

Spraak- en Taaltechnologische Essentiële Voorzieningen In het Nederlands (STEVIN)

<http://taalunieversum.org/stevin>

STEVIN Jaarverslag 2010 - d.d. 15 augustus 2011

In dit jaarverslag is een overzicht opgesomd van de activiteiten die van 1 januari t/m 31 december 2010 zijn uitgevoerd in het kader van het STEVIN Programma. Daarnaast is de oorspronkelijke begroting en de realisatie voor het jaar 2010 opgenomen en toegelicht. De realisatie is gebaseerd op de financiële afrekening zoals aangeleverd door de NTU. Verder zijn in dit jaarverslag ook de nieuw in 2010 aangegane committeringen opgenomen die nog niet in het financiële overzicht van de NTU staan. Dit betreft kleine netwerksubsidies die over het algemeen door de NTU pas bij uitbetaling in de afrekening opgenomen worden.

1. Beheer, voortgangscontrole en afsluiting STEVIN-onderzoeksprojecten

In het afgelopen jaar is vooral aandacht besteed aan de voortgangscontrole en het afsluiten van projecten. De voortgangscontrole en eindcontrole worden uitgevoerd op basis van de *workflows* die het Programmabureau (PB) met de TST-Centrale in 2007 hebben opgesteld. De controles worden gedaan door de portefeuillehouders (PF's) die aan elk project zijn toegewezen, te weten: per project twee PC-leden en de TST-Centrale vertegenwoordiger in de STEVIN-PC die specifiek aandacht besteedt aan de IPR-zaken en technische zaken.

De PF's hebben voor alle lopende projecten voortgangs- dan wel eindfiches opgesteld die, nadat ze waren geaccordeerd door de PC, zijn voorgelegd aan het bestuur.

In totaal zijn er gedurende de looptijd van het STEVIN-programma na 3 open rondes en 3 tenders 19 grote onderzoeks- en ontwikkelprojecten gehonoreerd.

- Reeds volledig afgerond voor 31/12 2009: 8 projecten
uit Oproep 1: *COREA, IRME, D-COI, JASMIN en AUTONOMATA*
uit Tender: 1 *SPRAAK*
uit Oproep 2: *N-Best en STEVINcanPRAAT*
- Afgerond in 2010: 3 projecten
uit Oproep 2: *DPC, DAESO*
uit Tender 2: *CORNETTO*
- Nog lopend: 8 projecten
uit Oproep 1: geen
uit Oproep 2: *MIDAS, LASSY*
uit Tender 3: *SoNaR*
uit Oproep 3: *DISCO, AUTONOMATA TOO, DAISY, PACOMT, DUOMAN*

Het project *DPC* is afgerond met dien verstande dat er voor 2011 alleen nog een bedrag resteert dat aan de externe validator moet worden overgemaakt zodra die zijn werk heeft afgerond.

Momenteel zijn er nog 8 lopende STEVIN O&O projecten. In het kader van de voortgangscontrole zijn bij 7 projecten vorig jaar site visits uitgevoerd. Voor het 8^e project, *PaCoMT*, moet nog een site visit gepland worden. Sommige projecten kampen met enige vertraging oplopend tot 7 maanden. Deze is met name te wijten aan een tragere start vanwege de problemen bij de invulling van vacatures. Er zijn geen serieuze technische problemen geconstateerd.

Het *DISCO* project heeft vanwege het faillissement van de firma Polderland enige vertraging opgelopen. Het PB heeft uitgebreid overleg gevoerd met de andere deelnemers van het consortium. Uiteindelijk is Knowledge Concepts - die de kennis van Polderland heeft overgenomen - in het consortium gestapt. Er is in 2010 een verlenging aangevraagd en toegekend.

De meeste projecten zouden in 2011 afgerond moeten kunnen worden met een heel bescheiden uitloop in 2012: de einddatum van het *SoNaR*-project is vastgesteld op 1 december 2011. Na afloop geldt ook voor *SoNaR* de regulier voorziene periode van 3 maanden om met de TST-Centrale de overdracht en afsluiting te realiseren.

Naast deze O&O projecten is in 2010 ook een aantal demonstratieprojecten en educatieve projecten afgerond. Meer informatie daarover staat in sectie 3.

2. Intellectual Property Rights (IPR) en Prijzen politiek

Gezien de ervaringen die tot nu toe zijn opgedaan binnen STEVIN en de aanbevelingen van het International Assessment Panel (IAP) in de midterm evaluatie, bleek een optimalisering van de bestaande IPR-regeling in de rede te liggen. In eerste instantie heeft dat geleid tot verruiming van de mogelijkheden voor het aanvragen van ontheffing van de staande STEVIN IPR-regeling zodat resultaten als open source beschikbaar gemaakt worden.

Daarnaast had de PC in 2009 een notitie opgesteld over IPR-beleid voor een eventueel nieuw programma. Het uitgangspunt van de nieuwe regeling blijft TST-materialen zo maximaal mogelijk beschikbaar te maken zowel voor onderzoekers als voor het bedrijfsleven. Dit geldt natuurlijk in elk geval voor de door de overheid gefinancierde TST-materialen, maar waar mogelijk voor alle (Nederlandstalige) TST-infrastructuur. De notitie is in de vergadering van het TST-bestuur op 29 april 2010 besproken. Besloten is te wachten met een besluit tot er input vanuit de IPR-workshop verkregen zou zijn.

Voor de reeds beschikbare STEVIN-resultaten zijn inmiddels de niet commerciële/commerciële prijzen vastgesteld: AUTONOMATA namencorpus (€0, €4.000), AUTONOMATA g2p-toolkit (€500, op aanvraag), COREA coreferentiecensus (€0, €3.000), IRME DuELME-lexicon (€0, €2.500), JASMIN-CGN spraakcorpus (€0, €2.500), D-COI-corpus (€0, €12.000), CORNETTO (€0, €15.000), NBest (€0, op aanvraag), LASSY (€0, nvt) en tussenresultaten SoNaR ((€0, nvt). Alle vastgestelde prijzen zijn op de website van de TST-Centrale gepubliceerd en zijn exclusief eventuele verzend- en handlingkosten.

3. Flankerende activiteiten

3.1 Begripsomschrijving en doelstellingen

In het kader van het STEVIN-programma wordt, in aanvulling op de onderzoeks- en ontwikkelprojecten, een aantal flankerende activiteiten gefinancierd. Hieronder vallen:

- de organisatie van symposia, bijeenkomsten en andere publicitaire evenementen, alsook het (financieel) ondersteunen van STEVIN-gerelateerde bijeenkomsten. Dat laatste gebeurt onder meer via het toekennen van netwerksubsidies voor de organisatie van dergelijke bijeenkomsten;
- de STEVIN-nulmeting (om de stand van zaken op het vlak van Nederlandstalige taal- en spraaktechnologie in kaart te brengen bij aanvang van het programma, en na afloop de impact van het STEVIN-programma te meten), de tussentijdse wetenschappelijke evaluatie en de eindevaluatie;
- het ondernemen van initiatieven die zijn gericht op het stimuleren van de vraag naar (Nederlandstalige) taal- en spraaktechnologie. In dit verband worden vanuit het STEVIN-programma onder meer ook demonstratieprojecten gefinancierd en masterclasses voor het bedrijfsleven;
- het ondernemen van initiatieven om taal- en spraaktechnologie onder de aandacht te brengen van (schoolgaande) jeugd en studenten opdat deze hierin geïnteresseerd raken en een professionele carrière op dit domein willen starten. In dit verband worden vanuit het STEVIN-programma onder meer educatieve projecten gefinancierd als ook initiatieven die tot doel hebben taal- en spraaktechnologie te verankeren op HBO's.

De flankerende activiteiten in het kader van het STEVIN-programma worden geïnitieerd en uitgewerkt door de *Werkgroep Flankerende Activiteiten* (kortweg *WGFA*), een subcommissie van de PC, aangevuld met de makel- en schakel verantwoordelijke van de NTU, enkele andere Vlaamse en Nederlandse vertegenwoordigers van andere taal- en spraaktechnologische projecten en vertegenwoordigers van (mogelijke) gebruikersgroepen van (Nederlandstalige) taal- en spraaktechnologie. Het idee hierachter is dat flankerende activiteiten en met name het organiseren van taal- en spraaktechnologische symposia en dergelijke, zoveel mogelijk met alle TST-actoren in Nederland en Vlaanderen gezamenlijk zou moeten geschieden. De door de WGFA geconcipeerde plannen worden voorgelegd aan de PC die ze vervolgens ter goedkeuring voorlegt aan het TST-bestuur.

In 2009 was, gezien de eindfase waarin het STEVIN-programma zich bevond, al besloten om na het vertrek van mw. Debbie Kenyon-Jackson zowel de WGFA als de PC te laten voorzitten door de voorzitter van de PC,

professor Jan Odijk. In de loop van 2010 is daarnaast besloten om de vergaderingen van de PC en de WGFA gezamenlijk te laten plaatsvinden.

In de secties 3.2.1 tot en met 3.2.7 worden de activiteiten nader toegelicht en gemeld welke specifieke activiteiten in 2010 hebben plaatsgevonden.

3.2 *Specifieke flankerende activiteiten*

3.2.1. Eigen activiteiten

NTU kosten

Binnen het begrotingskader voor "eigen activiteiten" is een vaste post "vergaderkosten" (inclusief vacatiegelden) en "kantoorkosten/accountantskosten NTU" opgenomen.

In 2010 hebben de volgende STEVIN vergaderingen plaatsgevonden:

- STEVIN Programmacommissie: 4 vergaderingen, in januari, april, juni en november 2010 (de laatste samen met de WGFA);
- STEVIN Werkgroep Flankerende Activiteiten: 1 vergadering, in november 2010 (samen met de PC);
- STEVIN IPR Werkgroep: geen bijeenkomsten;
- TST-bestuur: 2 vergaderingen: april en september 2010.

Het grootste deel van de kosten betreft vacatiegelden voor leden van het TST-bestuur, de PC, de WGFA en de Prijzencommissie als ook de vergaderkosten van het TST-bestuur en de Prijzencommissie. De vergaderkosten van de PC en de WGFA worden betaald door NWO en Agentschap NL vanuit het programmabureaubudget.

TST op Wikipedia

Via deze actie zal zoveel mogelijk informatie over TST op Wikipedia gezet worden. Een groep representanten uit het Vlaamse en Nederlandse taal- en spraaktechnologieveld heeft te kennen gegeven voor een beperkte vergoeding de informatie in Wikipedia te willen invoeren. De totale kosten inclusief een kleine donatie aan Wikipedia waren begroot op € 6.000. In de loop van 2010 is besloten dat mede vanwege het feit dat een vervolg op het Educaproject TST op Kennislink werd gehonoreerd, de begrote kosten naar beneden bijgesteld konden worden, tot € 2.000. Het project zal in 2011 afgerond worden.

IPR workshops / overleg met uitgevers over gebruik (IPR) van hun teksten:

In lopende STEVIN projecten is gebleken dat het van belang is om contacten met uitgevers en eigenaren van bronmateriaal voor het regelen van het gebruik van teksten te coördineren. Ook in het kader van het in 2009 in Nederland door OCW met 9M€ gesubsidieerde CLARIN project wordt men met deze problematiek geconfronteerd. De STEVIN Programmacommissie zou in samenwerking met de TST Centrale, de IPR Commissie, de NTU en vertegenwoordigers van het CLARIN project een aantal workshops coördineren en realiseren in 2009 en 2010. Maximaal € 20.000 van het budget bestemd voor eigen activiteiten was hiervoor gereserveerd. In tegenstelling tot de plannen hebben zowel in 2009 als in 2010 geen workshops plaatsgevonden omdat de noodzakelijke experts vanwege andere werkzaamheden te weinig tijd hadden. Mede daarom is besloten het aantal workshops terug te brengen tot één workshop - te organiseren in 2011 - en het budget naar beneden bij te stellen.

Initiatieven ten behoeve van toekomstig TST-onderzoek, ontwikkeling en vraagstimulering

Het STEVIN programma eindigt formeel in 2011. Tot op heden is het niet gelukt een vervolg te realiseren. Met name aan Nederlandse zijde is nieuw geld niet te vinden. Maar ook is er behoefte aan een goed plan dat de nieuwe fondsenverwerving zou ondersteunen en motiveren.

In dit kader zou de STEVIN Programmacommissie in samenwerking met de WGFA in 2009 en 2010 een aantal kleine *Roadmap Workshops* (maximaal 10 deelnemers per workshop) organiseren. Hiervoor was in totaal een bedrag van € 14.000 gereserveerd voor 2009-2010. Daarnaast heeft de NTU, uit het eigen budget gefinancierd, aan prof. Lou Boves een opdracht gegeven om - in overleg met een aantal Vlaamse en Nederlandse experts - een toekomstvisie met bijbehorend werkplan voor de taal- en spraaktechnologie op papier te zetten.

In tegenstelling tot de plannen heeft in 2009 en ook in 2010 nog geen workshop plaatsgevonden. Ook is de eerste concepttekst van Boves nog niet gereed.

STEVIN-programmadag

De STEVIN-programmadag is gehouden op 14 oktober 2010 in Antwerpen. Dit is een jaarlijks georganiseerde interne dag voor de direct betrokkenen van het STEVIN-programma plus de contactpersonen voor die STEVIN-projecten bij de TST-Centrale. Het idee was om voor deze programmadag met name ook gebruikers uit te nodigen. Daartoe waren duopresentaties van projectonderzoekers en gebruikers in het dagprogramma opgenomen naast twee presentatie van twee internationale experts (Fabio Ciravegna en Steve Renals). Parallel aan de expert presentaties was een speciale gebruikerssessie van het project LASSY. De deelname van echte gebruikers uit de neerlandistiek was helaas beperkt ondanks dat de dag aangekondigd was in mailinglijsten speciaal gericht op die doelgroep.

3.2.2. Beheer, opvolging en afsluiting demonstratieprojecten en zoeken naar extra financiering

Het doel van de demonstratieprojecten is het stimuleren van de vraag naar (Nederlandstalige) taal- en spraaktechnologie. Het betreft kortlopende voorbeeld- en/of demonstratieprojecten (maximale projectduur is 15 maanden), waarin gebruik wordt gemaakt van "proven technology". Daarbij wordt vooral gemikt op de ontsluiting van nieuwe markten en het toepassen van bestaande taal- en spraaktechnologische applicaties in andere - nieuwe - domeinen. In 2008 is een start gemaakt met het actiever monitoren van de projecten. Daartoe is in de WGFA per project een portefeuillehouder aangewezen. In totaal zijn er gedurende de looptijd van het STEVIN-programma in 3 rondes 14 demonstratieprojecten gehonoreerd.

- Reeds volledig afgerond voor 31/12 2009: 8 projecten
uit Demo-oproep 1: *Rechtsorde, GemeenteConnect, Kentekenlijn*
uit Demo-oproep 2: *Web Assess, Klinkende Taal, Audiokrant, Rechtspraakherkenning*
uit Demo-oproep 3: *HATCI*
- Afgerond in 2010: 4 projecten
uit Demo-oproep 2 en 3: *SpeelSpiek, YourNews (voorheen EasyInfo), Woody, NEON.*
- Nog lopend: 2 projecten
uit Demo-oproep 2 en 3: *PRIMUS, AAP.*

Vanwege het faillissement van de firma Polderland, zijn er problemen ontstaan bij het afronden van de demonstratieprojecten *Aap* en *PRIMUS*. Het PB heeft in 2009 overleg gevoerd met het consortium van het project *AAP* (Alfabetisering Anderstaligen Plan). Ook met het consortium van *PRIMUS* is het PB in overleg gegaan, in beide gevallen over een mogelijke (beperkte) doorstart in 2010. Eind 2010 heeft het TST-bestuur besloten akkoord te gaan met de doorstart van beide projecten in 2011.

Ook dit jaar heeft met name de NTU weer veel aandacht besteed aan de "etalage" op de STEVIN website waarin de resultaten van de demonstratieprojecten gepresenteerd worden.

Zie: <http://taalunieversum.org/taal/technologie/stevin/etalage/>.

De demonstratieprojecten spelen een belangrijke en succesvolle rol bij het verwezenlijken van de STEVIN doelstellingen en met name voor het beter zichtbaar maken van de haalbaarheid en de waarde van de inbedding van taal- en spraaktechnologie in bepaalde applicaties en diensten. Ook in 2010 is weer tevergeefs gepoogd om aan Nederlandse zijde nieuwe financieringsmogelijkheden te vinden voor demonstratieprojecten.

3.2.3. Nulmeting, tussenevaluatie en monitoring

Om de voortgang en de impact van het STEVIN-programma te kunnen meten is het van belang het STEVIN-programma, en de verschillende projecten die binnen dit programma worden gefinancierd, doeltreffend te monitoren en te evalueren. Daartoe is in de 1^e fase van het STEVIN-programma een nulmeting uitgevoerd. De kosten daarvan bedroegen € 20.000.

In 2008 is er een wetenschappelijke tussenevaluatie uitgevoerd door het International Assessment Panel. De aanbevelingen die daaruit zijn voortgekomen zijn voor zover mogelijk geïmplementeerd. De kosten van de tussenevaluatie, die samen met het NWO IMIX programma was georganiseerd, bedroegen circa € 31.000, waarvan ruim de helft is betaald uit het IMIX budget.

Dat betekent dat er voor de eindmeting nog een bedrag van maximaal € 60.000 beschikbaar was.

De NTU heeft in 2009 samen met EWI en in overleg met de andere financiers de opzet van en opdracht voor de uitvoering van de eindmeting en eindevaluatie geformuleerd. De opdracht daarvoor is gegund aan Technopolis. Het PB heeft begin 2010 in samenwerking met de PC een uitgebreid evaluatiedossier voorbereid bestaande uit een Zelfevaluatie (38 pagina's), een Factfile (83 pagina's) en een bundeling van de voortgangs- en eindrapportages van alle STEVIN-projecten.

In de periode februari - juli 2010 heeft Technopolis vervolgens gewerkt aan de eindmeting en eindevaluatie volgens het door het TST-bestuur goedgekeurde protocol. Op basis van het evaluatierapport heeft het TST-bestuur geconcludeerd dat:

1. de belangrijkste doelstellingen in hoge mate zijn behaald;
2. de gezamenlijke investeringen van Nederland en Vlaanderen in taal en spraaktechnologie op de lange termijn tot besparingen leiden, dat de expertise en middelen maximaal benut worden, dat doublures vermeden worden en de kwaliteit van de resultaten verhoogd is;
3. nieuwe steun voor de ontwikkeling van taal- en spraaktechnologie in een gezamenlijk programma noodzakelijk blijft.

Daarnaast constateerde het TST-bestuur dat het Technopolis rapport eerder beschrijvend van aard is en dat een oordeelsvorming over de vaststellingen in het rapport op punten ontbreekt.

Het volledige eindevaluatierapport en het onderliggende dossier is in te zien via de STEVIN-website (<http://taalunieversum.org/taal/technologie/stevin/programma/#evaluaties>).

3.2.4. Toekenning van netwerksubsidies

Netwerksubsidies zijn bedoeld voor de ondersteuning van de organisatie van (wetenschappelijke) bijeenkomsten die zijn gericht op het bevorderen van de netwerkvorming en de interactie binnen het taal- en spraaktechnologische veld in (en tussen) Vlaanderen en Nederland, en/of die kunnen bijdragen tot een vruchtbare uitwisseling van ideeën en resultaten tussen de verschillende onderzoeks- en gebruikersgroepen op het gebied van (Nederlandstalige) taal- en spraaktechnologie. Ook kan deelname aan congressen en beurzen, in de vorm van sponsorovereenkomsten, worden gesteund om de naambekendheid van het STEVIN-programma te vergroten. Voor de jaren 2010 en 2011 was hiervoor in totaal nog € 15.000.gereserveerd.

In 2010 hebben 5 activiteiten plaatsgevonden die door STEVIN financieel zijn gesteund.

- Netwerksubsidie CLIN 20, 2010 € 1.500 (reeds toegekend in 2009)
- Netwerksubsidie DIR 2010 € 500 (reeds toegekend in 2009)
- Netwerksubsidie TABU-dag 2010 € 1.400
- Netwerksubsidie Gridline TST-stand op Overheid en ICT € 2.500
- Netwerksubsidie LOT 2010 € 750

In 2010 zijn nog 3 nieuwe netwerksubsidies toegekend die waren ingediend in de continue openstaande oproep. Deze activiteiten zullen pas in 2011 plaatsvinden en tot uitbetaling leiden.

- Netwerksubsidie CLIN 21, 2011 € 2.000
- Netwerksubsidie EAMT 2011 € 1.500
- Netwerksubsidie DIR 2011 € 600

3.2.5. Externe communicatie

In 2010 hebben alle STEVIN-geledingen wederom de onderzoeks- en ontwikkelresultaten uit het STEVIN-programma waar mogelijk onder de aandacht gebracht op (inter)nationale conferenties. Ook zijn projectconsortia naar aanleiding van het Midterm advies van het IAP gestimuleerd om hun resultaten in internationaal hoogstaande tijdschriften te publiceren. Omdat een belangrijk deel van de projecten betrekking heeft op de ontwikkeling van basisinfrastructuur, blijkt dit echter niet altijd eenvoudig te realiseren. Een overzicht van alle STEVIN publicaties wordt bijgehouden op de STEVIN website. Het overzicht van de STEVIN publicaties uit 2010 is aan het eind van dit verslag te vinden.

Als onderdeel van de algemene NTU website (<http://taalunieversum.org/>) bestaat een specifieke STEVIN website die door de NTU wordt onderhouden. Van 1 januari tot 31 december 2010 heeft de website 5922 bezoeken gehad.

Begin 2010 was besloten dat er een communicatieplan voor de besteding van de rest van het budget gereserveerd voor externe communicatie opgesteld zou worden door de PC/WGFA dat ter goedkeuring voorgelegd zou worden aan het TST-bestuur. Daarvoor was een bedrag van ruim € 40.000 gereserveerd. In de loop van het jaar werd besloten om geen nieuw plan te maken maar om de activiteiten te concentreren rondom het grote eindevenement. Volgend jaar zal gekeken worden of er nog budget over is voor andere communicatieactiviteiten. In 2011 zal ook duidelijk gedefinieerd worden wat de rol en de verantwoordelijkheden zijn van het communicatiegroepje.

3.2.6 Interne communicatie: STEVIN-website voor intern gebruik (WIKI)

K.U.Leuven (Van hamme) heeft de interne WIKI site opgezet en is verantwoordelijk voor het hosten en onderhouden van deze site. De kosten hiervoor bedragen 2.000 euro per jaar. Er is voldoende geld gereserveerd om de site ook na de afloop van het STEVIN programma nog een tijd in de lucht te houden. Het PB heeft er op toegezien dat de STEVIN projecten de noodzakelijke informatie via de WIKI beschikbaar maken en dat ook de informatie op het WGFA-gedeelte bijgewerkt wordt.

3.2.7. Educatieve activiteiten (educaprojecten en masterclasses)

In het kader van de educatieve activiteiten zijn twee specifieke subsidieregelingen opgesteld: educaprojecten en masterclasses. Via meerdere kanalen is voor deze beide subsidiemogelijkheden geadverteerd (TST-nieuwsbrief, NoTaS-leden, alle aanwezigen van de beide Taal-in-Bedrijf evenementen) en in specifieke Nederlandse en Vlaamse onderwijsbladen.

Educatieve projecten (min. 3 maanden en max. 12 maanden) bedoeld om in een educatieve omgeving (middelbare school, hogeschool, universiteit, museum, ...) schoolgaande jongeren en beginnende studenten bewust te maken van de mogelijkheden van taal- en spraaktechnologie:

- als multi-inzetbare technologie van de toekomst;
- als onderwerp van een (wetenschappelijke) opleiding;
- als sector voor een eventuele professionele loopbaan.

In 2007-2009 zijn er drie oproepen voor educatieve voorstellen opengesteld. In totaal zijn er uit drie rondes, drie projecten gehonoreerd: Kennislink I, Diademo en Kennislink II. De eerste twee projecten zijn in 2010 succesvol afgerond. Het derde project zal pas in 2011 eindigen.

Masterclasses zijn bedoeld om de belangstelling voor taal- en spraaktechnologische toepassingen en onderzoek binnen overheden en bedrijfsleven te vergroten. In 2008 en 2009 zijn twee oproepen gepubliceerd. In totaal zijn er twee masterclasses gehonoreerd: ICT & Dyslexie, TST in NL Overheidsdiensten. De beide masterclasses zijn in 2010 gegeven. Het afrondingsproces voor de beide masterclasses is in 2010 gestart en zal in 2011 worden afgerond. Uitbetaling heeft voor de masterclasses al wel in 2010 plaatsgevonden.

4. Financiële verantwoording- STEVIN uitgaven

In deze sectie wordt een overzicht gegeven van de uitgaven/toekenningen gedaan van 1-1-2010 tot en met 31-12-2010 en een toelichting op het verschil tussen de goedgekeurde begroting voor 2010 en de realisaties.

In het STEVIN jaarwerkplan 2011 is een volledig meerjarig financieel overzicht opgenomen over de periode 2004 t/m 2011.

In de realisatie zijn opgenomen:

- 1) de daadwerkelijk ontvangen bijdragen van de financiers;
- 2) daadwerkelijke betalingen aan projectkosten en gedane committeringen voor O&O en demoprojecten;
- 3) committering voor flankerende activiteiten.

Committeringen voor kortlopende educatieve projecten, masterclasses en netwerksubsidies worden door de NTU pas bij de uitbetaling in de realisaties opgenomen maar deze worden hieronder wel apart opgesomd per deelbudget en zijn ook verwerkt in de meerjarenbegroting zoals opgesteld door het Programmabureau en opgenomen in het STEVIN Werkplan 2011.

Nadere toelichting inkomsten

In het kader van het Vlaams-Nederlandse STEVIN programma is in 2004 door de gezamenlijke financiers afgesproken dat het STEVIN programma volgens de standaard verdeling 1/3 - 2/3 gefinancierd zou worden. Daarnaast werd overeengekomen dat een klein deel van het Nederlandse bijdrage zou bestaan uit een bijdrage in nature vanuit het NWO IMIX-programma dat eind 2002 was gestart. Het IMIX-programma was opgestart met een bedrag van € 2.000.000 dat NWO oorspronkelijk beschikbaar had gesteld voor de start van het BaTaVo programma dat door de ineenstorting van L&H niet was doorgedaan. Afgesproken is dat de matching bijdrage vanuit NWO IMIX aan het Vlaams-Nederlandse STEVIN programma € 250.000 zou bedragen. De IMIX projecten zijn reeds in 2003 gestart en in 2009 afgerond. In 2010 hebben EZ en NWO het laatste deel van hun bijdrage overgemaakt aan de NTU. EWI en OCW hadden dat al in 2009 gedaan.

	Herziene begroting 2010 d.d. maart 2010 (€)	Realisatie 1/1/2010 t/m 31/12/2010 (€)
<u>Inkomsten</u>		
MVG-EWI	0	0
EZ (IOP en CIC)	500.000	500.000
NWO-GW, NWO-EW en NWO-AB	326.086	326.986
OCW	0	0
Totaal inkomsten	826.986	826.986

Uitgaven

reeds toegekend NWO IMIX

betalingen projecten 1e oproep

betalingen projecten 2^e oproep - tender 100.000 100.000

betalingen projecten 2e oproep - open call 862.700 603.601

betalingen projecten 3^e oproep - tender 318.000 0

betalingen projecten 3e oproep - open call 1.000.000 690.263

betalingen demonstratieprojecten 94.366 52.529

uitgaven flankerende activiteiten 200.000 163.942

uitgaven onderhoud en beheer 40.000 40.000

uitgaven Programmabureau 100.000 25.000

NTU coördinatorkosten 0 0

totaal uitgaven	2.715.066	1.675.355
------------------------	------------------	------------------

Rente-inkomsten	26.112	55.178
------------------------	---------------	---------------

Nadere toelichting Rente-inkomsten

De rente-inkomsten vallen bijna ruim 2x zo hoog uit dan begroot: € 55.178 in plaats van € 26.112 met name omdat betalingen later worden gedaan dan begroot. Gezien het feit dat projecten later bedragen opvragen of later dan gepland worden afgesloten, kan men verwachten dat ook in de komende periode nog behoorlijke rente-inkomsten ingeboekt kunnen worden. De rente-inkomsten zijn berekend op dagbasis à rato van 2,5% (tot 31/12/2010).

Nadere toelichting uitgaven

Toelichting uitgaven O&O projecten en demonstratieprojecten

Van de projectkosten (O&O projecten, tenderprojecten, demonstratieprojecten) is reeds k€ 7.814 van de in de verschillende rondes gecommiteerde k€ 9.668 uitbetaald. Voor de begroting en de realisatie 2010 zie tabel hieronder.

Betaling/Committing	Herziene begroting d.d. maart 2010 (€)	Realisatie 2010 (€)
projecten 1e oproep		
projecten 2e oproep - tenders	100.000	100.000
projecten 2e oproep - open call	862.700	603.601
projecten 3e oproep - tender (SoNaR)	318.000	0
projecten 3e oproep - open call	1.000.000	690.263
demonstratieprojecten	94.366	52.529

Toelichting uitgaven Onderhoud en beheer, Programmabureau en NTU coördinatorkosten

Uitgaven	Herziene begroting 2010 d.d. maart 2010 (€)	Realisatie 2010 (€)
uitgaven onderhoud en beheer	40.000	40.000
uitgaven Programmabureau	100.000	25.000
NTU coördinatorkosten	0	0

De geplande uitgaven voor onderhoud en beheer (TST-Centrale) komen overeen met het daarvoor begrote bedrag. NWO heeft de lumpsum toegekend voor de uitvoering van het PB nu volledig gedeclareerd. Het Agentschap moet nog 75k€ declareren. NB het PB werkt net als de projecten op basis van resultaat, er worden geen werkelijke kosten gedeclareerd.

Toelichting uitgaven Flankerende activiteiten

De uitgaven in het kader van het flankerend beleid zijn dit jaar wederom lager uitgevallen dan begroot. Naast die hieronder vermelde daadwerkelijk uitgaven is in 2010 ook een aantal nieuwe posten gecommiteerd (een nadere specificatie is verderop bij de verschillende subbudgetten te vinden).

Specificatie uitgaven flankerende activiteiten 2010

Activiteit	Herziene* begroting 2010-2011(€)	Realisatie 2010 (€)	soort uitgave
Vergader- en kantoorkosten betaald door NTU	40.000	11.528	eigen activiteiten
STEVIN project TST op wikipedia	2.000	0	eigen activiteiten
IPR workshops PC	20.000	0	eigen activiteiten
Roadmap workshops WGFA	14.000	0	eigen activiteiten
STEVIN programmadag 2010	5.000	6.069	eigen activiteiten
Eindevenement	27.500	0	eigen activiteiten
Netwerksubsidies	15.000	7.150	netwerksubsidies
Kosten eindmeting/evaluatie	60.000	59.450	nulmeting en evaluatie
Kosten WIKI 2010/11/12	6.000	4.000	interne communicatie
Externe Communicatie	47.297	6.000	externe communicatie
Educatieve activiteiten	0	69.028	educa activiteitem
Totaal uitbetaald	236.797	163.225	

* Deze herziene begroting is in maart 2010 goedgekeurd door het TST-bestuur.

Nadere toelichting betalingen uit budget eigen activiteiten

Voor kantoor- en vergaderkosten NTU werd in het werkplan voor de periode 2010 - 2011 een bedrag van in totaal € 40.000 gereserveerd. Daarvan is ongeveer een kwart uitgegeven in 2010. Ervan uitgaande dat het aantal vergaderingen aan het eind van het programma zal dalen en dientengevolge ook de kosten die daarmee samenhangen zal het budget voor het komende jaar naar beneden bijgesteld kunnen worden. Het grootste deel van de kosten betreft vacatiegelden voor leden van het TST-bestuur, de PC, de WGFA en de Prijzencommissie als ook de vergaderkosten van het TST-bestuur en de Prijzencommissie. De vergaderkosten van de PC en de WGFA worden betaald door NWO en Agentschap NL vanuit het programmabureaubudget.

Het project TST op Wikipedia is nu ingebed in een vervolgaanvraag op het eerder gehonoreerde TST op Kennislink project (zie overzicht committeringen gedaan in 2009 die nog niet tot uitbetaling hebben geleid. De begrote kosten zijn daarom teruggebracht tot € 2.000. In tegenstelling tot de planning hebben dit jaar geen IPR-workshops of Roadmap-workshops plaatsgevonden. Voor de organisatie van die workshops die nu in 2011 moeten plaatsvinden zijn wel specifiek verantwoordelijke PC-leden aangewezen die de organisatie in samenwerking met het Programmabureau zullen oppakken. De kosten voor de interne STEVIN programmadag hebben dit jaar het vorig jaar naar beneden bijgestelde bedrag overschreden. Dat is met name veroorzaakt door de niet begrote kosten van de buitenlandse gastsprekers.

Nadere toelichting betalingen uit budget netwerksubsidies

In 2010 zijn de volgende netwerksubsidies toegekend (en uitbetaald):

- Netwerksubsidie CLIN 20, 2010 € 1.500 (reeds toegekend in 2009)
- Netwerksubsidie DIR 2010 € 500 (reeds toegekend in 2009)
- Netwerksubsidie TABU-dag 2010 € 1.400
- Netwerksubsidie Gridline TST-stand op Overheid en ICT € 2.500
- Netwerksubsidie LOT 2010 € 750

In 2010 zijn nog 3 nieuwe netwerksubsidies toegekend die waren ingediend in de continue openstaande oproep. Deze activiteiten zullen pas in 2011 plaatsvinden en tot uitbetaling leiden.

- Netwerksubsidie CLIN 21, 2011 € 2.000
- Netwerksubsidie EAMT 2011 € 1.500
- Netwerksubsidie DIR 2011 € 600

Nadere toelichting betalingen uit budget nulmeting en evaluatie

De opdracht aan een consultant die de externe evaluatie/eindmeting kan uitvoeren is in 2009 opgesteld en uiteindelijk toegekend aan Technopolis die het werk in de loop van 2010 heeft uitgevoerd. Zowel het voorschot als de eindbetaling hebben in 2010 plaatsgevonden.

Nadere toelichting betalingen uit budget interne communicatie

De KU Leuven heeft € 4.000 ontvangen om de STEVIN wiki-pagina nog tot in 2012 te hosten. Door een vergissing van het PB was er ten onrechte € 2.000 te veel begroot.

Nadere toelichting betalingen uit budget externe communicatie: DIXIT

In 2010 is alleen de STEVIN-bijdrage (€ 6.000) uitbetaald voor de nieuwe DIXIT die eind 2009 was verschenen.

Nadere toelichting betalingen uit budget educatieve activiteiten

In 2010 heeft de eindbetaling plaatsgevonden van de educaprojecten Kennislink en Diademo en van de masterclasses ICT en Dyslexie en TST in Nederlandstalige overheidsdiensten.

- Educaronde project Kennislink (toegekend in 2007) € 27.500
- Educaronde project Diademo (toegekend in 2008) € 8.028
- Masterclass ICT en Dyslexie (toegekend in 2008) € 17.500
- Masterclass TST NI talige overheidsdiensten (toegekend in 2009) € 16.000

Overzicht publicaties en activiteiten STEVIN 2010 en 1^e helft 2011

Een cumulatief overzicht is opgenomen in de STEVIN FactFile en op de STEVIN website).

In tijdschriften

1. [DAESO] Erwin Marsi and Emiel Kraemer (2010), Detecting semantic similarity using syntactic tree alignment, in preparation, submitted to *Computational Linguistics* journal.
2. [DAESO] Erwin Marsi and Emiel Kraemer (2010), Buidling a parallel monolingual treebank for semantic overlap detection, in preparation, submitted to *Language Resources and Evaluation* journal.
3. [DAISY] Saint Dizier, Patrick & Moens, Marie-Francine, Knowledge and Reasoning for Question Answering: Research Perspectives. *Information Processing & Management* (in press). 2011
4. [DAISY] Kolomiyets, Oleksandr and Moens, Marie-Francine A Survey on Question Answering Technology from an Information Retrieval Perspective. *Information Sciences* (in press). 2011
5. [DAISY] Moens, Marie-Francine & Saint Dizier, Patrick Introduction to the Special Issue on Question Answering. *Information Processing & Management* (2011), (in press).
6. [DISCO] J. van Doremalen, C. Cucchiarini, H. Strik (to appear), Optimizing automatic speech recognition for low-proficient non-native speakers. *EURASIP Journal on Audio, Speech, and Music Processing*. [PDF]
7. [DPC] Macken L., De Clercq O., Paulussen H. (2011), Dutch Parallel Corpus: a Balanced Copyright-Cleared Parallel Corpus, *META* 56 (2)
8. [DUOMAN] Weerkamp W., Balog K., de Rijke M. (2010), Blog Feed Search with a Post Index, *Information Retrieval Journal*, submitted.
9. [DUOMAN] Meij E J., Trieschnigg D., de Rijke M., Kraaij W. (2010), Conceptual Language Models for Domain-specific Retrieval, *Information Processing & Management*, to appear.
10. [DUOMAN] Tsagkias E., Larson M., de Rijke M. (2010), Predicting Podcast Preference: An Analysis Framework and its Application, *Journal of the American Society for Information Science and Technology*, 61 (2), 2010, in press.
11. [DUOMAN] Hofmann K., Balog K., Bogers T., de Rijke M. (2010), Contextual Factors for Finding Similar Experts, *Journal of the American Society for Information Science and Technology*, in press.
12. [DUOMAN] Hofmann K., Balog K., Bogers T., de Rijke M. (2010), Contextual Factors for Finding Similar Experts, *Journal of the American Society for Information Science and Technology*.
13. [DUOMAN] Meij E., Trieschnigg D., de Rijke M., Kraaij W. (2010), Conceptual Language Models for Domain-specific Retrieval, *Information Processing and Management*.
14. [DUOMAN] Tsagkias E., Larson M., de Rijke M. (2010), Predicting Podcast Preference: An Analysis Framework and its Application, *Journal of the American Society for Information Science and Technology*.
15. [IRME] Grégoire, N. (2010), 'DuELME: A Dutch Electronic Lexicon of Multiword Expressions', *Journal of Language Resources and Evaluation*, special issue on Multiword Expressions 44(1/2), 23-40. DOI 10.1007/s10579-009-9094-z
16. [MIDAS] Gemmeke, J., H. Van hamme, B. Cranen, L. Boves (2010), Compressive Sensing for Missing Data Imputation in Noise Robust Speech Recognition. *IEEE-Journal of selected topics in Signal Processing*, pp 272 - 287.
17. [MIDAS] Jort F. Gemmeke, Tuomas Virtanen, Antti Hurmalainen (2011), "Exemplar-based sparse representations for noise robust automatic speech recognition", accepted for publication in *IEEE Transactions on Audio, Speech and Language processing*.
18. [MIDAS] Jort Florent Gemmeke, Bert Cranen, Ulpu Remes (2011), "Sparse imputation for large vocabulary noise robust ASR", in *Computer Speech & Language*, vol. 25, no. 2, pp 462-479.
19. [MIDAS] Gemmeke J., Cranen B., Remes U., (2011) (2011), Sparse imputation for large vocabulary noise robust ASR, *Computer Speech & Language*, 25 (2): 462-479.
20. [MIDAS] Huang H., Liu Y., Gemmeke J., ten Bosch L., Cranen B., Boves L. (2011), Globality-Localy Consistent Discriminant Analysis for Phone Classification, *Proceedings of the 12th Annual Conference of the International Speech Communication Association (Interspeech 2011)* , International Speech Communication Association.
21. [MIDAS] Huang H., ten Bosch L., Gemmeke J., Cranen B., Boves L. (2011), Exploiting local and global structures for timit phone classification, *Proceedings of the 19th European Signal Processing Conference (EUSIPCO 2011)*, pp. 1485-1489.
22. [MIDAS] Hurmalainen A., Jort Gemmeke J., Tuomas Virtanen T. (2011), Non-negative matrix deconvolution in noise robust speech recognition, *Proceedings of the International Conference on Audio, Speech and Signal Processing (ICASSP 2011)*.

In conference proceedings

1. [AUTONOMATA TOO] Reveil, B., Martens, J-P., D'hoore, B. (2009): "How speaker tongue and name source language affect the automatic recognition of spoken names", in Proceedings Interspeech2009, Brighton, UK: [Paper \(.pdf\)](#)
2. [AUTONOMATA TOO] Heuvel, H. van den, Reveil, B., Martens, J-P., D'hoore, B. (2009): "Pronunciation-based ASR for names", in Proceedings Interspeech2009, Brighton, UK: [Paper \(.pdf\)](#)
3. [AUTONOMATA TOO] Schraagen, M., Bloothoof, G. (2010), "Evaluating repetitions, or how to improve your multilingual ASRsystem by doing nothing", in Proceedings LREC2010, Malta: [Paper \(.pdf\)](#)
4. [AUTONOMATA TOO] Reveil, B., Martens, J-P., Van den Heuvel, H. (2010), "Improving proper name recognition by adding automatically learned pronunciation variants to the lexicon", in Proceedings LREC2010, Malta: [Paper \(.pdf\)](#)
5. [CORNETTO] Valentin Jijkoun and Katja Hofmann (2009), Generating a Non-English Subjectivity Lexicon: Relations That Matter. Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009). PDF
6. [CORNETTO] Erik Tjong Kim Sang and Katja Hofmann (2009), Lexical Patterns or Dependency Patterns: Which Is Better for Hypernym Extraction?. Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009).
7. [DAESO] Edgar Meij and Maarten de Rijke (2010), Detecting novel contributions to news items, in preparation, to be submitted to *International Conference on Knowledge Discovery and Information Retrieval*.
8. [DAESO] Erwin Marsi and Emiel Krahmer (2010c), Automatic detection of semantic similarity in comparable text through syntactic tree matching, submitted to the *48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*.
9. [DAESO] Erwin Marsi and Emiel Krahmer (2010), Automatic analysis of semantic similarity in comparable text through syntactic tree matching. Proceedings of COLING 2010, Beijing. [[pdf](#)]
10. [DAESO] Balog K., Bron M., de Rijke M. (2010), Category-based Query Modeling for Entity Search, *Proceedings of the 32nd European Conference on Information Retrieval (ECIR 2010)*, Springer, March, 2010.
11. [DAESO] Tsagkias E., Weerkamp W., de Rijke M. (2010), News Comments: Exploring, Modeling, and Online Predicting, *Proceedings of the 32nd European Conference on Information Retrieval (ECIR 2010)*, Springer, March, 2010.
12. [DAISY] De Belder, Jan & Moens (2010), Marie-Francine Integer Linear Programming for Dutch Sentence Compression. In A. Gelbukh (Ed.). *CICLing 2010 (Lecture Notes in Computer Science 6008)* (pp. 711-723). Berlin: Springer. (Best Student Paper Award)
13. [DAISY] De Belder J., Deschacht K., Moens M.-F. (2010), Lexical simplification, Proceedings of Itec2010, 1st International Conference on Interdisciplinary Research on Technology, Education and Communication.
14. [DAISY] Daniël de Kok (2010), Feature Selection for Fluency Ranking, in: *Proceedings of the 6th International Conference on Natural Language Generation (INLG 2010)*, Trim, Ireland.
15. [DISCO] Van Doremalen J., Cucchiarini C., Strik H. (2010), Using Non-Native Error Patterns to Improve Pronunciation Verification, Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech 2010), International Speech Communication Association.
16. [DUOMAN] Bron M., Balog K. and de Rijke M. (2010), Ranking Related Entities: Components and Analyses, Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10).
17. [DUOMAN] Bron M., He J., Hofmann K., Meij E J., de Rijke M., Tsagkias E. and Weerkamp W. (2010), The University of Amsterdam at TREC 2010: Session, Entity, and Relevance Feedback, TREC 2010 Working Notes, NIST.
18. [DUOMAN] de Rijke M., Balog K., Bogers T. and van den Bosch A. (2010), On the Evaluation of Entity Profiles, Proceedings of the Conference on Multilingual and Multimodal Information Access Evaluation (CLEF 2010).
19. [DUOMAN] He J., Balog K., Hofmann K., Meij E., de Rijke M., Tsagkias E., Weerkamp W. (2010), Heuristic Ranking and Diversification of Web Documents, Proceedings of the Eighteenth Text REtrieval Conference (TREC 2009).
20. [DUOMAN] Jijkoun V B., de Rijke M., Weerkamp W. (2010), Generating Focused Topic-specific Sentiment Lexicons, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010), ACL, pp. 585-594.

21. [DUOMAN] Meij E., He J., Weerkamp W., de Rijke M. (2010), Topical Diversity and Relevance Feedback, Proceedings of the Eighteenth Text REtrieval Conference (TREC 2009).
22. [DUOMAN] Tsagkias E., Weerkamp W., de Rijke M. (2010), News Comments: Exploring, Modeling, and Online Predicting, Proceedings of the 2nd European Conference on Information Retrieval (ECIR 2010), Springer, pp. 109-203.
23. [DUOMAN] Tsagkias E., de Rijke M., Weerkamp W. (2011), Linking Online News and Social Media, Proceedings of the Fourth ACM Web Search and Data Mining (WSDM), ACM.
24. [LASSY] Erik Tjong Kim Sang and Katja Hofmann (2010), Lexical Patterns or Dependency Patterns: Which Is Better for Hypernym Extraction? In: Proceedings of CoNLL-2009, Boulder, CO, USA, June 2009. [pdf]
25. [MIDAS] Kris Demuynck, Xueru Zhang, Dirk Van Compernelle, Hugo Van hamme (2010), "Feature versus Model Based Noise Robustness", *Interspeech 2010*.
26. [MIDAS] Wambacq P., Demuynck K. (2011), Efficiency of Speech Alignment for Semi-Automated Subtitling in Dutch, Proceedings of the 14th International Conference on Text, Speech and Dialogue (TSD).
27. [MIDAS] Jort Florent Gemmeke and Tuomas Virtanen (2010), "Artificial and online noise dictionaries for noise robust speech recognition", *Interspeech 2010*.
28. [MIDAS] Jort F. Gemmeke, Ulpu Remes, Kalle J. Palomäki (2010), "Observation uncertainty measures for sparse imputation", *Interspeech 2010*.
29. [MIDAS] Jort Gemmeke and Tuomas Virtanen (2010), "Noise robust exemplar-based connected digit recognition," Proceedings of ICASSP 2010, Dallas, USA
30. [MIDAS] Yang Sun, Jort Florent Gemmeke, Bert Cranen, Louis ten Bosch, Lou Boves (2010), "Using a DBN to integrate sparse classification and GMM-based ASR", *Interspeech 2010*.
31. [MIDAS] Virtanen T., Gemmeke J., Hurmalainen A. (2010), State-based labelling for a sparse representation of speech and its application to robust speech recognition, Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech 2010).
32. [MIDAS] J. F. Gemmeke, A. Hurmalainen, T. Virtanen and Y. Sun (2010), "Toward a practical implementation of exemplar-based noise robust ASR," *in Proc. EUSIPCO*, 2011, pp. 1490-1494.
33. [MIDAS] H. Huang, Y. Liu, J. F. Gemmeke, L. ten Bosch, B. Cranen and L. Boves (2011), "Globality-Locality Consistent Discriminant Analysis for Phone Classification," *in Proc. Interspeech 2011*.
34. [MIDAS] H. Huang, L. ten Bosch, J. F. Gemmeke, B. Cranen and L. Boves (2011), "Exploiting local and global structures for timit phone classification," *in Proc. EUSIPCO 2011*, pp. 1485-1489.
35. [MIDAS] Antti Hurmalainen, Jort F. Gemmeke, Tuomas Virtanen (2011), "Non-negative matrix deconvolution in noise robust speech recognition", *ICASSP 2011*.
36. [MIDAS] H. Kallasjoki, U. Remes, J. F. Gemmeke, T. Virtanen and K. J. Palomäki (2011), "Uncertainty measures for improving exemplar-based source separation," *in Proc. Interspeech 2011*.
37. [MIDAS] K. Mahkonen, A. Hurmalainen, T. Virtanen and J. F. Gemmeke (2011), "Mapping Sparse Representation to State Likelihoods in Noise-Robust Automatic Speech Recognition," *in Proc. Interspeech 2011*.
38. [MIDAS] Y. Sun, J. F. Gemmeke, B. Cranen, L. ten Bosch and L. Boves (2011), "Improvements of a dual-input DBN for noise robust ASR," *in Proc. Interspeech 2011*.
39. [MIDAS] Y. Sun, J. F. Gemmeke, B. Cranen, L. ten Bosch and L. Boves (2011), "Early fusion of sparse classification and GMM for noise robust ASR," *in Proc. EUSIPCO 2011*, pp. 1495-1499.
40. [PaCoMT] Vandeghinste V. and Martens S. (2010), Bottom-up transfer in Example-based Machine Translation, Proceedings of the 14th Annual conference of the European Association for Machine Translation Proceedings (EAMT 2010), European Association for Machine Translation.
41. [SONAR] Reynaert M., Oostdijk N., De Clercq O., van den Heuvel H. and de Jong F. (2010), Balancing SoNaR: IPR versus Processing Issues in a 500-Million-Word Written Dutch Reference Corpus, Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010).

In workshop proceedings

1. [DAESO] Martijn Spitters, Remko Bonnema, Mihai Rotaru, Jakub Zavrel (2010), A System for Large-Scale Detection of Semantic Overlap in Taxonomies (provisional title), to be submitted to the ACL workshop on Geometrical Models of Natural Language Semantics (ACL GEMS 2010).
2. [DAESO] Martijn Spitters, Remko Bonnema, Mihai Rotaru, and Jakub Zavrel, Bootstrapping information extraction mappings by similarity-based reuse of taxonomies. Reuse and Adaptation of Ontologies and Terminologies Workshop, EKAW, Lisbon, Portugal, 2010. [pdf]
3. [DAISY] Arias, Javier, Deschacht, Koen & Moens, Marie-Francine (2010), Content Extraction from Multilingual Web Pages. In Aly R., Hauff, C. Hiemstra, D., Huybers, T. & De Jong, F. (Eds.), *Proceedings of the 9th Dutch-Belgium Information Retrieval Workshop*. University of Twente 2009.

4. [DAISY] Daniël de Kok and Gertjan van Noord, A sentence generator for Dutch (2010), in: Proceedings of the 20th Meeting of Computational Linguistics in the Netherlands.
5. [DAISY] De Belder, Jan & Moens, Marie-Francine (2010), Sentence Compression for Dutch Using Integer Linear Programming. In *Proceedings of DIR 2010 10th Dutch-Belgian Information Retrieval Workshop* (pp. 16-21). Radboud Universiteit Nijmegen.
6. [DISCO] Cucchiarini C., van Doremalen J., Strik H. (2010), Fluency in non-native read and spontaneous speech, Proceedings of the Joint Workshop of the 5th Workshop on Disfluency in Spontaneous Speech and The 2nd International Symposium on Linguistic Patterns in Spontaneous Speech (DiSS-LPSS).
7. [DISCO] van Doremalen J., Cucchiarini C., Strik H. (2010), Phoneme Errors in Read and Spontaneous Non-Native Speech: Relevance for CAPT System Development, Proceedings of the Speech and Language Technology in Education workshop (SLaTE-2010), International Speech Communication Association
8. [DISCO] Strik H., Colpaert J., van Doremalen J. and Cucchiarini C. (2010), Language resources and CALL applications: speech data and speech technology in the DISCO project, Proceedings of the first LREC Workshop on Supporting eLearning with Language Resources and Semantic Data, pp. 1-6.
9. [DISCO] Strik H., van de Loo J., van Doremalen J., Cucchiarini C. (2010), Practicing syntax in spoken interaction: Automatic detection of syntactical errors in non-native utterances, Proceedings of the Speech and Language Technology in Education.
10. [DPC] Paulussen Hans & Lieve Macken (2010), "Annotating the Dutch Parallel Corpus". In: Proceedings of the Workshop on Annotation and Exploitation of Parallel Corpora AEPC 2010. Editors: Lars Ahrenberg, Jörg Tiedemann and Martin Volk. NEALT Proceedings Series, Vol. 10 (2010), 63-72.
11. [DUOMAN] Balog K., Meij E J., de Rijke M. (2010), Entity Search: Building Bridges between Two Worlds, *Semantic Search 2010*, submitted.
12. [DUOMAN] Balog K., Bron M., de Rijke M. and Weerkamp W. (2010), Combining Term-based and Category-based Representations for Entity Search, Proceedings of the Focused Retrieval and Evaluation. 8th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX 2009).
13. [DUOMAN] He J. and de Rijke M. (2010), An Exploration of Learning to Link with Wikipedia: Features, Methods and Training Collection, Proceedings of the Focused Retrieval and Evaluation. 8th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX 2009).
14. [DUOMAN] Jijkoun V., de Rijke M., Weerkamp W., Ackermans P., Geleijnse G. (2010), Mining User Experiences from Online Forums: An Exploration, Proceedings of the NAACL HLT 2010 workshop on Computational Linguistics in a World of Social Media, ACL, pp. 17-18.
15. [DUOMAN] Weerkamp W., Balog K., de Rijke M. (2010), A Two-Stage Model for Blog Feed Search, Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval (SIGIR 2010), ACM, pp. 877-878.
16. [DUOMAN] Weerkamp W., Tsagkias E. and de Rijke M. (2010), From Blogs to News: Identifying Hot Topics in the Blogosphere, Proceedings of the Eighteenth Text REtrieval Conference (TREC 2009).
17. [LASSY] Martijn Spitters, Remko Bonnema, Mihai Rotaru, Jakub Zavrel (2010), A System For Large-Scale Bootstrapping of Synonyms in Taxonomies. Presented at Computational Linguistics In The Netherlands (CLIN), Utrecht, The Netherlands, February 5 2010. [[pdf](#)]
18. [LASSY] Erwin Marsi and Emiel Kraemer (2010), Detecting semantic similarity using syntactic tree alignment. Presented at Computational Linguistics In The Netherlands (CLIN), Utrecht, The Netherlands, February 5 2010. [[pdf](#)]
19. [LASSY] Iris Hendrickx, Redundancy reduction for multi-document summarization (2010, Presented at NLX group, Faculty of sciences, University of Lisbon, Lisbon, 11 December 2009. [[pdf](#)]
20. [LASSY] Paulussen Hans & Lieve Macken (2010), "Annotating the Dutch Parallel Corpus". In: *Proceedings of the Workshop on Annotation and Exploitation of Parallel Corpora AEPC 2010*. Editors: Lars Ahrenberg, Jörg Tiedemann and Martin Volk. NEALT Proceedings Series, Vol. 10 (2010), 63-72.
21. [MIDAS] Wang Y., Van hamme H. (2011), Gaussian Selection Using Self-Organizing Map for Automatic Speech Recognition, Proceedings of the Workshop on Self-Organizing Maps "Advances in Self-Organizing Maps", Lecture Notes in Computer Science 6731, Springer, pp. 218-221.
22. [PaCoMT] Martens, Scott, and Vincent Vandeghinste (2010), An Efficient, Generic Approach to Extracting Multi-Word Expressions from Dependency Trees. In Proceedings of the CoLing Workshop: Multiword Expressions: From Theory to Applications (MWE 2010). ([pdf](#))
23. [PaCoMT] Vandeghinste, V., and Martens, S. (2010), Bottom-up transfer in Example-based Machine Translation. In Proceedings of EAMT 2010. European Association for Machine Translation. Saint-Raphael ([pdf](#))
24. [PaCoMT] Tiedemann, J. & Kotzé, G. (2009), Building a Large Machine-Aligned Parallel Treebank. Proceedings of TLT-8. ([pdf](#))

- [PaCoMT] Vandeghinste, V. & Martens, S. (2009), Top-down Transfer in Example-based MT. Proceedings of 3rd Workshop on EBMT. pp. 69-76. (pdf)

Manuals

- [SONAR] De Clercq O. and Reynaert M. (2010), SoNaR Acquisition Manual, LT3 Technical Report LT3 10-02, Hogeschool Gent, Gent.
- [DPC] Piet Desmet, Willy Vandeweghe, Hans Paulussen, Lieve Macken, Maribel Montero Perez, Orphée De Clercq, Lidia Rura, Julia Trushkina and Antoine Besnehard (2010), Dutch Parallel Corpus User Manual.
- [LASSY] Gertjan van Noord, Ineke Schuurman, Gosse Bouma (2011), Lassy Syntactische Annotatie, Revision : 19077.
- [LASSY] Gertjan van Noord (2011), User Guide: How to use Alpino (update).
- [SPRAAK] Update SPRAAK V1.0 documentation, (2011).

Book editing

- [LASSY] Van Eynde, F., A. Frank, K. de Smedt, G. van Noord (eds) (2010), Proceedings of the Seventh International Workshop on Treebanks and Linguistic Theories (TLT 7). January 23-24, 2009, Groningen, The Netherlands. LOT Occasional Series

Book contributions

- [CORNETTO] Vossen P. (2010), WordNet: principles, developments and applications, Dictionaries, an International Encyclopedia of Lexicography, Supplementary volume: Recent developments with special focus on computational lexicography, Handbook of Linguistics and Communications (HSK), Mouton de Gruyter, Berlin.
- [CORNETTO] Vossen P., Fellbaum C. (in press), "Universals and Idiosyncracies in Multilingual WordNets", in: Handbook Multilingual Lexicography, Oxford University Press.
- [DAESO] Erwin Marsi, Emiel Kraemer, Iris Hendrickx, Walter Daelemans (2010), On the limits of sentence compression by deletion, In: Empirical Methods in Natural Language Generation. Lecture Notes in Computer Science 5980, Springer, Berlin Heidelberg, to appear.
- [LASSY] Van Noord G. (2010), Self-trained Bilexical Preferences to Improve Disambiguation Accuracy, In Bunt H., Merlo P. and Nivre J., (eds.), Trends in Parsing Technology, Springer, pp. 183-200.
- [MIDAS] Wang Y., Van Segbroeck M., Van hamme H., (2010), Robust Large Vocabulary Continuous Speech Recognition Based on Missing Feature Techniques, In Ramirez J. and Górriz J., (eds.), Recent Advances in Robust Speech Recognition Technology, Bentham Science Publishers, pp. 141-154.
- [MIDAS] Gemmeke J., Van Segbroeck M., Wang Y., Cranen B., Van hamme H. (2011), Using missing data techniques: Handling of real-world data, In Kolossa D., Haeb-Umbach R., (eds.), Robust Speech Recognition of Uncertain or Missing Data, Springer, chapter 7.
- [MIDAS] Gemmeke J., Remes U. (2011), Missing data techniques: Feature reconstruction, In Virtanen T., Rita Singh R., Raj B., (eds.), Techniques for Noise Robustness in Automatic Speech Recognition, Wiley.

Books

- [MIDAS] Gemmeke J. (2011), Noise Robust ASR: Missing data techniques and beyond, (Ph.D. thesis), Nijmegen.

DEMOS

- [CORNETTO] Erik Tjong Kim Sang (2011), Cornetto Dutch Set Demo
- [CORNETTO] Erik Tjong Kim Sang, (2011) Cornetto Lexical Search Demo
- [LASSY] Erik Tjong Kim Sang (2010), LASSY Search Demo

Activiteiten georganiseerd/medegefinancierd door het STEVIN-programma in 2010 en in de 1^e helft van 2011

25 januari 2010	Nijmegen	10th Dutch-Belgian Information Retrieval Workshop (DIR 2010)
5 februari 2010	Utrecht	20e Computational Linguistics in the Netherlands workshop (CLIN 20)
27-29 april	Utrecht	Gridline TST-stand op Overheid en ICT beurs
3-4 juni 2010	Groningen	TABU-dag 2010
14-25 juni 2010	Nijmegen	LOT zomerschool 2010
14 oktober 2010	Antwerpen	STEVIN programmadag
4 november 2010	Breukelen	Masterclass TST in Nederlandstalige overheidsdiensten
15 november 2010	Grave	Masterclass ICT en Dyslexie
11 februari 2011	Gent	21e Computational Linguistics in the Netherlands workshop (CLIN 21)
30-31 mei 2011	Leuven	15th Annual Conference of the European Association for Machine Translation (EAMT2011)
4 februari 2011	Amsterdam	11th Dutch-Belgian Information Retrieval Workshop (DIR 2011)