Developing a software product rarely finishes. It is maintained, refactored, repurposed, updated, extended, etc. Especially with creative products, where the functional requirements are more fluid perhaps, it is always tempting to change things.

For the purpose of this doctoral project, the artefact `pata.physics.wtf` is a snapshot of a product in constant motion. The state of the code at the time of
§ ?? submission of this thesis is described in chapter **??** and further elaborated on
§ ?? in the **??** chapter. But it may very well continue to evolve.

Here, in this chapter I will lay out some of the potential further work for this project. This may continue on a private basis or in a more academic environment.

0.1 Performance

Startup The website can be slow to load. Currently speed performance was not a priority during development. In fact it is not built for speed from the ground up. Each time the server restarts, the indexing process takes place from scratch
§ ?? (see chapter **??**). This takes time. Google and other big web search engines
§ ?? do this continuously in the background to keep data up to date. The index is currently cached after startup but perhaps preprocessing it and storing it more permanently in a database would help speed up the start. However this may not be necessary, as it only affects the server startup.

Query Response The time it takes from the user entering a query term and the system displaying the results page varies between unnoticable short and impatiently long. This is due to the pataphysicalisation process. This requires calls to external and internal **API!** (**API!**)s such as Flickr and WordNet. See
▦ ?? analysis on speed issues in table **??**.

Preprocessing Corpora At this point the texts in the corpora consist of al-
§ ?? most unedited plaintext ('.txt') files[1] (see chapter **??**). Newlines and whitespace formatting varies, as does language and quality of spelling. Generally, chapter headings, chapter numberings, etc. were left untouched. The Shakespeare corpus contains poetry and plays for example. With the plays, scene information, stage directions, and voice details were kept. This means sentences that appear in the results of the search tool can contain peripheral words such as in this example: "...Athens and a wood near it ACT I ..." from *A Midsummer Night's Dream* or this example: "...Exit SHERIFF Our abbeys and our priories shall pay This expedition's charge ..." from *King John*. This could be addressed by

[1] For text files downloaded from Project Gutenberg, the Gutenberg specifc copyright notices have been removed to only contain the relevant body of text

Go to TOC

preprocessing the individual texts in advance and removing any text that might interfere with the readablility of results.

Image Sizes At the moment images are retrieved at one specified size through the various **API!** calls even though they are displayed at various different sizes depending on their location in the image spiral (unless they are displayed as a list). This process could certainly be optimised. Smaller image sizes could be accessed via the **API!**s.

0.2 Design

§ **??** **Responsive Spirals** Currently the image and video spirals (see chapter **??**) are fixed size. This means that when the webpage is resized the spiral stays the same size and is left-aligned on the page. Ideally it would be better to scale the spiral with the width of the browser page. This could be achieved using percentage widths, although it would require a lot of work to adapt the current
§ **??** code for the spirals (see chapter **??**).

Scalable Image Sizes As mentioned above, images are retrieved at one size through the various **API!** calls. Because images in the spiral have different sizes according to where in the spiral they are located, they are scaled up or down directly in the **HTML!** (**HTML!**) code. This means that some of the images look distorted and pixelated if they have to be scaled up or down too much.

Square Aspect Ratio Another issue is the aspect ratio of images and videos. For the spiral they need to be square. They are currently distorted as opposed to cropped. It might be possible to specify an option in the **API!** calls to only retrieve square images which would help this problem.

Responsive Poems A similar problem to the responsive spirals exists with the display of the Queneau poems. The random poems are centered on the page but the Queneau poems require a lot more formatting and styling to render and currently this is achieved my left-aligning them and having a fixed 'absolute' position on the page. Ideally this would also be centered as in the random poems.

Paginate Results For the text-by-source and text-by-algorithm search as well as the image- or video-as-list searche results, it may improve the loading speed of the results page to split the results into smaller chunks and display them on several pages instead of one long scrolling page. This is called pagination.

Go to TOC

Random Sentences Adding to the source of random sentences used in the top and bottom banner on the website should be an ongoing endevour. The current
§ **??** list of sentences used is shown in appendix **??**.

0.3 Text

Result Sentences Currently the way result sentences are retrieved for the text
§ **??** search is based on punctuation (see chapter **??**). This means once a pataphysicalised keyword has been found, the system retrieves up to 10 words prior until it reaches a punctuation mark and the same for after. The idea here was to get suitable sentence fragments. This could be changed to rely on **POS!** (**POS!**) tags for example or simply retrieving complete sentences.

Stopwords When the index is created only words that are not considered stop-
§ **??** words are added. We could modify the list of stopwords (see appendix **??**) to include a few more uninteresting words. Or we could simply remove everything but nouns for example. This would drastically influence the results produced by the system.

Rhyming Scheme One of the biggest points for future work is to introduce a rhyming scheme for the poetry results. This might involve some more **NLP!**
§ **??** (**NLP!**) during the creation of the index. It would make the poems much more readable. This could include pronounciation **POS!** tags or other **IPA!** (**IPA!**) like data (for example using an **API!** like Wordnik (**Wordnik2016**) or a library like **NLTK!** (**NLTK!**)). So a word in the index dictionary might contain the following items.

```
(``tree'': [``l_00'': [24,566,4990], ``s_14'': [234,5943]],
      ↪  ``[tri]'')
```

§ **??** By doing **POS!** tagging with pronounciation data, we could retrieve sentences that match the sound of the last word of the previous line for example.

0.4 Pataphysicalisation

WordNet The vocabulary in WordNet is limited. According to it's website (**Princeton2010**) it contains 117,000 'synsets'[2] This affects two of my algorithms (namely the Syzygy
§ **??** and Antinomy algorithms). See also discussion in chapter **??**. An option might

[2] Synonyms—"words that denote the same concept and are interchangeable in many contexts"—are grouped into unordered sets called synsets.

Go to TOC

be to somehow widen the amount of word matches by including different word-types/forms and relationships, such as troponyms, homonyms and heteronyms. Using these could introduce a whole new kind of pataphysical result.

Homonyms are pronounced the same but mean something else (e.g. 'write' and 'right'). Heteronyms are words that are spelled the smae but have a different meaning (e.g. 'close to the edge' and 'to close the door'). Homophones are often used to create puns (and remember—puns are syzygys of words), for example "past your eyes" and "pasteurize".

> You can tune a guitar, but you can't tuna fish. Unless of course, you play bass.
> attributed to Douglas Adams

Antinomy The antinomy algorithms relies on WordNets antonyms. A lot of words simply do not have an opposite and no fallback is currently defined. This means a lot of the time the antinomy function will not produce any results. Andrew Dennis implemented the algorithm in the same way, as discussed in § **??** chapter **??**. It would be great to come up with a better way of dealing with this concept to ensure results are produced everytime.

Stemming Stemming could increase the number of results found by all al- § **??** gorithms. (See chapter **??**). A danger of increasing the output of the pataphysicalisation is always that results become more boring. Currently queries such as 'clear' and 'clearing' are treated as separate entities and would produce different results. Stemming would turn both of these words into the stem 'clear' and they would return the same results. Now it becomes immediatly clear (no pun intended) though that this might not always be desirable as just illustrated in this sentence: the root meaning of 'clear' can be very different to the meaning of 'clearing'.

Queneau's poems It would be nice to actually add Queneau's poems (**Queneau1961**) § **??** into the Faustroll corpus as little easter egg (see chapter **??**).

Image Algorithms The image and video search currently rely on external **API!**s § **??** (see chapter **??**). One option to approach this in a totally differnet way would be to write algorithms that analyse and pataphysicalise the actual image or video data themselves. This might involve manipulating histograms or pixel maps.

§ **??** **Maximum Obscurity** N-grams are a **NLP!** technique introduced in chapter **??**. The idea is that it allows for prediction of likely word pairs, meaning if the word 'sunny' often occurs just before the word 'day' in a given training text or corpus

Go to TOC

then the probability for this particular n-gram is higher than say for 'sunny dog'. This can be increased to predict the probability of longer chains of words. On can immediately see the attraction of abusing this to generate pseudo sentences or even of creating a formula similar in nature but for example ranking obscure combinations of words higher than common ones. So for example instead of
Σ ?? having a **MLE!** (**MLE!**) (see equation **??**) we could have a 'Maximum Obscurity Estimation' which returns the highest probabilty for word sequences that happen the rarest.

Pataphysical Entropy Similarly, we could could play with maximum entropy
§ ?? models as shown in chapter **??** together with **POS!** tagging by rigging given probability for tags. There are endless possipilities of abusing these kinds of techniques. This is also very reminiscent of **OULIPO!** (**OULIPO!**) techniques.

Grammars We could create a whole new language grammar based on pata-
§ ?? physical principles. Examples of using a standard grammar (see chapter **??**) for generating 'random' text are as follows.

ArtyBollocks
: Generates artist statements (**Winter2016**).

DadaEngine
: A system for generating random text from grammars (**Dada2016**).

SciGen
: Generates random Computer Science research papers (**Stribling2016**).

§ ?? **Uncreativity** In chapter **??** I discussed the concepts of uninspiration and aberration by Wiggins and Ritchie (**Wiggins2006**; **Ritchie2012**).

Pataphysical Search Framework We have explained earlier that evaluating creative search results largely depends on the user's understanding of relevance, usefulness and creativity. We can define pataphysical concepts and evaluation criteria in terms of the **CSF!** (**CSF!**) to provide a common ground for this understanding. The suggested definitions should provide more suitable criteria for the analysis of creative search results rather than traditional results.

Table 3 – Extension to the **CSF!** to provide a notation for pataphysical concepts. Norm and value are defined in a traditional sense, with the other pataphysical concepts being defined around those two. Table 3 shows various definitions for pataphysical principles that can be applied as criteria in search processes. There are of course more ways in which they could be expressed but as a demonstration this should suffice. Combined with a set of evaluative rules inspired by

Go to TOC

Table 0.1 – My caption

Name	**Equation**
Universal set of concepts	U
Aberration	$B \notin N_\alpha(X) \ \wedge \ B \neq \emptyset$
Perfect Aberration	$V_\alpha(B) = B$
Productive Aberration	$V_\alpha(B) =\neq \emptyset \wedge \ \neq B$
Pointless Aberration	$V_\alpha(B) = \emptyset$
Hopeless Uninspiration	$V_\alpha(X) = \emptyset$
Conceptual Uninspiration	$V_\alpha(N_\alpha(X)) = \emptyset$
Generative Uninspiration	$elements(A) = \emptyset$

Ritchie (2001), this forms an essential part of creative exploratory search systems. Some example evaluative criteria are shown here: The norm and value ratings by themselves have lower importance than the pata rating, the average pata rating should be suitably high, results with good pata rating should form a significant proportion of the results, and there should be a good mixture of the different pataphysical concepts present in the results. Producing, interpreting and evaluating search results based on these ideas, together with a healthy knowledge and apprehension about the system's purpose, should be much more suitable and "fair" than trying to apply conventional measures such as precision and recall to something so subjective.

https://wordnet.princeton.edu/wordnet/man/wngloss.7WN.html for glossary

fix all chapter XYZ mentions

if pataphysicalisation happens at different locations...

 0.1 They are all possible and will maybe be done all at the same time. We can consider the possibility of a 'patametric index', rather than a parametric index or a 'patasaurus' (pataphysical thesaurus/ontology).

copy and adapt figure

Go to TOC

Table 0.2 – My caption

Name	**Equation**	**Condition[redacted]tion**	
Norm	$N_\alpha(X) = \{c \in X \mid N(c) > \alpha\}$	$N \in [0,1]^X$	
Value	$V_\alpha(X) = \{c \in X \mid V(c) > \alpha\}$	$V \in [0,1]^X$	
Pata	$P_\alpha(X) = \{c \in X \mid c \in (CLI(X) \cup ANT(X) \cup SYZ(X) \cup ANO(X) \cup ABS(X))\}$		
Clinamen[redacted]	$\{c \in X \mid N_{0.9}(N_{0.1}(c))\}$	$X \subseteq U$	smallest possible aberration to make the biggest difference
Antinomy[redacted]	$= \{c \in X \mid V(N_0(c)) > \alpha\}$	$X \subseteq U$	reachable, abnormal concepts with value
Anomaly[redacted]	$) = \{c \in X \mid N(c) < \alpha\}$	$X \subseteq U$	reachable concepts outside the norm
Absolute[redacted]		$V_1(N_1(X)) \neq$[redacted]r value and $\emptyset$	norm must be perfectly matched
Syzygy 1	$SYZ(query) = \bigcup_{n=0}^{3} elements(Q(N,V)^n(query))$		concepts reachable within 3 steps from the query
Syzygy 2	$SYZ(X) = S'(X)$	$S_{obj} \rightarrow S^{meta} \rightarrow S'obj$	transformed set of concepts

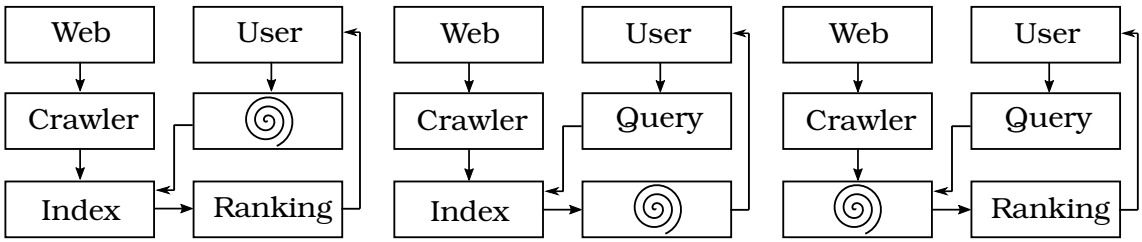

Figure 0.1 – Pataphysicalisation

Go to TOC

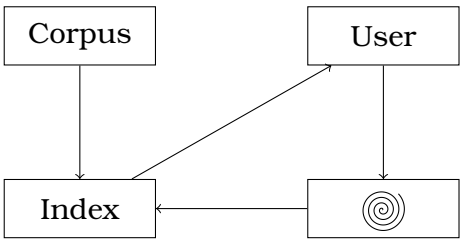

Figure 0.2 – Pataphysical system architecture

More APIs Currently X **API!**s are used[3]. This could be increased to include more varied sources of data. Sites like Flickr are heavily based on user tags ('folksonomies') which can be unreliable and a bit random at times.

Web search The use of **API!**s could also include web search results rather than just images and videos. This would needs its own interface section and a suitable display style for the results. The biggest problem for this is **API!** restrictions. Alternatively a ready-made index or crawl could be used but these are typically many terrabytes in size and have a cost attached. Crawling the Web myself is not an option due to the computational power, time and space required to do so.

Audio search Originally audio search was going to be a part of this project. This has been abandoned due to time constraints. However it could be added using an **API!** such as SoundClouds. Technically the pataphysicalisation could work similar to the image and video searches, meaning it would be based on user tags. One idea would be to search in audio waves.

More algorithms It would be nice to implement some more algorithms for the search tool. This could include the two additional algorithms suggested by An-
§ **??** drew Dennis (see chapter **??**) or developing more of my own. This could involve implementing some of the other pataphysical principles, such as equivalence or anomaly. Or it could consist of implementing some of the more famous **OULIPO!** techniques. The repetoire of them is huge (see appendix XYZ).

Custom API It would be great to develop a custom **API!** for this the search tool. This would allow other people to use the search remotely without going through the interface and to use the results as they want. This would have been benefi-
§ **??** cial for the Digital Opera project and certainly for other researchers/developers
§ **??** like Adnrew Dennis.

Focus group It might be interesting to look at opinions of various people (general public and experts) about the interpretation/evaluation framework. This

[3] Flickr, Getty, Bing, MicrosoftTranslator and YouTube

Go to TOC

could be done by asking them to provide their own definition of computer creativity and then to analyse and evaluate a product (such as `pata.physics.wtf`) accoriding to their own criteria. Then follow this up by getting the same people to use my proposed framework to compare the results. This would include asking them about wether or not they thought that using the framewokr was beneficial to them or confusing.

Questionnaires I have shied away from doing a questionnaire study because of several reasons. One is that due to the creative and subjective nature of the artefact, opinions on it may vary wildly and I don't see how I could derive useful unbiased data from that. Yes, it depends what questions you ask. But even if I managed to get some half-decent data, what would that tell me? Half of the people like my site, the other half don't?

Eye-tracking To study the effects of using different styles of presenting the same results an eye-tracking experiment could be done. This would involve setting up participants with the necessary equipment and then introduce them to the website and moniter their eye movements as they naviagte the site. This could also provide details about how long users spend on each results page, what kind of style of results they prefer, etc. Some may prefer image or video search over the text search while others may not be interested in that at all. Generally of course one has to take into account that this is a creative piece of work and not everybody will like it. It has no clear immediate purpose and that may put users off.

Performance Benchmarks?

Go to TOC