



UNIVERSITEIT VAN AMSTERDAM

Proposing related documents within the knowledge graph of StarFish

Second Year's Project of the Bachelor Artificial Intelligence

ROBBERT VAN GINKEL 60606060

JORN PETERS 60606060

LOTTE WEERTS 10423303

Supervisor: RAQUEL FERNANDEZ

Client: PERCEPTUM

Abstract

This is totally an abstract. Like totally

Contents

	Page
1 Introduction	1
2 Product overview	3
3 Implementation	4
4 Conclusion	5

1 Introduction

This report describes the results of the Second Year's project of the Perceptum team. The project focused on creating a *document link recommender system* to the StarFish website.

StarFish, one of the projects of Perceptum, is a website that aims to share knowledge about the education domain by means of a connected graph. People from all around the world should get access to this knowledge graph in a simple, personalized manner. The nodes in this graph are documents and they are connected with links. These documents can be of all sorts of types - e.g. a good practice, information, a question. Each document has a set of tags associated with it, which describe the different aspects of educational innovation. StarFish is community-driven: both the content of the documents as the links between documents are determined by the users of StarFish.

The drawback of a community driven knowledge graph is that not all the users know the entire document base. Especially when the knowledge base grows it becomes impossible for a user, since one does not know the existence of one or more linkable documents. A possible solution could be to make use of administrators, which can devote more time in getting to know all the documents, but that approach has two main drawbacks. First of all, this would mean that some central authority determines whether or not two documents should be linked. This is not in line with the idea of a community-driven knowledge base. Secondly, if the knowledge base grows even further, it becomes impossible also for an administrator to keep track of all documents. Imagine one person having to link all pages on Wikipedia - an impossible job.

In order to overcome the problem of linking documents in a large knowledge base, this process should be automated. This project therefore focuses on automating making the connections between documents. Though ideally these connections should be made completely automatic, a first step would be to create a recommendation system. When a user adds a new document, he or she can choose from a list of proposed documents the documents he or she deems relevant. This means that the recommender system does not have to work perfect, but should work reasonably well enough. Defining 'well enough', however, is also a part of this project. Thus, the product vision of the system can be described in the following concise way:

Figure 1: Caption

Figure 2: Caption 2

Product vision:

For StarFish users

who search for and edit knowledge in starfish

the starfish document linker

is a starfish core system addition

that finds related documents

unlike moderated or individual linking

our product uses algorithms and data to suggest document links

Within the time span of this project multiple ways of recommending links between documents have been explored. The results of these explorations will be discussed in this report.

Algorithm 1 SLIC Segmentation

Require: K = number of super pixels

- 1: Initialize K cluster centres $C_k = [l_k, a_k, b_k, x_k, y_k]^T$ by sampling pixels at regular grid
 - 2: Perturb cluster centres in an $n \times n$ neighbourhood, to the lowest gradient position
 - 3: **repeat**
 - 4: **for all** cluster centre C_k **do**
 - 5: Assign the best matching pixels from a $2S \times 2S$ square neighbourhood around the cluster centre according to the distance measure in eq. 1
 - 6: **end for**
 - 7: Compute new cluster centres and residual error E (L1 distance between previous centres and recomputed centres)
 - 8: **until** $E \leq threshold$
 - 9: Enforce connectivity
-

2 Product overview

ALGORITHMS!!!!

MATHSSS!!!!

$$\begin{aligned} d_{lab} &= \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \\ d_{xy} &= \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \\ D_s &= d_{lab} + \frac{m}{S} d_{xy} \end{aligned} \tag{1}$$

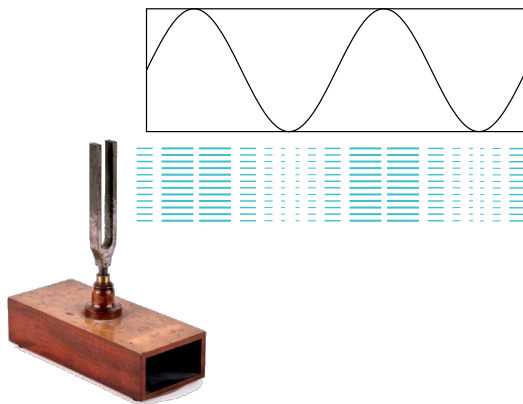


Figure 3: Sound waves propagate through a medium such as air molecules

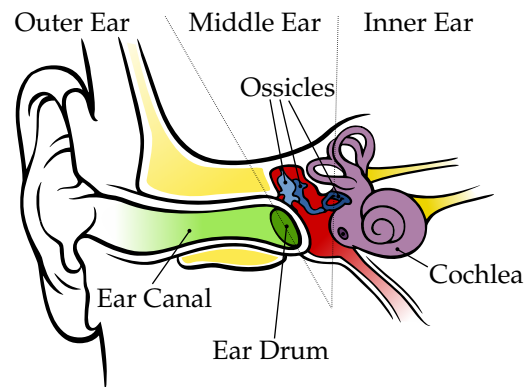


Figure 4: Internals of the ear

3 Implementation

Een ander vet iets dat je kan doen is svg in Inkscape openen en opslaan als pdf met latex output! Dan kan je dus latex tekst hebben binnen plaatjes zoals in figuur 4.

4 Conclusion

Bla.

References

Robrecht, Harrie, and Lotte. Totally bitching paper. *Journal*, 1(1):1–5, January 2013.