

Help, ik wil Wikipedia veranderen

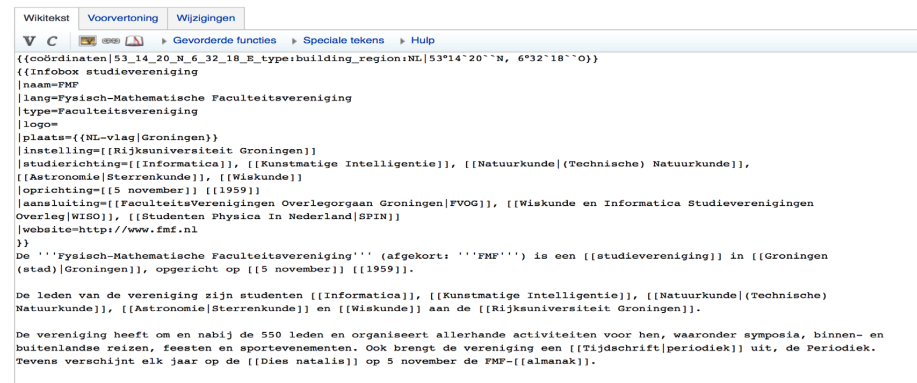
DOOR JAN PAUL POSMA

Wikipedia heeft een probleem, het aantal actieve vrijwilligers neemt al enkele jaren langzaam af en nieuwe vrijwilligers gaan steeds vaker al na korte tijd weg [3]. Hier zijn verschillende redenen voor, en een van de belangrijkste is dat het steeds moeilijker is geworden Wikipedia te bewerken. Wanneer je op de “Bewerk” knop klikt op een willekeurige pagina van Wikipedia, dan verwacht je een interface te zien zoals Microsoft Word of Google Docs. Precies dezelfde pagina als voorheen, maar dan met een tekstcursor en een toolbar voor het opmaken van tekst en het invoegen van links en media. Maar niets is minder waar. Je krijgt een groot tekstvak voorgeschoteld met daarin tekst die ons bèta’s nog het meest doet denken aan een programmeertaal.

Voor mijn bachelorproject heb ik onderzocht wat een goede en snelle manier zou kunnen zijn om het bewerken van Wikipedia makkelijker te maken. Waarom gebruikt Wikipedia niet allang een visuele editor zoals bijvoorbeeld Google Docs? Dit blijkt alles te maken te hebben met *wikitext*, de taal die je voor je neus krijgt als je een pagina bewerkt. Wikitext is de taal waarin wikipagina’s,

pagina’s op bijvoorbeeld Wikipedia, worden opgeslagen en bewerkt. Zodra een pagina wordt opgeslagen, wordt deze wikitext omgezet in HTML, de taal die een browser gebruikt om een pagina te tonen.

Vlak na de internetbubbel maar jaren voor men van Web 2.0 sprak, in 2001, werd Wikipedia opgericht. Toen was het een logische keuze om een taal als wiki-



```

Wikitext  Voorvertoning  Wijzigingen
V C  → Gevorderde functies  → Speciale tekens  → Hulp
{{coördinaten|53_14_20_N_6_32_18_E_type:building_region:NL|53°14'20"N, 6°32'18"O}}
{{Infobox studievereniging
|name=FMF
|lang=Fysisch-Mathematische Faculteitsvereniging
|type=Faculteitsvereniging
|logo=
|plaats={{NL-vlag|Groningen}}
|instelling=[[Rijksuniversiteit Groningen]]
|studierichting=[[Informatica]], [[Kunstmatige Intelligentie]], [[Natuurkunde|(Technische) Natuurkunde]],
[[Astronomie|Sterrenkunde]], [[Wiskunde]]
|oprichting=[[5 november]] [[1959]]
|aansluiting=[[Faculteitsverenigingen Overlegorgaan Groningen|FVOG]], [[Wiskunde en Informatica Studieverenigingen
Overleg|WISO]], [[Studenten Physica In Nederland|SPIN]]
|website=http://www.fmf.nl
}}
De ''Fysisch-Mathematische Faculteitsvereniging'' (afgekort: ''FMF'') is een [[studievereniging]] in [[Groningen
stad|Groningen]], opgericht op [[5 november]] [[1959]].

De leden van de vereniging zijn studenten [[Informatica]], [[Kunstmatige Intelligentie]], [[Natuurkunde|(Technische)
Natuurkunde]], [[Astronomie|Sterrenkunde]] en [[Wiskunde]] aan de [[Rijksuniversiteit Groningen]].

De vereniging heeft om en nabij de 550 leden en organiseert allerhande activiteiten voor hen, waaronder symposia, binnen- en
buitenlandse reizen, feesten en sportevenementen. Ook brengt de vereniging een [[Tijdschrift|periodiek]] uit, de Periodiek.
Tevens verschijnt elk jaar op de [[Dies natalis]] op 5 november de FMF-[[almanak]].

```

FIGUUR 1 Wikitext zoals een editor van wikipedia er nu nog mee moet werken. Je kunt allerlei complexe codes zien die gebruikt worden voor het opbouwen van complexere elementen van een wiki-pagina. Een informatievak heeft bijvoorbeeld de code Infobox.

tekst te gebruiken. Microsoft Word bestond uiteraard al wel, maar een online variant zoals Google Docs nog niet. De technieken en standaarden waren simpelweg nog niet ontwikkeld. Browsers kenden al wel grote invoervelden en dus werden die gebruikt om de pagina te bewerken. Om de tekst toch nog wat vorm te kunnen geven, werden codes verzonden, zoals “vetgedrukt”, “cursief”, [[link]] en ==koptekst==. Deze werden vervolgens omgezet naar HTML door middel van substitutie, een algoritme verving bijvoorbeeld eerst elke “ door en , de HTML tags die vetgedrukte tekst markeren. Vervolgens werd ook cursief gedrukte tekst vervangen, eveneens als links en alle andere elementen van zo’n wikipagina.

Omdat Wikipedia gestaag groeide werd het nodig nieuwe codes te verzinnen voor veelvoorkomende taken, zoals het bijhouden van een lijst met referenties en het plaatsen van zogenaamde informatievakken, de vakken informatie die rechtsboven op pagina’s staan. Voor de vrijwilligers van Wikipedia waren dit onmisbare toevoegingen, maar het gevolg was wel dat de wikitext-taal steeds complexer werd.

Om de drempel lager te maken voor potentiële vrijwilligers, is het nodig een interface aan te bieden die algemeen bekend is, zoals een visuele editor. Aan de andere kant kan wikitext ook niet helemaal afgedankt worden, er zijn immers tienduizenden actieve vrijwilligers op Wikipedia en andere sites die hiermee gewend zijn te werken. Een betere oplossing is dat pagina’s nog steeds op worden geslagen in wikitext, maar dat de gebruiker de keuze krijgt te bewerken met een visuele editor of op de traditionele manier in wikitext.

Het bouwen van een visuele editor bovenop wikitext is echter geen eenvoudige klus. Dat komt door de manier waarop wikitext wordt omgezet naar HTML, met behulp van opeenvolgende substituties. De meest gebruikte manier om een visuele editor te bouwen is namelijk met behulp van een tussenliggende representatie, een documentboom of syntaxboom. In zo’n boom wordt een pagina hiërarchisch gerepresenteerd. Een pagina bevat een aantal paragrafen, zo’n

paragraaf bevat zinnen en een zin bevat stukjes tekst, links en opgemaakte tekst. Helaas is wikitext niet zo mooi in een hiërarchisch model te vatten. Naast substituties worden namelijk ook bepaalde correcties gedaan, zodat er geen “foute” wikitext bestaat. Zo is de volgende constructie prima mogelijk: “vetgedrukte [[link]]”. Dit is equivalent aan “vetgedrukte [[link]]”, waarbij een link-element zich binnen een opmaakelement bevindt. In het eerste voorbeeld is die hiërarchie er niet: het link-element bevindt zich immers maar voor een gedeelte in het opmaak-element.

Nu is dit een relatief flauw voorbeeld, maar er zijn meer complexere gevallen die in de praktijk veel voor blijken te komen. Wikipedia werkt eraan om een dergelijke tussenrepresentatie wel mogelijk te maken, maar aangezien dit een moeilijke klus is kan dit nog een tijd duren [4].

In mijn bacheloronderzoek heb ik gekeken of andere interfaces mogelijk zijn die makkelijker in gebruik zijn, maar die niet zo moeilijk te bouwen zijn als een visuele editor. Uiteindelijk is hier een compromis uitgekomen: je krijgt wel de pagina zelf te zien, maar bewerken doe je door de wikitext van kleinere elementen aan te passen. Dit werkt als volgt. De gebruiker krijgt de pagina te zien, waarbij individuele elementen een blauwe achtergrondkleur hebben gekregen. Paragrafen hebben een eigen balkje in de kantlijn. Door de muis over zo’n element te bewegen licht deze op en na een klik verschijnt een tekstvak met de bijbehorende wikitext. Nadat deze wikitext is veranderd, gaat het tekstvak weer weg en zijn de veranderingen onmiddellijk te zien op de pagina.

Voor deze methode is het niet nodig de wikitext in een tussenstructuur om te zetten, met alle problemen van dien. Het is alleen nodig om voor een aantal elementen te weten wat de wikitext is die daarbij hoort. Hiervoor gebruik ik een truc, *annotatie*. Voordat de wikitext wordt omgezet naar HTML, wordt wat tekst toegevoegd. Paragrafen zijn bijvoorbeeld goed te herkennen, dus aan het begin en eind van een paragraaf worden zelfverzonnen codes toegevoegd. Die codes

- (a) Dit is een voorbeeldzin. Een zin met een `[[link]]`.
- (b) `<paragraaf-0>`
`<zin-0>`Dit is een voorbeeldzin.`</zin-0>`
`<zin-1>`Een zin met een `[[link]]`.`</zin-1>`
`</paragraaf-0>`
- (c) • `paragraaf-0`: [1..51]
 • `zin-0`: [1..25]
 • `zin-1`: [25..51]
- (d) `<p>`
`<paragraaf-0>`
`<zin-0>`Dit is een voorbeeldzin.`</zin-0>`
`<zin-1>`Een zin met een `[[link]]`.`</zin-1>`
`</paragraaf-0>`
`</p>`
- (e) `<p id="paragraaf-0">`
``Dit is een voorbeeldzin.``
``Een zin met een `[[link]]`.``
`</p>`

VOORBEELD 1 Annotatie tijdens het omzetten van wikitekst naar HTML: (a) de oorspronkelijke wikitekst, (b) de wikitekst na annotatie, (c) de lijst met posities, (d) de HTML na omzetting en (e) de uiteindelijke HTML.

Wikimedia Foundation

Het is ontzettend gaaf om software te schrijven die binnen enkele maanden door honderden miljoenen mensen gebruikt wordt. De Foundation is altijd op zoek naar contractors die zich als vrijwilliger bewezen hebben, dus de beste manier om te beginnen is om simpelweg wat bugs te fixen om zo bekend te raken met de software, de community en om ideeën op te doen voor grotere projecten. Ook kun je meedoen aan de Google Summer of Code, waarbij je in de zomermaanden aan een groter project werkt. Hopelijk tot ziens op een van de developers-meetings!

zijn zo gekozen dat ze na het omzetten naar HTML nog steeds ongewijzigd in de pagina te vinden zijn. Bovendien zijn alle toegevoegde codes uniek, zodat uiteindelijk de paragraaf in de HTML te koppelen is aan de oorspronkelijke wikitekst. Dit kan gedaan worden met allerlei elementen, zoals secties, zinnen, templates, referenties, etc. Dit annoteren gebeurt in voorbeeld 1.

De zelfverzonnen, geannoteerde codes worden uiteindelijk weer omgezet in echte HTML codes zodat de browser het ook begrijpt. Wanneer een gebruiker nu klikt op bijvoorbeeld de paragraaf, dan kan de oorspronkelijke wikitekst erbij gevonden worden. Immers staat voor ieder element de begin- en eindpositie van de wikitekst in een lijst. De wikitekst kan dan getoond worden en na bewerking kan de oude wikitekst met behulp van dezelfde lijst gesubstitueerd worden met de nieuwe wikitekst. Om de pagina vervolgens weer aan de gebruiker te tonen moet het hele verhaal opnieuw uitgevoerd worden. De nieuwe wikitekst krijgt een nieuwe annotatie, vervolgens wordt alles opnieuw omgezet naar HTML, etc.

Het probleem van hiërarchie is hiermee nog niet gelijk opgelost, want HTML vereist strikte hiërarchie. Het voorbeeld van boven zou dan nog steeds ongeldig zijn:

```
<zin-0>
  <vet-0>''vetgedrukte
    <link-0>[[link''</vet-0>]]</link-0>.
</zin-0>
```

Immers dient link-0 gesloten te worden voordat vet-0 gesloten mag worden. Er is echter een verbluffend makkelijke oplossing in deze gevallen: alle conflicterende elementen kunnen simpelweg verwijderd worden! Als een link soms niet afzonderlijk te bewerken valt, is dat geen probleem, omdat het dan toch mogelijk is de gehele zin te bewerken:

```
<zin-0>''vetgedrukte [[link'']].</zin-0>
```

Er is altijd wel een element op een hoger niveau

Fysisch-Mathematische Faculteitsvereniging

 53°14'20"N, 6°32'18"O (Kaart)



FIGUUR 2 De interface met blauwe balkjes om een zin, paragraaf of sectie te bewerken. De langwerpige, meest linkse blauwe balk is voor het bewerken van de hele pagina. Aan de rechterkant van de pagina zie je het resultaat van het informatievak dat in figuur 1 te zien is.

waardoor het mogelijk blijft de tekst te bewerken, het hoogste niveau is immers de gehele pagina en die kan nergens mee conflicteren. Bovendien conflicteren niveaus zoals secties en paragrafen maar zelden. Omdat de gebruiker de wikitekst uiteindelijk toch moet bewerken, kan die net zo goed soms iets langer en complexer zijn. Dat is bij een traditionele visuele editor niet mogelijk.

Je kan je afvragen of dit alles wel zin heeft als gebruikers toch wikitekst moeten bewerken, het kan nog steeds afschrikwekkend en te lastig zijn. Daarom heb ik dit onderzocht in samenwerking met GRNET, een Grieks onderzoeksinstituut, door 'echte' gebruikers in te schakelen [5]. De conclusie van dit onderzoek was grotendeels positief, gebruikers konden namelijk makkelijker aanpassingen maken aan een wikipagina. Sommige geavanceerdere opdrachten die we de gebruikers gaven bleven lastig, vooral omdat men aangeleerde technieken van visuele editors probeerde toe te passen, zoals het verslepen van zinnen en paragrafen, waar de interface niet op gebouwd is.

Ondanks de positieve resultaten zul je deze nieuwe interface voorlopig niet op Wikipedia zien, want hij is

nog niet af. In mijn scriptie worden wel de algoritmes en ideeën voor verdere ontwikkeling beschreven [5], maar omdat de Wikimedia Foundation enkele maanden geleden eindelijk is begonnen met het ontwikkelen van een echte visuele editor [4], heb ik besloten eerst af te wachten hoe dat uitpakt. Maar wie weet klik je in de toekomst nog eens op "Bewerken" en zie je de interface die ik heb ontwikkeld! •

Interessante links

- [1] SWEBLE Blog, sweble.org
- [2] Mediawiki, mediawiki.org

Referenties

- [3] March 2011 Update, Wikimedia website, strategy.wikimedia.org/wiki/March_2011_Update
- [4] Future, Wikimedia website, www.mediawiki.org/wiki/Future
- [5] Greek Research and Technology Network, WYSIWYG support for Wikipedia using the Inline Editor (internal report), 2011.
- [6] Posma, J.P., In-line Editing: a New Approach to Editing Wikis, Rijksuniversiteit Groningen