

IPASS – Plan van aanpak

Optical Character Recognition

Beschrijving:

Het thema van mijn IPASS zal OCR zijn. OCR staat voor Optical Character Recognition, dit is de techniek waarbij tekst uit een foto van tekst door middel van patroonherkenning herkend en opgeslagen wordt. De eerste vormen van OCR komen uit de jaren 20, maar pas in de jaren 70 werd te toepassing van OCR populair.

Voor mijn IPASS zal ik een onderzoek doen naar OCR toepassingen in de programmeertaal Python. Ik zal 5 Python-OCR-libraries uitzoeken en deze met elkaar vergelijken. Ik zal deze libraries met elkaar scores geven op basis van hun performance in twee situaties, bij het scannen van een natuurlijke foto en bij het scannen van een document. Voor het scannen van natuurlijke foto's zal er een algemene dataset van foto's gebruikt worden. Ik zal de libraries analyseren op ieders methode van het tesseracten van de foto's. De libraries zullen ieder een score krijgen per situatie, deze score zal waarschijnlijk gebaseerd worden op basis van accuraatheid en op snelheid.

Eisen:

Aan het einde van het IPASS zal er een lijst zijn met minstens 5 Python-OCR-Libraries, deze zullen allemaal getest zijn met de twee test situaties, bij deze testresultaten zal een onderzoeksverslag bijgevoegd worden. In dit onderzoeksverslag zal per library een uitleg gemaakt worden met welke methode deze de foto's analyseert.

Planning:

Datum	Doel van week
27 mei / 2 juni	Plan van Aanpak gereed
3 juni / 9 juni	Onderzoek naar verschillende libraries en algoritmes
10 juni / 16 juni	Keuze maken voor libraries
17 juni / 23 juni	Implementaties testen
24 juni / 30 juni	Conclusie van het onderzoek schrijven
1 juli / 7 juli	Onderzoek presenteren

Risico's:

Bij dit IPASS zouden zich een aantal risico's kunnen voordoen. In tijdnood raken zou een van deze risico's kunnen zijn als er niet goed aan de planning gehouden wordt. Een ander risico dat zich voor zou kunnen doen is het verliezen van motivatie, nou is de kans dat dit zal gebeuren bij mij nihil. Een ander risico zou zijn een IPASS dat niet uitdagend genoeg zou zijn, maar na een overleg met de begeleiders van het IPASS is dit risico ook uitgesloten.