

GEAVANCEERDE COMPUTERONDERSTEUNING VAN KWANTITATIEF ONDERZOEK

Hajo Rijgersberg, Jan Top, Mari Wigham

Wageningen UR - Information Management,
hajo.rijgersberg@wur.nl, jan.top@wur.nl, mari.wigham@wur.nl

Onderzoek is niet meer denkbaar zonder gebruik van de computer. Met name kwantitatief onderzoek vindt in belangrijke mate op de computer plaats. Toch is de computerondersteuning van kwantitatief onderzoek niet ideaal. Er zijn meerdere, ongekoppelde systemen nodig en veel informatie blijft impliciet en dus niet toegankelijk voor automatische verwerking. De oplossing schuilt in semantische ontsluiting van de kwantitatieve informatie. Daartoe ontwikkelen we een kwantitatief vocabulaire, passen dit vocabulaire toe in ontwikkelde tools en evalueren het gebruik van deze tools met kwantitatieve onderzoekers.

Kwantitatief onderzoek vindt in de meeste, zo niet alle, wetenschappelijke disciplines plaats. De bestudeerde fenomenen en hun eigenschappen worden uitgedrukt met behulp van kwantitatieve concepten zoals getallen, meetschalen, eenheden, mathematische relaties en operaties, tabellen, grafieken, etc. Het belangrijkste aspect van kwantitatief onderzoek is wellicht dat de kennis niet kwalitatief van aard is, maar preciezere relaties en afhankelijkheden impliceert. Bovendien wordt door standaardisatie van kwantitatieve concepten en ijking van meetapparatuur een bepaalde graad van objectiviteit van de wetenschappelijke informatie bereikt.

Kwantitatief onderzoek vindt natuurlijk voor een belangrijk deel op de computer plaats. Berekeningen worden uitgevoerd, gegevens worden bewerkt, grafieken worden gecreëerd. Toch is de computerondersteuning van kwantitatief onderzoek bepaald niet ideaal te noemen. Er zijn meerdere, ongekoppelde systemen nodig en veel informatie blijft impliciet en dus niet toegankelijk voor automatische verwerking. De verschillende computersystemen (LIMS, data-acquisitie, statistische pakketten) hebben gewoonlijk hun eigen vocabulaire, waardoor het onderzoek niet op een geïntegreerde manier kan worden ondersteund en uitgevoerd. Bovendien ontstijgt het gedeelde kwantitatieve vocabulaire in het algemeen niet het numerieke niveau, terwijl het kwantitatieve domein toch een rijk vocabulaire heeft van grootheden, dimensies, eenheden, etc. (Keller en Dungan 1999).

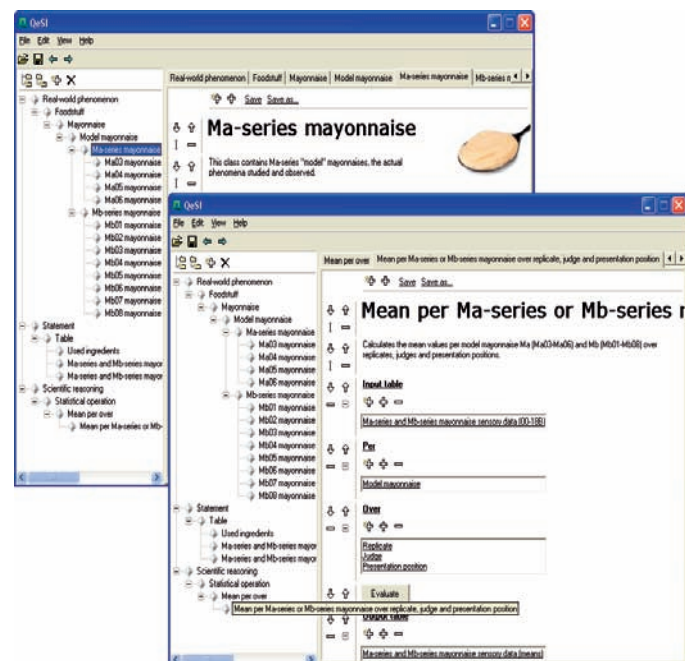
Als gevolg hiervan kan kwantitatieve informatie moeilijk worden gevonden, geïnterpreteerd en geëxecuteerd. De informatie staat vaak los van beschrijvende documenten, waardoor de transparantie van de informatie laag is. De computer kan niet worden ingezet in intelligente, ondersteunende processen omdat de gegevens te ambigu zijn voor de computer. Bijvoorbeeld, een computertool kan niet zien of een bepaalde relatie hypothetisch is dan wel bewezen, of verkregen met een bepaald meetinstru-

ment. Dit soort informatie is natuurlijk cruciaal, maar kan op dit moment niet formeel worden gespecificeerd. Het "mooiste" voorbeeld van problemen met kwantitatieve informatie in computers is misschien wel de neergestorte Mars Climate Orbiter. Door een foutief geïnterpreteerde eenheid bleken 125 miljoen dollar achteraf weggegooid geld.

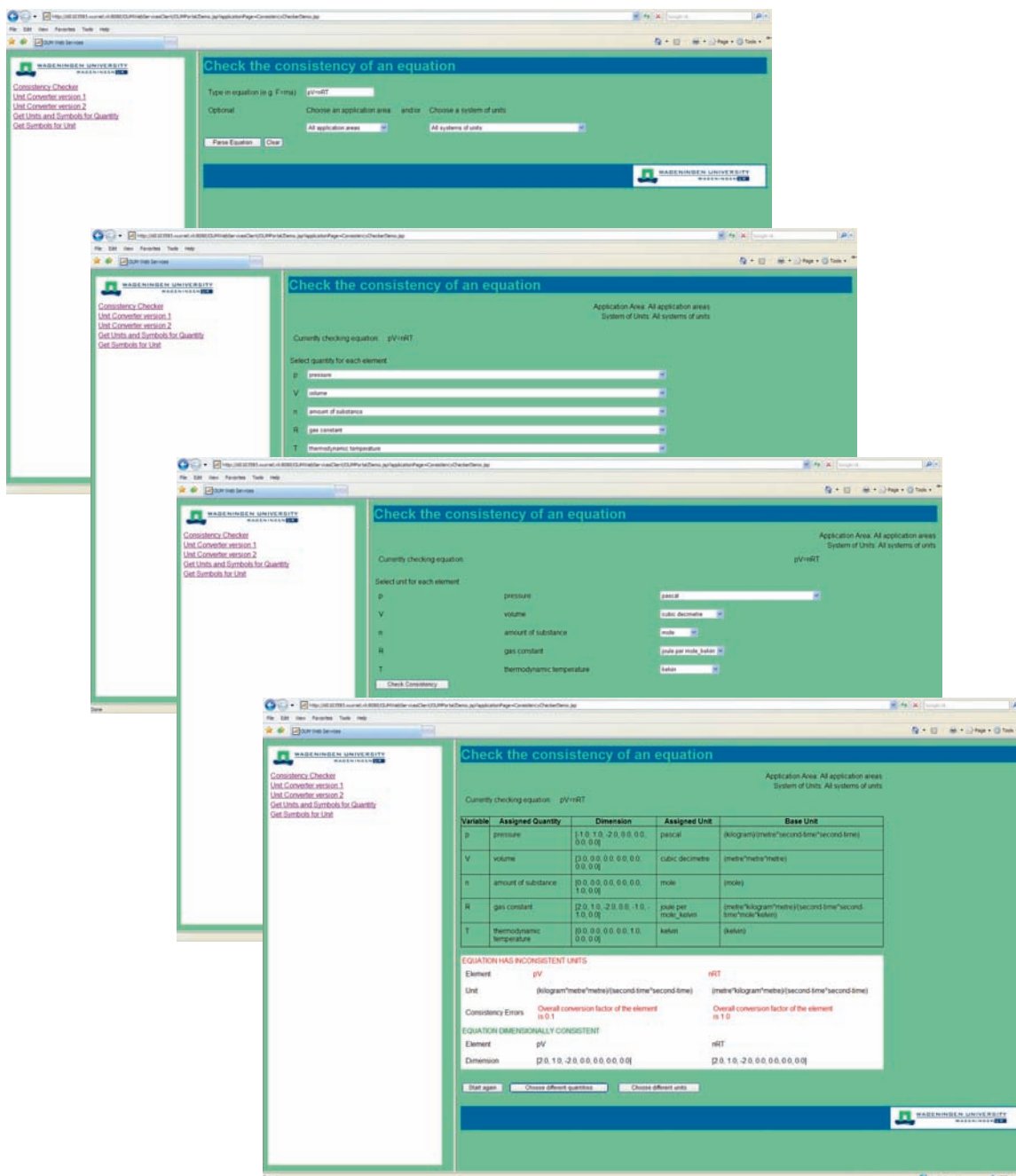
Semantiek

Aan bovenstaande problemen kan wat gedaan worden door de semantiek van kwantitatieve informatie te verbeteren. Alle relevante onderdelen moeten worden gedefinieerd, zodat computertools kunnen worden ontwikkeld die iets zinnigs met al die informatie doen. Wat betreft kwantitatieve informatie zal, behalve de kwantitatieve statements zelf, dus ook het onderzoeksproces, d.w.z. de verkrijgingsmethoden, moeten worden geformaliseerd (Rijgersberg *et al.* 2008).

Kijken we naar het (kwantitatieve) onderzoeksproces, dan zien we dat het een gestructureerd, vaak iteratief proces is waarbij bewijs wordt geëvalueerd, theorieën en hypothesen worden verfijnd, technische vorderingen worden gemaakt, etc. (Langley 2000). Onderzoek start in het algemeen met het formuleren van een onderzoeksvraag. Op basis van de onderzoeksvraag worden



Figuur 1. Prototype kwantitatieve e-science infrastructuur, QeSI, in zijn huidige vorm



Figuur 2. Dimensie- en eenheidconsistentie-checker-wepapp

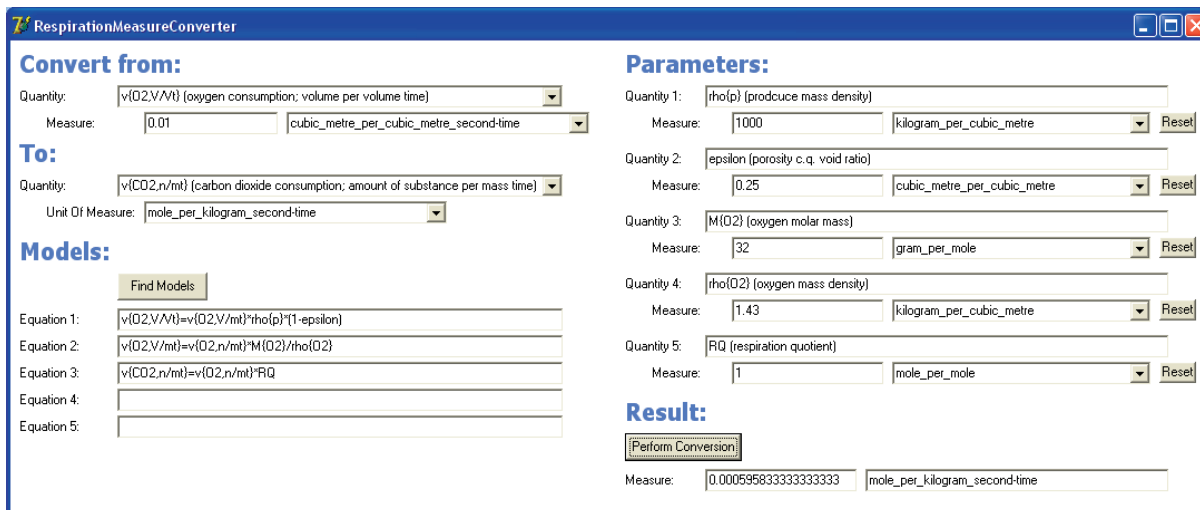
hypothesen geformuleerd, worden data en modellen gezocht, worden experimenten opgezet en onder al dan niet gestandaardiseerde omstandigheden metingen uitgevoerd. Deze metingen worden bewerkt en op basis van deze gegevens wordt een hypothese getest en vervolgens veralgemeniseerd. De onderzoeksvraag wordt beantwoord, het onderzoek wordt gepubliceerd en al dan niet door anderen gereproduceerd. Alle genoemde wetenschappelijke *redenerstappen* worden gemaakt op basis van (uitspraken over) de beschouwde fenomenen. In kwantitatief onderzoek zijn dit soort statements kwantitatief van aard. *Mathematische constructen* zijn nodig om ze te beschrijven. De concepten die direct relateren aan de beschouwde fenomenen zijn *grootheden*, *eenheden*, *meetschalen*, etc. Redenerstappen specifiek voor kwantitatief onderzoek zijn *berekeningen*. Berekeningen vinden gewoonlijk op de computer plaats op basis van methoden die, behalve mathematische constructen, ook zijn opgebouwd uit *programmaconstructen* (if, then, while, etc.). Vaak zijn deze

methoden gerelateerd aan specifieke *software*, zoals Matlab. Bovenstaande concepten hebben we in een ontologie gedefinieerd, genaamd OQR (Ontology of Quantitative Research). Het voordeel van een ontologie boven informatie los opslaan is dat de informatie gestandaardiseerd en centraal beschikbaar is in een toegankelijke vorm. OQR is nog in ontwikkeling, maar gereed zijnde delen ervan, en prototypische delen, passen we reeds toe in ontwikkelde tools, die we vervolgens ook hebben geëvalueerd met kwantitatieve onderzoekers, de beoogde doelgroep.

Tools

De tools die we ontwikkelen zijn onder te verdelen in vier groepen:

1. infrastructuur voor de specificatie van kwantitatief onderzoek en communicatie van kwantitatieve informatie tussen numerieke tools zoals Matlab,



Figuur 3. REspiration Measure Converter

2. basisondersteuning met betrekking tot eenheden, grootheden, dimensies, etc., zoals het checken van dimensie- en eenheidconsistentie van wiskundige formules, en het converteren van eenheden en meetschalen in bijvoorbeeld data,
3. het omrekenen van grootheden en eenheden in een bepaald domein op basis van wiskundige modellen,
4. het upgraden van reeds bestaande kwantitatieve informatie in spreadsheets.

Figuur 1 toont twee screenshots van de kwantitatieve infrastructuur, QeSI (Quantitative e-Science Infrastructure). In het linker-gedeelte van het scherm worden concepten uit OQR getoond. Het rechtergedeelte toont details van het geselecteerde concept. Berekeningsprocessen kunnen met behulp van QeSI worden uitgevoerd. Het systeem kleedt de semantisch rijke kwantitatieve informatie uit de ontologie uit tot numeriek niveau, communiceert het door naar de beoogde software, en upgrade de resultaten verkregen uit het berekeningsproces weer tot het semantische niveau.

Figuur 2 toont een dimensie- en eenheidconsistentiechecker. De gebruiker kan voor een formule grootheden en eenheden en meetschalen aangeven. Op basis van in OQR aanwezige dimensionale kennis is de tool in staat de formule te evalueren op dimensie- en eenheidconsistentie.

Figuur 3 toont de Respiration Measure Converter, een tool om grootheden en eenheden uit het productrespiratiedomein om te rekenen, op basis van wiskundige modellen uit dit domein. Wat betreft QeSI zijn we met een kwantitatief onderzoeker ontwikkeld in een iteratief proces van ideevorming, implementatie, evaluatie, aanpassing en ga zo maar door. Dit evaluatieproces heeft geleid tot waar we nu zijn. Nu moet de ontologie nog beter worden gepresenteerd, of liever aan de gebruiker worden gedoseerd. Terugkoppeling met de gebruiker heeft geleerd dat het presenteren van *studies*, ketens van redeneerstappen die steeds een bepaald stuk (kwantitatieve) kennis als output hebben, veelbelovend is en de onderzoeker houvast geven. In het systeem kun je wel al goed alles kwijt, een belangrijke voorwaarde. De andere twee soorten tools, de basistools en de omreken tools, zijn nu voldoende ontwikkeld voor evaluatie en eerste gebruik. Beide voorbeelden zijn met kwantitatieve onderzoekers geëvalueerd. De onderzoekers bevestigen de relevantie van dit soort tools voor kwantitatief onderzoek. De gebruikers geven aan dat

het nut van dit soort tools nog duidelijker zou worden als geïntegreerd zijn in veelgebruikte computer tools zoals spreadsheets. Deze kant willen wij binnenkort ook op. In een breder perspectief willen we bovenstaande tools en het toegepaste kwantitatieve vocabulaire op termijn integreren in Tiffany, een onderzoeksondersteunend systeem in ontwikkeling (Top & Broekstra 2008).

Conclusie

Semantiek is het codewoord in dit onderzoek. De kerngedachte is dat het numerieke niveau op de computer ontstegen wordt en daardoor de ondersteuning en uitvoering van het kwantitatieve onderzoek verbetert dan wel versnelt. We hebben kwantitatief vocabulaire ontwikkeld en passen dit vocabulaire toe in de ontwikkelde en te ontwikkelen tools. Deze tools zijn, al dan niet in een voortdurend iteratief proces, geëvalueerd met kwantitatieve onderzoekers. Het vocabulaire is in grote lijnen klaar, een begin is gemaakt aan de ontwikkeling van een serie tools, en de stap naar de beoogd gebruikers is gezet. De eerstverkrege feedback geeft aan dat de ingeslagen weg een juiste is. De volgende stap is het vocabulaire en de tools te integreren in veelgebruikte software zoals spreadsheets. Het is belangrijk de stap te zetten naar data-ondersteuning omdat die tak van sport door een groot deel van de onderzoekers bebezigt wordt. Dat zijn de gebruikers, meer dan bijvoorbeeld modellers, voor wie de beoogde kwantitatieve ondersteuning onontbeerlijk is. Een kracht van de ingeslagen weg zit hem in de ontwikkeling van enerzijds vocabulaire en anderzijds tools. Dankzij het beschikbare vocabulaire stellen we dat tool-bouwers beter in staat een hoger niveau van computerondersteuning te ontwikkelen.

Referenties

- Keller, R.M. and Dungan, J.L. 1999. Meta-Modeling: a Knowledge-Based Approach to Facilitating Process Model Construction and Reuse. *Ecological Modelling* 119: 89-116.
- Langley, P. 2000. The computational support of scientific discovery. *Int. J. Human-Computer Studies* 53: 393-410.
- Rijgersberg, H., Top, J. & Meinders, M. 2008. Use of a quantitative research ontology in e-science. In AAAI 2008 Spring Symposia. Palo Alto, Californië.
- Top, J. & Broekstra, J. 2008. Tiffany: sharing and managing knowledge in food science. Keynote in ISMICK, Brazilië.