheid de paper te verbeteren. Als dat in voldoende mate lukt, wordt de paper alsnog geaccepteerd voor de conferentie. Het eindcijfer is in dat geval 6.

Het eindcijfer is voornamelijk gebaseerd op de kwaliteit van de ingeleverde paper. De details van de beoordeling zijn te vinden in bijlage D, *Beoordeling Bachelorreferaat en procedure acceptatie/afwijzing papers*.

## 7 Conferentie

Details over de conferentie zijn te vinden op de site <http://referaat.cs.utwente.nl>. Een paar belangrijke zaken zijn:

- Deelnemers moeten zich vooraf inschrijven voor de conferentie, via de website van de conferentie.
- Deelnemende studenten kunnen gasten (familie, vrienden) uitnodigen voor de conferentie. Deelname is gratis.
- Er zijn UT-templates die gebruikt kunnen worden; zie <http://www.utwente.nl/huisstijl/downloads/>
- Studenten worden aangemoedigd om in het Engels te presenteren. De conferentie is een goede gelegenheid om te oefenen met presenteren in het Engels. Het is niet verplicht om in het Engels te presenteren, tenzij de begeleider niet Nederlandstalig is. Voor de slides van de presentatie is het gebruik van Engels wel verplicht.
- Elke track kan een 'best paper' aanwijzen. Best papers worden plenair gepresenteerd. Auteurs van best papers ontvangen een Best Paper Award.
- De duur van de presentaties van de best papers is 15 minuten, gevolgd door 5 minuten discussie. De overige papers worden gepresenteerd in parallelsessies. De duur van deze presentaties is in totaal 15 minuten: 10 minuten voor de spreker, gevolgd door 5 minuten discussie.

## Bijlage A: Onderzoeksvoorstel: opzet en review

Het onderzoeksvoorstel is (naast je paper en je presentatie) een van de belangrijke deliverables van het BSc referaat. Je hebt een kwartiel de tijd voor. Het onderzoeksvoorstel is een plan van aanpak waarin je beschrijft, (a) wat je gaat onderzoeken, (b) waarom je dit gaat onderzoeken, (c) hoe je dit gaat doen en (d) wat er al gebeurd is op dit gebied. Daarnaast bevat je onderzoeksvoorstel een (e) planning en (f) een literatuurlijst. Echter, deze ingrediënten worden niet in deze volgorde opgeschreven.

De structuur van het onderzoeksvoorstel is als volgt:

1. Titel, auteurs
2. Abstract (Samenvatting)
3. Introductie
4. Probleemstelling/ vraagstelling
  - aka Onderzoeksdooel en onderzoeksvragen
5. Achtergrond
  - voorkennis
6. Bestaand onderzoek (related work)
7. Onderzoeksmethode
8. Beoogde resultaten/ product
9. Planning
10. Literatuurlijst (references)
11. (Optioneel) Appendices

Enkele opmerkingen:

- Er bestaan variaties op deze structuur: soms is er achtergrondkennis nodig om de onderzoeksvragen te kunnen begrijpen, of moet je veel bestaand onderzoek bespreken voordat je je probleemstelling kunt uitleggen. In dat geval verandert de volgorde van de hoofdstukken 4,5 en 6. Soms komt related work aan het eind, om een goede vergelijking te kunnen maken met de resultaten van het paper.
- Het onderzoeksplan wordt meestal niet in deze volgorde geschreven. Vaak begin je met de onderzoeksvragen, dan onderzoeksmethoden en dan gerelateerd werk. Daarna schrijf je de introductie, conclusie en abstract. De literatuurlijst vul je zodra zich een artikel aandient. Definitieve titel wordt ook vaak laat bepaald, daarvoor gebruik je met werktitels.

## De componenten van een onderzoeksvoorstel

### 1. Titel

- moet de lading dekken
- is “catchy”

### 2. Abstract. Bevat 4 ingrediënten

- **onderwerp:** welk probleem wordt opgelost?
- **motivatie:** waarom is het belangrijk dat dit probleem wordt opgelost?
- **methode:** hoe wordt het probleem opgelost?
- **resultaten:** wat zijn de (beoogde) resultaten?

Of, anders geformuleerd:

- **situatie:** wat is de bestaande situatie?
- **probleem:** waarom is de bestaande situatie onbevredigend?
- **motivatie:** waarom is het belangrijk dat de bestaande situatie wordt opgelost?
- **oplossing:** hoe (met welke methode) gaan we de situatie oplossen?
- **resultaten:** wat zijn de (beoogde) resultaten?

De motivatie is een externe motivatie, i.e. je moet aangeven waarom het onderzoek interessant is voor anderen. Een interne motivatie (‘het vak halen’) is niet relevant voor de lezer.

### 3. Introductie

- De introductie bevat eigenlijk dezelfde componenten als het abstract, maar nu veel meer uitgewerkt.
- je introduceert *alle* belangrijke concepten, methoden, resultaten die je gebruikt in je paper: het mag niet zo zijn dat er een belangrijk feit in je paper staat, dat niet eerst in de introductie is geïntroduceerd.
- geeft een vooruitblik op vraag/ probleemstelling, (te gebruiken) onderzoeksmethoden (beoogde) resultaten.

**4. Onderzoeksdoel en -vragen.** Het onderzoeksdoel geeft aan welke probleem je met je onderzoek wil oplossen; de onderzoeksvragen zijn de vragen die je moet beantwoorden om de onderzoeksdoel te bereiken. Net als andere projectdoelen, dienen onderzoeksdoelen (gecombineerd met onderzoeksvragen) **SMART** te zijn, dus

- **specific:** zo concreet mogelijk
- **measurable:** kwantificeerbaar; je kunt achteraf zeggen of doelen gehaald zijn
- **attainable:** dit wordt o.a. gewaarborgd door de onderzoeksmethode
- **relevant:** lost belangrijk of interessant probleem op, innovatief
- **time-bound:** vandaar de planning in je proposal

Voor een manier om aan goede vragen te komen, zie de alinea over *Design Science*.

**5. Achtergrond.** Deze sectie beschrijft wat moet de lezer weten op de rest van het paper de kunnen begrijpen? Hierin introduceer je standaardterminologie, -notatie, etc. Dit hoofdstuk hoeft niet origineel te zijn, want dit is niet de bijdrage van jouw onderzoek. Met een goede bronverwijzing mag je standaard definities daarom zonder meer overnemen.

**6. Bestaand onderzoek ('related work' in het Engels):** geeft aan (1) wat er al bekend is over jouw probleem en (2) wat jouw aanpak toevoegt aan bestaande resultaten. Tijdens je voorstel begin je met een initiële verkenning van bestaande literatuur, maar je vult deze sectie aan zodra er zich relevant werk aandient.

#### 7. Onderzoeksmethode.

- **effectief:** lost de onderzoeksvragen inderdaad op
- **transferable:** je kunt jouw methode aan een medestudent geven, de resultaten zijn dan gelijkwaardig aan de resultaten die je zelf zou behalen
- **haalbaar:** dit geef je aan door voorbeelden te beschrijven.
- **validerend:** je onderzoeksmethode kwantificeert in hoeverre jouw onderzoeksresultaten een oplossing zijn voor het gestelde probleem.

**8. (Beoogde) Resultaten/product.** In het onderzoeksvoorstel beschrijf je beoogde resultaten: wat heb ik in mijn handen als het onderzoek gedaan is?

- **controleerbaar:** een *peer* (iemand met dezelfde achtergrond als jij) kan beoordelen of de of jouw resultaten correct zijn.
- **herhaalbaar:** Voor experimenteel onderzoek impliceert controleerbaarheid ook herhaalbaarheid: Hiervoor is het belangrijk dat producten die jij maakt tijdens je onderzoek (modellen, code, benchmarks etc) zo veel mogelijk openbaar zijn, en de set up (computer settings, etc).

**Verder geef je zo veel mogelijk voorbeelden!**

#### 9. Planning

Een planning geeft houvast bij het uitvoeren van je onderzoek: wat doe je wanneer. Daarnaast toont het aan dat jouw onderzoek haalbaar is in de beschikbare tijd. Ik adviseer een tabel met 4 kolommen.

weeknummer	activiteit	deliverable	deadline
1-2	modelleren	Uppaal model van xyz protocol	17 februari
2-3	programmeren	component A van tool B	17 februari

Tip: natuurlijk ruim je in je planning tijd voor het schrijven van je paper. Het is echter belangrijk dat je vanaf dag 1 begint met schrijven: documenteer zo veel mogelijk wat je doet, en houd je resultaten op een systematische manier bij.

**10. Referenties.** Verwijs hierbij naar bestaande bronnen uit te literatuur.

- Zorg dat je referenties, volledig, up-to-date, en betrouwbaar zijn.
- Hou je aan het voorgeschreven format (zie template op Blackboard)

**Zie je onderzoeksvoorstel als een businessplan.** Een goede strategie om je onderzoeksvoorstel te schrijven is om dit te zien als businessplan: stel je wilt geld lenen bij een bank om jouw onderzoek te financieren. Op welke gronden zal een bank dat doen? De bank wil onderzoek financieren dat (a) zinvol en (b) succesvol is. Ad (a): Je zult de geldschieters dus moeten overtuigen dat jouw onderzoek zinvol is, m.a.w. je moet met een goede motivatie komen waarom jouw onderzoek de moeite waard is. Een van de zaken die je zult moeten beargumenteren is dat jouw onderzoek niet al eerder gedaan is. M.a.w. je moet laten zien wat er op jouw gebied al gedaan is, en hoe jouw onderzoek daarop aansluit, m.a.w. wat zijn de overeenkomsten, verschillen met bestaand onderzoek. Ad (b) Banken willen risico's uitsluiten. Ze willen graag weten dat je waar kunt maken wat jij in jouw voorstel belooft. Jij kunt de bank overtuigen door met een goed doortimmerd plan te komen, d.w.z. een goede onderzoeksmethode die jouw probleem gaat oplossen. Dit doe je door zo concreet mogelijk te beschrijven wat je gaat doen: welke tools ga je gebruiken, welke case studies ga je analyseren.

**Transferabiliteit.** Eigenlijk moet het zo zijn dat, als je jouw voorstel aan een medestudent geeft, jouw medestudent het voorstel goed kan uitvoeren, en de beloofde resultaten boekt.

## Review van onderzoeksvoorstel

Om een onderzoeksvoorstel te reviewen, loop je alle componenten langs:

- Titel: dekt die de lading en is die 'catchy'?
- Abstract: bevat die de genoemde ingrediënten?
- Introductie: voldoet die aan de genoemde punten?
- Onderzoeksdelen en –vragen: zijn die SMART geformuleerd?
- Achtergrond: voldoet die?
- Bestaand onderzoek: adequaat?
- Onderzoeksmethode: voldoet die aan de gestelde punten?
- Beoogde resultaten/ product: voldoet die aan de gestelde punten?
- Planning: met activiteiten, deliverables en deadlines?
- Referenties: volledig? Vormgeving als voorgeschreven?

## **Bijlage B**

# **Structure of a paper**

Version 1, November 2010

No unambiguous recipe for writing a paper exists. But it is possible to provide some support for novices in this area. This document gives a possible structure of a scientific paper and characterizes the different components.

A paper generally consists of:

- an abstract
- an introduction
- a “something in between”
- conclusions
- references.

These components are discussed in the next sections.

## **1. Abstract**

An abstract is a shortened version of the paper and should contain all information necessary for the reader to determine:

- what the objectives of the study were
- how the study was done
- what results were obtained
- the contribution of the paper.

It should be possible to determine the major points of a paper by reading the abstract. Many readers of a scientific paper will only read the abstract to determine whether the paper is interesting to them. For this reason, this section should be written carefully and succinctly to have the greatest impact in as few words as possible.

## **2. Introduction**

The introduction consists of different parts.

First of all it introduces the context of the work and gives the motivation for research in this area. Meant is the ‘external’ motivation: what is the relevance of this subject. The ‘internal’ motivation, why you did choose this subject, is less relevant for the reader.

The introduction also describes the problem this paper focuses on. This leads to a number of research questions (usually 3 to 6). The research questions should be operational: they can be answered by the method you are using. It should be made clear what this paper contributes to the topic.

The approach or method should also be described: explain how you will answer the research questions. A description like ‘literature study’ is not precise and concrete enough; in case of a literature study describe which literature and how the literature can answer the questions.

At the end of the introduction the paper organization is given: a very short ‘road map’ to the rest of the paper.

## **3. ‘Something in between’**

The ‘in between’ depends heavily on the kind of paper. We will discuss some topics that in many cases will be useful.

### **3.1 Overview of literature**

A possible section ‘Literature’ discusses relevant results and conclusions of previously published studies. This section should be designed to lead to the paper’s research questions, rather than to show that the author has read a lot of literature. Discuss only findings that are relevant to your research questions. Usually the structure of this section is not a sequence of reviewed papers; the structure should rather reflect the research questions.



### **3.2 Method(s)/Approach**

Such a section gives a detailed description of what was actually done. It describes the methods which have been used and a motivation for using these methods.

### **3.3 Results**

If something has been measured, in this section the results are reported. Report only what you found; do not attempt to evaluate or explain the results in this section. This should be done in the section 'Discussion'.

### **3.4 Discussion**

This section has a relation to the section Results. The Results section reports the results, the discussion section explains and evaluates the results. How do the results compare to the expected results? How do they compare to literature? Can you explain possible differences?

You should interpret your results in light of other published results. You can refer to sources already mentioned earlier, but also new sources can be introduced. Make sure you provide accurate citations.

This section is also the place to identify weak points of your research.

## **4. Conclusions**

This is the place to return to the research questions. This section describes usually a general conclusion and also the answer to the different research questions as posed in the introduction. The contribution of the work can be repeated. Limitations of your work should also be mentioned in this section. You may also suggest future work.

## **5. References**

All references in the paper should be included in this section. Do not include references you did not use in the paper.

## Bijlage C

# Checklist peer review paper

Version November 2010

The first part of the checklist (points 1-5) deals with the structure of the paper and reflects the structure as mentioned in the document Writing a paper. The last part (points 6-8) deals with additional topics.

## 1. Abstract

Does the abstract describe

- the objectives
- the research method
- the results
- the contribution of the paper (what does the paper contribute to existing knowledge)?

## 2. Introduction

Does the introduction describe

- the motivation for the research
- the problem definition, the research questions, and the research method
- the context of the research (references to literature)
- the contribution of the paper to the research field
- the 'road map' to the rest of the paper?

*NB Some of these topics can also appear in another section, depending upon the structure of the paper.*

## 3. 'In between'

The part between the Introduction and Conclusions depends heavily on the kind of paper. A global review question is:

- is the structure of the sections of this part clear? Do sections follow logically from one another?

Usually one or more of the following sections can be found.

### 3.1 Literature

- Is this section focused at the research questions?

### 3.2 Methods/approach

- Is this section detailed enough that somebody else could repeat the study?
- Does it give a motivation for the chosen method/approach?

### 3.3 Results/Discussion

- Is there a clear separation between the results (what has been measured) and the interpretation and evaluation of the results?

## 4. Conclusions

Does the section Conclusions

- answer the research questions as posed in the introduction
- describe future work (optional)?

## 5. References

- Do the references have the correct format (check template)

- Do all referenced papers appear in this section?

## **6. Reasoning**

- Is the reasoning clear and logically sound?
- Are claims justified by evidence? Does the reasoning irrefutably lead to the final conclusions?

## **7. Writing style and format**

- Does the writer succeed in making the reader interested?
- Are the titles (first of all the paper title) clear and do they cover the content?
- Are sentences clear and grammatically correct?
- What about the carefulness (numbering and captioning of tables and figures, misspellings, lay out)?
- Is the used format in agreement with the template?

## **8. General**

- Mention one or two strong points and one or two weak points of the paper. (No new points; these points should follow from the foregoing)
- Give at least one recommendation for the writer to improve the paper.

## **Bijlage D**

### **Procedure acceptatie/afwijzing papers en beoordeling Bachelorreferaat.**

Versie januari 2014

#### **Accepteren/afwijzen papers**

Als een paper geaccepteerd wordt voor de conferentie, wordt deze op de conferentiesite geplaatst. Het eindcijfer is in dat geval minimaal een 6. Een paper kan voorwaardelijk worden geaccepteerd, als de kwaliteit net niet voldoende is. De student krijgt dan ongeveer 1 week de gelegenheid de paper te verbeteren. Als dat in voldoende mate lukt, wordt de paper alsnog geaccepteerd voor de conferentie. Het eindcijfer is in dat geval 6. Als een paper niet wordt geaccepteerd, moet de betreffende student in een volgend semester opnieuw starten. In principe kiest de student dan een ander onderwerp, maar als de begeleider daarmee instemt, kan worden afgesproken dat een student verder gaat met hetzelfde onderzoek.

#### **Beoordeling Bachelorreferaat**

Bij de beoordeling wordt naar de volgende aspecten gekeken:

- Kwaliteit van paper en peer reviews
- Wetenschappelijke kwaliteit van het werk
- Proces
- Mondelinge presentatie

Op het hierna volgende formulier worden deze aspecten uitgewerkt.

De begeleider bepaalt op basis van de criteria zoals genoemd in het formulier een eindcijfer, met een motivatie.

**Student Name:**

**Programme:**

**Titel of paper:**

**Student number:**

**Date:**

Assessment criteria	- -	-	+/-	+	++	Remark (optional)
<b>Quality of Paper and Peer reviews</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<b>Content Paper</b> <ul style="list-style-type: none"> <li>• Problem statement</li> <li>• Research questions</li> <li>• Context and existing work</li> <li>• Research method (selection, motivation, application)</li> <li>• Discussion and argumentation</li> <li>• Results and conclusions</li> <li>• Recommendations</li> </ul>						
<b>Structure Paper</b> <ul style="list-style-type: none"> <li>• Organization and structure</li> <li>• Clarity</li> <li>• References</li> <li>• Language</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<b>Peer reviews</b> <ul style="list-style-type: none"> <li>• Structure</li> <li>• Argumentation</li> <li>• Relevance</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<b>Scientific quality of work</b> <ul style="list-style-type: none"> <li>• Relevance</li> <li>• Originality</li> <li>• Complexity/volume of work</li> <li>• Contribution</li> <li>• Validation of design/product</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<b>Process</b> <ul style="list-style-type: none"> <li>• Independence</li> <li>• Creativity</li> <li>• Problem solving</li> <li>• Planning</li> <li>• Critical attitude</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<b>Oral presentation and defense</b> <ul style="list-style-type: none"> <li>• Clarity</li> <li>• Use of media</li> <li>• Answering questions</li> </ul>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
<b>FINAL MARK:</b>						

**Motivation for final mark:**

Supervisor:

Signature:

## **Inhoud**

### **TITEL**

Is beschrijvend

Is catchy

### **ABSTRACT**

beschrijft duidelijk wat het onderwerp is

beschrijft duidelijk wat opgeloste probleem is

maakt duidelijk waarom onderwerp / probleem relevant is

beschrijft duidelijk de gebruikte methoden

beschrijft zo concreet mogelijk wat de behaalde resultaten zijn

geeft aan wat nieuw is

### **INTRODUCTIE**

beschrijft (in meer detail) wat het onderwerp is

beschrijft duidelijk wat opgeloste probleem is

maakt duidelijk waarom onderwerp / probleem relevant is

beschrijft duidelijk de gebruikte methoden

beargumenteert waarom deze methoden (taal, technieken, tools) gebruikt zijn

beschrijft duidelijk wat de resultaten zijn

vergelijkt resultaten met (state-of-the-art) literatuur

geeft duidelijk aan wat nieuw is

beschrijft related work (of verwijst naar elders in paper)

is volledig: alle belangrijke info uit paper komt terug in intro

### **RELATED WORK**

(als onderdeel introductie of elders in paper)

beschrijft verschil huidig paper en bestaande literatuur

is up-to-date (zo recent mogelijk)

is volledig (geen belangrijk werk gemist)

### **PROBLEEM- en VRAAGSTELLING (in aparte sectie of onderdeel intro)**

het is duidelijk wat de probleemstelling uit het paper is

probleemstelling is specifiek

probleemstelling is measurable / kwantitatief volgens duidelijke criteria

probleemstelling is haalbaar (proposal)

probleemstelling is relevant

probleemstelling is time bound / urgent

### **ONDERZOEKSMETHODE(N)**

is effectief (lost onderzoeksvragen inderdaad op)

is transferable (een medestudent behaalt zelfde resultaten als auteur)

zijn volgens de wetenschappelijke methoden

### **BODY (resultaten)**

zijn verkregen volgens onderzoeksmethoden

zijn controleerbaar

zijn herhaalbaar

### **CONCLUSIONS**

beantwoorden beantwoorden onderzoeksvragen

geven richtingen aan voor nieuw onderzoek

## **Specifieke onderdelen**

### **REFERENTIES**

zijn consistent (steeds zelfde formaat)

bij webpagina's wordt jaartal aangegeven

verwijzen naar betrouwbare bronnen, met wetenschappelijke status

### **TOOLS en TOOLCOMPONENTEN**

tool structuur wordt geschreven met plaatje: component in box, data op pijlen

van ieder tool (component) wordt beschreven: input, output, methoden

het is duidelijk wat door auteur geprogrammeerd is, en wat van elders is

### **ALGORITHMEN, MODELLEN**

worden uitgelegd mbv regelnummers

van iedere regel is duidelijk wat het is

### **PLAATJES, DIAGRAMMEN, TRANSITIE-SYSTEMEN**

ieder plaatje wordt uitgelegd

Van ieder onderdeel in plaatje is duidelijk wat het voorstelt

### **FIGUREN en TABELLEN**

alle figuren zijn leesbaar en voorzien van duidelijke captions

al het cijfermateriaal wordt gepresenteerd in tabellen

bij meerder tabellen: resultaten worden geaggregeerd in overzichtstabel

de introductie verwijst vooruit naar belangrijkste resultaat-tabellen

### **PLANNING (alleen in proposal)**

is realistisch

voldoet aan formaat

## **Algemeen**

### **STRUCTUUR**

paper is geschreven volgens (reverse) pyramid (= belangrijke info komt eerst)

paper maakt structuur expliciet: geeft steeds aan wat komen gaat

paper bevat alleen info die nodig is om resultaten te begrijpen, geen zijwegen

### **GLOBAAL**

paper is consistent: er wordt steeds dezelfde notatie / terminologie gebruikt

introduceert alle begrippen die worden gebruikt

referenties naar literatuur zijn expliciet, duidelijk, correct

### **TAALGEBRUIK**

taalgebruik is zo specifiek / concreet mogelijk

er worden korte zinnen gebruikt

het Engels is correct

alle zinnen zijn volledig (werkwoord, persoonsvorm)

het taalgebruik is objectief

er wordt weinig naamwoordstijl gebruikt

taalgebruik is actief

verwijswoorden (die, dat ...) verwijzen correct naar zelfstandige naamwoorden

of zinsdelen (geen "dangling pointers")

... geen "dangling pointers"

### **UITLEG / BEGRIJPelijkheid**

waar mogelijk worden voorbeelden gebruikt ter illustratie

ik kan dit paper uitleggen aan medestudenten

iedereen met een BSc informatica kan dit paper begrijpen

licht moeilijke begrippen, concepten, formules, methoden toe met voorbeelden

... ik kan geen plaatsen verzinnen waar voorbeelden zouden kunnen helpen

deze voorbeelden zijn waar mogelijk plaatjes of diagrammen

### **CORRECTHEID**

dit paper is correct (en daar durf ik een kratje bier om te verwedden)

dit paper is innovatief

alle beweringen zijn onderbouwd met argumenten

argumenten zijn overtuigend en verkregen volgens wetenschappelijke

methoden

# Handling User Feedback in Compact Probabilistic XML Databases

Oliver Jundt  
 University of Twente  
 P.O. Box 217, 7500AE Enschede  
 The Netherlands  
 o.jundt@student.utwente.nl

## ABSTRACT

Many data applications encounter uncertain data and user feedback helps to resolve this uncertainty. The probabilistic XML data model provides basic means for managing uncertain data but no solution exists for handling user feedback when the database is compacted. By focusing on feedback-affected subtrees a solution is developed that works with and retains the compact form. Experiments with varying databases and queries assess the usability of the solution.

## Keywords

Probabilistic XML, User Feedback, Possible Worlds, Uncertainty, Expansion, Simplification

## 1. INTRODUCTION

Many data applications encounter a problem that is often called *uncertainty*. Uncertain data arises for example during data integration or sensor measurements because these processes are inherently error-prone. A common approach to the uncertainty problem is to ignore the uncertainty, pick one of the possibilities for the data and take it as the absolute but possibly incorrect truth. However, nowadays it is widely accepted that the knowledge about different possibilities is valuable information that should be retained [4].

As a result a variety of data models have been developed for managing uncertain data. *Probabilistic XML* is such an uncertain data model and because of its intuitive simplicity it is the model of choice for this research. The probabilistic XML model applies the *possible worlds* semantics for the interpretation of uncertain data. According to the semantics a possible world represents a database without uncertainty and multiple possible worlds build an uncertain database [2]. Consider figure 1 as an example. This probabilistic XML tree represents four possible worlds:

1. There is one person 'John' with telephone number 1111
2. There is one person 'John' with telephone number 2222
3. There is one person 'Mary' with telephone number 1111
4. There is one person 'Mary' with telephone number 2222

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

16<sup>th</sup> Twente Student Conference on IT January 27<sup>th</sup>, 2012, Enschede, The Netherlands.  
 Copyright 2012, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

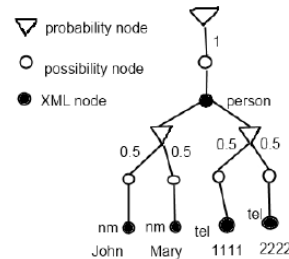


Figure 1. Simple probabilistic XML tree

Every possible world has a probability to be the correct representation of the real world. However, at most one world is indeed the correct representation. For better data quality it is necessary to eliminate possible worlds that are actually *incorrect*. In particular *user feedback* proved to be an effective mean for finding and eliminating possible worlds [5]. To explain user feedback, again consider the example from Figure 1. Assume that a user queried the uncertain database for the telephone number of John. Two possible worlds have information on the telephone number of John so the result will be Table 1.

Table 1. Query result for John's telephone number

nm	tel
John	1111
	2222

Now further assume that the user calls 2222 and finds that 2222 does not belong to a person called John. It is now desirable to eliminate all possible worlds which produced the result *nm='John', tel='2222'* because these possible worlds contradict the real world.

## 1.1 Problem Statement

It is common to store probabilistic XML databases in their *compact* form, where shared nodes between the possible worlds have been combined with a *simplification* algorithm [6]. The simplification creates a semantically equivalent tree but has the advantage to need significantly less space. Unfortunately the compactness leads to a problem when possible worlds need to be removed. For example deleting the tel node containing the number 2222 in Figure 1 is not going to work as intended. Doing so will also remove the possibility of Mary having 2222 as her number. In this case deleting the contradicting possible worlds without affecting other possible worlds requires prior transformation of the tree.



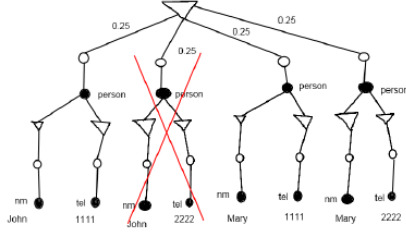


Figure 2. Expanded tree and cut-off

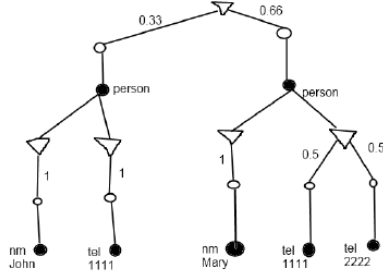


Figure 3. A compact version of the modified example tree

A naive solution to the problem is the complete *expansion* of the compact XML tree. This yields a new equivalent XML tree with a separate subtree for every possible world allowing for a simple cut-off (Figure 2). A subsequent simplification brings the tree back into the preferred compact form (Figure 3). Although the naive solution works, it is unfeasible because during expansion the tree explodes exponentially in size depending on the number of decision and possibility nodes. The simplification of the complete tree afterwards to retain the compact form also costs time.

Figure 4 shows an abstract view on this problem. Instead of making the detour over the completely expanded tree *EPT* it is more preferable to have function *f* that handles the user feedback and removes possible worlds but leaves the tree as compact as possible during the process.

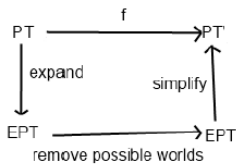


Figure 4. Abstract view of the problem

A closer look reveals that complete expansion is unnecessary if only a subtree is affected by the user feedback. A more sophisticated solution would thus only expand, modify and simplify the affected subtree. This is a new approach that has not been explored before. However, determining which subtrees are affected by a given user feedback and the expansion of probabilistic XML (sub)trees are subproblems that are not solved yet in the literature.

Therefore this research will be twofold. It first aims at solving the subtree finding and expansion problem. Then, as the second step, the new approach is implemented and its performance is analyzed under different conditions.

## 1.2 Research Questions

Based on the problem statement, the research addresses the following research questions:

1. How can the subtrees affected by user feedback be determined?
2. How does the push-up algorithm handle the expansion of a subtree?
3. How is new the solution implemented?
4. To what extent does the new solution avoid explosion in size under different conditions?

## 2. METHOD OF RESEARCH

The answer to the first question needs literature study. Further discussions with Maurice van Keulen will probably also provide valuable information on this topic. It is expected that the traces to the query results contain the information needed to determine the common subtree. Therefore it should not take too long to answer this question. The outcome will be an algorithm.

The answer to the second question needs a detailed study of the current simplification algorithm described by Veldman et al. [6]. The simplification works by a *push down* mechanism that pushes probability nodes down the tree. The algorithm also handles the correct normalization of the probabilities. It should be possible to revert the simplification algorithm into an expansion algorithm using a *push up* approach. To save some time it is necessary to contact Veldman or one of her colleagues and ask for access to the prototypes used in their work. Subsequent expansion, simplification and expansion should result in equivalent trees and is a mean for validating the expansion algorithm.

Based on the results from the previous questions it should be possible to implement the described solution and answer question 3. The implementation can be tested by repeated queries after user feedback, which should show that contradicting possible worlds were indeed removed. A formal verification is more preferable but will only be done if enough time is left at the end of the research.

The answer to the fourth question is derived empirically. The prototype is assessed using experiments with varying databases and artificial user feedback on queries, similar to the experiments performed by Maurice et al. in their work [5]. The test XML files for the experiments should vary in the number of nodes, the tree depth and number of possible worlds. Again time should be saved here by reusing existing test databases. It is expected that on average the new solution needs to expand smaller parts of the tree if the number of nodes and possible worlds is higher. The influence of tree depth is unclear. The queries used for the experiments should cover different tree distances between nodes because it is expected that the solution has its worst case when two nodes involved in a query result have the maximum distance in the tree. In this case the new solution performs as bad as the naive solution. The best case occurs if the nodes have a direct common parent node and no expansion is needed. For each experiment it is possible to measure the maximum node count during expansion and divide it by the original

number of nodes in the tree. The resulting growth factor gives an impression whether the new solution is sufficient for practical use and when the memory explosion is still a problem. The outcome for question 4 will be diagrams showing the growth factor under different conditions.

### 3. RELATED WORK

There exists related work on eliminating possible worlds based on real world knowledge.

A. de Keijzer and M. van Keulen describe a similar naive solution in their work on handling user feedback in probabilistic XML [5][2].

Koch et al. [3] present a way to handle new real world knowledge for relational data models. They express real world knowledge as conditions and therefore call the method *conditioning*. Although their approach works well for relational data models it cannot be easily applied to the probabilistic XML data model since the two models fundamentally differ in the way how the data is structured and how dependencies are represented.

Chen et al. [1] also deal with the relational model and a special case of conditions, so-called aggregate constraints.

### 4. RESEARCH SCHEDULE

The research schedule for this paper is given in Table 2.

Table 2. Research schedule	
Deadline	Task/Deliverable
Oct 28	Get access to test databases and current prototypes used for querying and simplifying probabilistic XML
Nov 4	Answer question 1 and develop subtree finding prototype
Nov 11	Answer question 2 and develop expansion prototype
Nov 18	Answer question 3 and assemble complete prototype
Nov 25	Prepare experiments
Dec 2	Do experiments and answer question 4
Dec 19	First draft paper
Jan 5	Second draft paper
Jan 17	Final camera-ready paper

### 5. REFERENCES

- [1] H. Chen, W. Ku, and H. Wang. Cleansing uncertain databases leveraging aggregate constraints. In *IEEE 26th International Conference on Data Engineering Workshops (ICDEW)*, pages 128–135, 2010.
- [2] A. de Keijzer and M. van Keulen. User feedback in probabilistic xml. 2007.
- [3] C. Koch and D. Olteanu. Conditioning probabilistic databases. *Proceedings of the VLDB Endowment*, 1(1):313–325, 2008.
- [4] M. Magnani and D. Montesi. A survey on uncertainty management in data integration. *Journal of Data and Information Quality*, 2(1), 2010.
- [5] M. van Keulen and A. de Keijzer. Qualitative effects of knowledge rules and user feedback in probabilistic data integration. *VLDB Journal*, 18(5):1191–1217, 2009.
- [6] I. Veldman, A. de Keijzer, and M. van Keulen. Compression of probabilistic xml documents. *Technical Report TR-CTIT-09-20, CTIT*, 2009.

## 7 steps to graduation

