

The screenshot shows the Memorix Maior DAM interface. On the left is a sidebar with metadata fields: Archiefnummer, Inventarisnummer, Volnummer, Contributor, Coverage, Creator, Datum, Beschrijving, Formaat, Identificatiecode, Taal, Publisher, Relation, Rights, Source, Subject, Titel, and Type. The main area has a search bar and tabs for 'Voorbeeld', 'Downloads', 'Versies', 'Records', 'Creative Commons', and 'Full Text'. The 'Full Text' tab is active, showing a text window with the following content:

Robin Hood is not at home Come in and drink with Little John. (An old drinking song)  
In the first weeks and months after the fight in the abbey, Robin and his men worked very hard. They cut down young trees and built homes in the centre of the forest. They made arrows from wood and caught forest deer for their food. They bought bread and milk from the villagers. Sometimes Robin liked to leave his men, and look for adventure. One fine autumn morning, he and his friend Will Scarlet woke early. 'It is a beautiful day,' said Robin. 'I think I will go for a walk to the river. Will you come with me?' 'I want to make some new arrows for my bow,' answered Will. 'But I will meet you at lunch time near the bridge.' So Robin left Will and walked quickly through the forest to the river and the little bridge. Robin stood in the middle of the bridge, and looked up at the trees. 'The autumn colours are beautiful,' he thought. Then he looked down at the water. 'I will catch a fish for our lunch,' he thought. 'Will and I can make a fire. I have some bread, and there is fruit on the trees. We will eat very well. Ah, this is a wonderful life!'

Below the text window is an 'Opslaan' button.

*Deze invulhulp is bedoeld voor deelnemers van de Brabant Cloud die gebruik maken van het collectiebeheersysteem Memorix Maior. Hierin wordt uitgelegd hoe de tekstextractie-functie van het DAM werkt.*

### Algemene informatie

Binnen Memorix Maior worden digitale bestanden geüpload in het DAM. Voor sommige PDF's is het mogelijk om de tekst uit de PDF te extraheren en doorzoekbaar te maken bij de presentatie.

Om de tekst te extraheren moet de tekst reeds in de PDF beschikbaar zijn. Dit kan op twee manieren:

- Het gaat om een tekstgebaseerde PDF. Dit betekent dat een digitaal tekstbestand (bijvoorbeeld Microsoft Word) is omgezet naar PDF.
- Het gaat om een eerder ge-ocr'de tekst. Een fysieke tekst is gescand en daarna met een eigen programma ge-ocr'd. (N.B. Memorix Maior bevat geen ocr-functionaliteit).

Het is dus niet bij alle PDF's mogelijk de tekst te extraheren. Dit kun je controleren door de originele PDF te openen en de tekst te selecteren. Selecteer je de regels tekst, dan kan Memorix Maior de tekst extraheren. Selecteer je de hele pagina, dan is dit niet mogelijk.



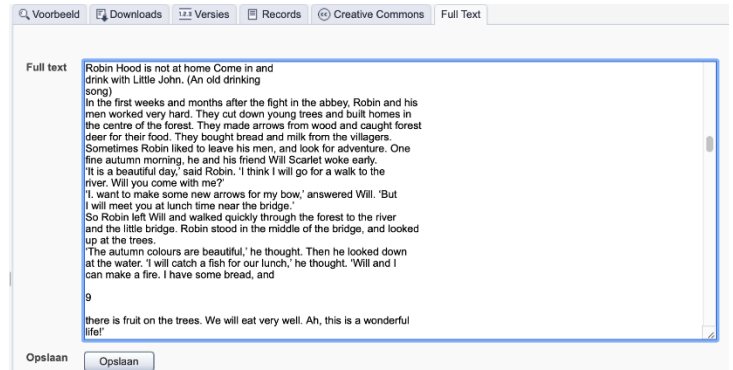
Kijk voor meer informatie op: <https://www.brabantcloud.nl/help>

Of mail naar: [Helpdesk@brabantcloud.nl](mailto:Helpdesk@brabantcloud.nl)

Versie: 1 | juni 2019

### Hoe werkt automatische tekst extractie:

1. Wanneer je een PDF upload naar het DAM wordt de tekst automatisch op de achtergrond geëxtraheerd.
2. Op de detailpagina van het bestand vindt je nu ook de geëxtraheerde tekst onder het tabblad 'Full Text'.  
De tekst in dit veld is aanpasbaar. Als je op 'Opslaan' klikt slaat Memorix Maior de gewijzigde tekst op en wordt deze, indien gepubliceerd, ook meegegeven in de presentatie. De geëxtraheerde tekst wordt daarmee gewijzigd. De tekst in de PDF wordt niet gewijzigd.
3. Op het moment dat je een PDF koppelt aan een record en publiceert wordt de geëxtraheerde tekst online doorzoekbaar (via Instant Website, Brabants Erfgoed en eventueel andere portalen en toepassingen).
4. Wanneer je de PDF loskoppelt van een record blijft de (gewijzigde) tekst beschikbaar in het veld "Full Text". Deze tekst is namelijk niet gekoppeld aan het record, maar het bestand in het DAM. Wanneer je deze PDF aan een ander record koppelt, zal dit nieuwe record dus de betreffende tekst meekrijgen.
5. De geëxtraheerde tekst, met of zonder aanpassingen, gaat verloren wanneer je de PDF uit het DAM verwijdert. Wanneer je de PDF opnieuw uploadt vindt er opnieuw tekstextractie plaats en zal in het veld "Full Text" dus de originele tekst, zonder aanpassingen, te vinden zijn.
6. Op het moment dat je een bestaande PDF in het DAM [vervangt door een nieuwere versie](#) van de PDF treedt de automatische tekst extractie weer in werking. Uit de nieuwe versie wordt de tekst geëxtraheerd en beschikbaar gesteld in het veld 'Full text'. De geëxtraheerde tekst van de oorspronkelijke PDF incl. de eventuele handmatig doorgevoerde wijzigingen gaan dus verloren.



Kijk voor meer informatie op: <https://www.brabantcloud.nl/help>

Of mail naar: [Helpdesk@brabantcloud.nl](mailto:Helpdesk@brabantcloud.nl)

Versie: 1 | juni 2019