

Voortgangsgesprek 29-08-2003

Richard Brinkman

1 Notulen vorige gesprek

- Wim gaat z'n AIO op een meer gestructureerde wijze begeleiden. Iedere promovendus heeft zijn eigen archief bij Sandra. Van iedere promovendus wordt verwacht dat hij uiterlijk een dag voor een voortgangsgesprek twee A4'tjes inlevert: eentje met een verslag van het vorige gesprek en eentje met het werk dat is uitgevoerd tussen de gesprekken en een prognose van waar je naar toe wilt.
- Wim gaat zich meer richten op het raakvlak security en information retrieval. Ik ben zijn eerste AIO op dit gebied, Iowan¹ de tweede en misschien komt er nog een derde.
- Wim vond het jammer dat ik naar de "verkeerde" summerschool ben gegaan. Hij vindt het belangrijk om de juiste mensen te vinden. De sprekers van de summerschool zijn niet echte security mensen en dus ook niet de juiste personen om papers mee te schrijven. Hij gaat dit ook bespreken met Pieter.
- Wim raadde aan om Ling meer te gebruiken. Een gesprek voeren met haar is misschien wat lastig, maar ze werkt erg hard en kan heel goed schrijven. Ze kan me dan ook heel goed leren om gestructureerd een paper op te zetten.
- Wim zou graag een rapportje zien over mijn geëxperimenteer met de parameters van mijn prototype [1]. Dit rapportje moet ik laten registreren bij het CTIT. Ook de probleemstelling over de lineaire search wil hij graag als een rapportje zien. Zonder rapportjes kan hij mij namelijk geen raad geven, anders dan: "Kijk eens naar B-tree's en R-tree's"
- Over het aanpassen van Hacıgümüş [2] raadde hij me aan om eerst zelf na te denken en dan pas te kijken hoe Ling en Wim het oplossen in hun workshop artikel [3].
- Wim heeft vakantie van half juli tot 22 augustus. Hij zou graag zo snel mogelijk na de vakantie weer een gesprek met mij hebben.

2 Gedaan sinds vorige gesprek

- Mijn experimenten met de parameters van het protocol van Song, Wagner en Perrig [1] zijn beschreven in een rapportje. Het moet nog bij het CTIT

¹Iowan heeft inmiddels afgezegd

geregistreerd worden. Tot die tijd staat het op <http://www.cs.utwente.nl/~brinkman/publications/prototype.pdf>.

- Ik heb het protocol aangepast aan de boomstructuur van XML documenten. Simpel gezegd komt het er op neer dat ik alle *i*'tjes (het index nummer van een blok in de volledige tekst) heb vervangen door de tag-naam. Hiermee ben ik onafhankelijk geworden van de sequentiële volgorde van de woorden. Daarnaast hebben woorden nu een flexibele lengte (in tegenstelling tot de vaste n van het originele artikel). De boomstructuur wordt in verband met de efficiëntie opgeslagen in een relationele database. Er wordt gebruik gemaakt van pre, post en parent waardes die als index dienen. Ik ben momenteel bezig het lineaire prototype uit te breiden naar deze boomstructuur. De eerste resultaten lijken erg hoopvol. Afhankelijk van de query worden hele snelle resultaten geboekt (in de orde van grootte van seconden op een giga byte grote database).

3 Prognose

- Het herschreven protocol heeft nog een aantal nadelen die verholpen moeten worden. Nu worden de pre, post en parent waardes bijvoorbeeld als plaintext in de database opgenomen. En daarmee is de structuur van de boom bekend (alleen de waardes van de nodes natuurlijk niet).
- Het herschreven prototype kan alleen nog maar XPath expressies aan die geen condities en functies aanroepen. Die moeten nog toegevoegd worden.
- Gekeken moet worden of XPath expressies met een // of een * erin niet efficiënter uitgevoerd zouden kunnen worden (met behulp van een extra index, ofzo).

Referenties

- [1] Dawn Xiaodong Song, David Wagner, and Adrian Perrig. Practical techniques for searches on encrypted data. In *IEEE Symposium on Security and Privacy*, pages 44–55, 2000.
- [2] Hakan Hacigumus, Balakrishna R. Iyer, Chen Li, and Sharad Mehrotra. Executing SQL over encrypted data in the database service provider model. In *SIGMOD Conference*, 2002.
- [3] Ling Feng and Willem Jonker. Efficient querying over encrypted xml data.