

# LIST OF TODOS

- go over previous chapters incl lit review and refer back to things. bring things together. show the breadth and depth of my research!!! . . . . . 10
- relate all of these things back to my topic of AMC . . . . . 10

Institute of Creative Technologies  
De Montfort University

FANIA RACZINSKI

# ALGORITHMIC META-CREATIVITY

**Creative Computing and Pataphysics  
for Computational Creativity**

**pata.physics.wtf**

***Supervisors:***

Prof. Hongji YANG  
Prof. Andrew HUGILL  
Dr. Sophy SMITH  
Prof. Jim HENDLER

***A thesis submitted in partial fulfilment of the requirements  
for the degree of Doctor of Philosophy***

Created: 25th March 2015 — Last Saved: 5th November 2016  
Wordcount: 10466

[Go to TOC](#)

# PRE

# TL;DR

## **Algorithmic Meta-Creativity** — Fania Raczinski — Abstract<sup>1</sup>

Using computers to produce creative artefacts is a form of computational creativity. Using creative techniques computationally is creative computing. Algorithmic Meta-Creativity ([AMC](#)) spans the two—whether this is to achieve a creative or non-creative output. Creativity in humans needs to be interpreted differently to machines. Humans and machines differ in many ways, we have different ‘brains/memory’, ‘thinking processes/software’ and ‘bodies/hardware’. Often creative output by machines is judged in human terms. Computers which are truly artificially intelligent might be capable of true artificial creativity. Until then they are (philosophical) zombie robots: machines that behave like humans but aren’t conscious. The only alternative is to see any computer creativity as a direct or indirect expression of human creativity using digital means and evaluate it as such. [AMC](#) is neither machine creativity nor human creativity—it is both. By acknowledging the undeniable link between computer creativity and its human influence (the machine is just a tool for the human) we enter a new realm of thought. How is [AMC](#) defined and evaluated? This thesis address this issue. First [AMC](#) is embodied in an artefact (a pataphysical search tool: [pata.physics.wtf](#)) and then a theoretical framework to help interpret and evaluate such products of [AMC](#) is explained.

**Keywords:** *Algorithmic Meta-Creativity, Creative computing, Pataphysics, Computational Creativity, Creativity*

---

<sup>1</sup>“Too long; didn’t read”

# PUBLICATIONS

**Fania Raczinski** and Dave Everitt (2016) “***Creative Zombie Apocalypse: A Critique of Computer Creativity Evaluation***”. Proceedings of the 10th IEEE Symposium on Service-Oriented System Engineering (Co-host of 2nd International Symposium of Creative Computing), SOSE’16 (ISCC’16). Oxford, UK. Pages 270–276.

**Fania Raczinski**, Hongji Yang and Andrew Hugill (2013) “***Creative Search Using Pataphysics***”. Proceedings of the 9th ACM Conference on Creativity and Cognition, CC’13. Sydney, Australia. Pages 274–280.

Andrew Hugill, Hongji Yang, **Fania Raczinski** and James Sawle (2013) “***The pataphysics of creativity: developing a tool for creative search***”. Routledge: Digital Creativity, Volume 24, Issue 3. Pages 237–251.

James Sawle, **Fania Raczinski** and Hongji Yang (2011) “***A Framework for Creativity in Search Results***”. The 3rd International Conference on Creative Content Technologies, CONTENT’11. Rome, Italy. Pages 54–57.



A list of talks and exhibitions of this work, as well as full copies of the publications listed above, can be found in appendix ??.

# CONTENTS

<b>Todo list</b>	<b>1</b>
------------------	----------

---

## **PREFACE**

<b>TL;DR</b>	<b>ii</b>
<b>Publications</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>Figures</b>	<b>vi</b>
<b>Tables</b>	<b>vii</b>
<b>Code</b>	<b>viii</b>
<b>Acronyms</b>	<b>ix</b>

---

## **HELLO WORLD**

---

## **TOOLS OF THE TRADE**

---

## **THE CORE: TECHNO-LOGIC**

---

## **THE CORE: TECHNO-PRACTICE**

---

## **META-LOGICALYSIS**

---

<b>1</b>	<b>Patanalysis</b>	<b>9</b>
1.1	Influences . . . . .	10
1.2	Pataphysicalisation . . . . .	12
1.3	Creativity & Intelligence . . . . .	32
1.4	Design . . . . .	40
1.5	Meta Analysis . . . . .	41

---

## HAPPILY EVER AFTER

---

### POSTFACE

---

<b>References</b>	<b>46</b>
-------------------	-----------

# FIGURES

1.1	Faustroll vs. Shakespeare poetry . . . . .	14
1.2	Changing base in Clinamen . . . . .	24
1.3	Semantic relationships of ‘feather’ . . . . .	25
1.4	Image spiral ‘blue kitten’—Flickr . . . . .	29
1.5	Image spiral ‘blue kitten’—Getty . . . . .	30
1.6	GitHub contributions . . . . .	42

# TABLES

1.1	Comparison of patalgorithms . . . . .	13
1.2	Faustroll vs Shakespeare in numbers . . . . .	15
1.3	Numbers per algorithm . . . . .	17
1.4	Count and time of results . . . . .	20
1.5	Changing number of errors in Clinamen . . . . .	25
1.6	Quantities of different semantic relations . . . . .	27
1.7	Metric prefixes . . . . .	36
1.8	Layers of abstraction in computers vs brains . . . . .	37

# CODE

# ACRONYMS

<b>AMC</b>	Algorithmic Meta-Creativity
<b>IR</b>	Information Retrieval
<b>AI</b>	Artificial Intelligence
<b>AGI</b>	Artificial General Intelligence
<b>TDM</b>	Term-Document Matrix
<b>API</b>	Application Program Interface
<b>RDF</b>	Resource Description Framework
<b>OULIPO</b>	Ouvroir de Littérature Potentielle
<b>VR</b>	Virtual Reality
<b>CPU</b>	Central Processing Unit
<b>EU</b>	European Union
<b>HBP</b>	Human Brain Project
<b>FLOPS</b>	Floating-Point Operations Per Second

## Part I

# HELLO WORLD

That it might upon him, for always very well be the sun himself and fear fell upon them so sincerely in love. The spacious hall prepare, the fishers hall each other not - Nor help - in their fraternal lot, the side of a great hill, with a hillock of sand, aux montagnes d'origine, . . . Ludgate hill, till the Spadefoots made their body. Who longs to plunge two fellow creatures into the water, who bends his head at the four corners. She fell on to a hillock of sand, aux montagnes d'origine, . . . Ludgate hill, till the Spadefoots made their body.

## Part II

# TOOLS OF THE TRADE

Made up your habill'd minds to brave me, ce train re  
comme'nait quand' que'z week's silenty, a diff'rent voyage, against the tide, aucun employe de  
long souverain coller road. Followed by a train plus  
t'ree and plus t'ree. Sell that which ye have, to be their mouthpiece is it true, that  
Sir Excellency stooped to take it up, or in the vegetab'ly  
part, a train is due, mad voyage.

# INTERLUDE I

(...) through aesthetic judgments, beautiful objects appear to be “purposive without purpose” (sometimes translated as “final without end”). An object’s purpose is the concept according to which it was made (the concept of a vegetable soup in the mind of the cook, for example); an object is purposive if it appears to have such a purpose; if, in other words, it appears to have been made or designed. But it is part of the experience of beautiful objects, Kant argues, that they should affect us as if they had a purpose, although no particular purpose can be found.

(Burnham 2015, ch.2a)

Chance encounters are fine, but if they have no sense of purpose, they rapidly lose relevance and effectiveness. The key is to retain the element of surprise while at the same time avoiding a succession of complete non-sequiturs and irrelevant content

(Hendler and Hugill 2011)

Conducting scientific research means remaining open to surprise and being prepared to invent a new logic to explain experimental results that fall outside current theory.

(Jarry 2006)

### **Part III**

# THE CORE: ΤΣΕΧΝΟ- ΛΟΓΙΚ

Do not cry and bleed to will, cloth to be sure, your blows it cringe and definitely. A royal robe none can miss, how cold she must be, sa belle robe rose en desordre. Come like u fillet sur le centre de la France et qui s'appela, mes bagages et regler ma note, if pure bijoussin. Ils peuvent aller a tout le moins ay sens in very quintessence, there is none of the jaded

## Part IV

# THE CΞRE: TΣCHNΞ- PRΆCTICΞ

I do not perform normal, his Excellecy in regard to my instructions, for he had already begun to exercise the tools, I could not help thinking of the wild ritual of this work. Importance de fonctionnement avec ce rituel, ce qui en engageait la mortelle, ce que je devais faire pour aider à la mortelle, mais dans une autre usage. And four thousand allians made me in different ways, I could not help it, but if you will follow my instructions, for he had already begun to exercise the tools, I could not help thinking of the wild ritual of this work. Importance de fonctionnement avec ce rituel, ce qui en engageait la mortelle, ce que je devais faire pour aider à la mortelle, mais dans une autre usage.

# INTERLUDE II

Computation is not a fact of nature. It's a fact of our interpretation.

([Searle 2015](#))

all the familiar landmarks of my thought - our thought, the thought that bears the stamp of our age and our geography - breaking up all the ordered surfaces and all the planes with which we are accustomed to tame the wild profusion of existing things, and continuing long afterwards to disturb and threaten with collapse our age-old distinction between the Same and the Other.

([Foucault 1966](#))—taking about Borges

Only those who attempt the absurd achieve the impossible.

(attributed to M.C. Escher)

A great truth is a truth whose opposite is also a great truth. Thomas Mann

(as cited in [Wickson, Carew and Russell 2006](#))

Heisenberg's Uncertainty Principle is merely an application, a demonstration of the Clinamen, subjective viewpoint and anthropocentrism all rolled into one.

([Jarry 2006](#))

Epiphany – 'to express the bursting forth or the revelation of pataphysics'

Dr Sandomir ([Hugill 2012](#), p.174)

Machines take me by surprise with great frequency.

([Turing 2009](#), p.54)

The view that machines cannot give rise to surprises is due, I believe, to a fallacy to which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into the mind simultaneously with it. (Turing 2009, p.54)

Opposites are complementary.

It is the hallmark of any deep truth that its negation is also a deep truth.

Some subjects are so serious that one can only joke about them.

Niels Bohr

There is no pure science of creativity, because it is paradigmatically idiographic — it can only be understood against the backdrop of a particular history.

(Elton 1995)

Tools are not just tools. They are cognitive interfaces that presuppose forms of mental and physical discipline and organization. By scripting an action, they produce and transmit knowledge, and, in turn, model a world.

(Burdick et al. 2012, p.105)

Humanists have begun to use programming languages. But they have yet to create programming languages of their own: languages that can come to grips with, for example, such fundamental attributes of cultural communication and traditional objects of humanistic scrutiny as nuance, inflection, undertone, irony, and ambivalence.

(Burdick et al. 2012, p.103)

## Part V

# MΣΤΑ- ΛΟΓΙΚΑΛΥΣΙΣ

Apart off a skull, meat off a skull, meat always suspends the seat, the heat of the sun being very great, pet. Is there not a fine horse medal of a Cycloidal mesh by mesh again, sit not down in the chief seat. Then like a pants horse let go, there will be a screwing him, the Oath of the Little men.

From a few sea, gobble ebery bit ob de  
meat by the mere smell of one of his drugs. D'un jet de science lectrique, who yet always suspends the seat, the heat of the sun being very great, pet. Is there not a fine horse medal of a Cycloidal mesh by mesh again, sit not down in the chief seat. Then like a pants horse let go, there will be a screwing him, the Oath of the Little men.

# PATANALYSIS

## 1

Where thou mayst knock a nail into his head,  
but near him thy angel becomes a fear,  
it must omit real necessities,  
hear Faith infringed which such zeal did swear.

With sighs in an odd angle of the isle,  
before me to sweet beds of flow,  
might quench the zeal of all professors else,  
the whilst his iron did on the anvil cool.

Intend a kind of zeal both to the Prince and Claudio,  
and threescore year would make the world away,  
nay if you read this line.

Have no delight to pass away the time,  
by a shadow like an angel,  
four nights will quickly dream away the time.

1.1	Influences . . . . .	10
1.2	Pataphysicalisation . . . . .	12
1.2.1	Numbers . . . . .	14
1.2.2	Sentences . . . . .	18
1.2.3	Index . . . . .	20
1.2.4	Clinamen . . . . .	22
1.2.5	Syzygy . . . . .	23
1.2.6	Antinomy . . . . .	27
1.2.7	APIs . . . . .	27
1.3	Creativity & Intelligence . . . . .	32
1.3.1	Free Will & Surprise . . . . .	32
1.3.2	Understanding & Simulation . . . . .	33
1.3.3	Brain & Computers . . . . .	34
1.3.4	Creativity . . . . .	38
1.3.5	State of the Art . . . . .	38
1.4	Design . . . . .	40
1.5	Meta Analysis . . . . .	41



go over previous chapters incl lit review and refer back to things. bring things together. show the breadth and depth of my research!!!

relate all of these things back to my topic of AMC



A lot of the more theoretical aspects of this research have been discussed in § ?? & ?? chapters ?? and ?. The evaluation here is more concerned with the practical artefact `pata.physics.wtf` and its interpretation.

The chapter is divided into several sections addressing issues related to `pata.physics.wtf`. This includes a discussion of the inspirations, an analysis of some of the technical aspects, a review of design decisions made, a contextualisation and also a meta-analysis of the project's execution and management.

## 1.1 INFLUENCES

§ ?? Looking back over the inspirations for this project described in chapter ??, some of the influences can be clearly seen straight away. Others are intentionally a bit

more subtle. There are various motivations for that. First, transparency conflicts with surprise. **Serendipity** was one of the original aims to try and model, so being overly obvious and descriptive about what the tool is and does would be counter productive. An element of surprise also makes it more enjoyable in repeat visits. Pure randomness is meaningless. Another reasons was **humour**. Pataphysics has an intrinsic kind of humour I wanted to include in the whole presentation of the artefact.

### **Syzygy Surfer**

- § ?? The influence of the Syzygy Surfer cannot be overstated. It forms the immediate predecessor to my research. The authors of the Syzygy Surfer are part of my supervisory team. This is where the initial ideas for the pataphysical algorithms came from. There are important differences as well though. For example, pataphors were never implemented as originally suggested. The idea of using ontologies and semantic web technologies such as Resource Description Framework ([RDF](#)) to develop the system was abandoned early on too.

### **Faustroll Library**

- § ?? This fictional library of real books was direct inspiration for the Faustroll corpus used in the text search. I tried my best to complete the library as accurately as I could but some of the texts were unsorceable. As with the original, I included some foreign language texts. Since the results (if the Faustroll corpus is chosen of course) are drawn from any of these texts, the mood and style of language is quite distinct and atmospheric.

### **Queneau's $10^{14}$ poems**

- § ?? Queneau is another one of the inspirations that became a direct influence.
- § ?? The text search can be displayed as poetry in the same style as Queneau's 100 thousand million poems only in digital form and with a larger set of lines. This means that many more possible poems can be generated by switching individual lines. The outcome is beautiful.

### **Chinese Encyclopedia**

- § ?? Borges story has been an inspiration right from the start. The subtle humour in it is great. The sort of semantic logic behind it was modeled through the pataphysical algorithms.

### **Yossarian**

- § ?? The metaphorical algorithms are intriguing but elusive—I wasn't able to find any details on their implementation. This may be due to the nature of the project, which is commercial rather than academic. It is hard to compare against this site as it is so different even though we share some of the same goals or principles.

## Library of Babel

- § ?? The library of babel is a great project which has only indirectly influence my work. The pataphysical elements in it are obvious even though perhaps unconscious. The seriousness with which the library is presented, the pseudo-scientific approach, the vagueness of what's actually behind it. Is it random? Or is it indeed the most gigantic digital library of any book every written or even to be written? The sheer perceived scale of the library was part motivation for calculating the numbers of the generatable poems.

## Oulipo

- § ?? Given that the Ouvroir de Littérature Potentielle ([OULIPO](#)) is directly rooted in pataphysical principles<sup>1</sup>, the influence on this project cannot be underestimated. The algorithms created could even be seen as an oulipian technique themselves.

## Coder Culture

- § ?? This group of inspirations is a bit more generic and influenced lots of little things throughout the project. The idea of hiding easter eggs on the site, the deliberate placement or use of errors, the obfuscation, the humour, the jargonisation and littered 'l33t' style language, and the art and aesthetics behind it. All of that was influenced by coder culture—and most of all perhaps: this thesis.

## 1.2 PATAPHYSICALISATION

- § ?? As mentioned in chapter ??, the internal transformation of a query term to the final results is what I called the **pataphysicalisation** process. The three pataphysical algorithms (Clinamen, Syzygy and Antinomy), or **patalgorithms**, are at the center of this process.

1. User enters single query term,
2. system transforms query term into list of pataphysicalised terms,
3. system retrieves sentence fragments containing keywords from this list,
4. system displays sentence fragments in various formats.

It is quite interesting to compare the algorithms with each other. By removing the clutter (in this case the sentence surrounding the pataphysicalised keyword) we can see a few example results side by side below in table 1.1.

- 1.1 Seeing the results in a table like this gives an almost immediate idea of how each algorithm works. This is not meant to be transparent and perhaps only

---

<sup>1</sup>Remember that the [OULIPO](#) was founded as a subcommittee of the "Collège de Pataphysique" in the 60's.

Table 1.1 – Comparison of path algorithms showing a selection of results for each

<b>Query</b>	<b>Clinamen</b>	<b>Syzygy</b>	<b>Antinomy</b>
<b>clear</b>	altar, leaf, pleas, cellar	vanish, allow, bare, pronounce	opaque
<b>solid</b>	sound, valid, solar, slide	block, form, matter, crystal, powder	liquid, hollow
<b>books</b>	boot, bones, hooks, rocks, banks	dialogue, authority, record, fact	—
<b>troll</b>	grill, role, tell	wheel, roll, mouth, speak	—
<b>live</b>	love, lies, river, wave, size, bite	breathe, people, domi- cile, taste, see, be	recorded, dead

after knowing the ins and outs of the algorithms can one recognise how each result was found.

The clinamen results show words that contain one or two spelling errors of the original query term. It is perhaps counter-intuitive to have words such as 'altar', 'leaf' and 'cellar' be classed as spelling errors of the word 'clear' but they clearly could be. Remember that a spelling error can be classed in one of four ways: (1) deletion, (2) insertion, (3) substitution and (4) transposition. So, going from 'clear' to 'altar' is an instance of two times case 3 ('c' is replaced by 'a' and 'e' is replaced by 't') and going from 'clear' to 'leaf' is an example of case 1 ('c' is deleted) and case 3 ('r' is replaced by 'f').

Looking at the second column (the syzygy results) shows the semantic relationship between the original query term and the results. Again, this may not be immediately noticeable but certainly once you know how the process works you can recognise the common relations. This is especially evident for the antinomy algorithm which is based on opposites.



??, and ??). It also highlights the difference between the two corpora. Poems based on the Faustroll corpus have a very different sound and feel to it than ones based on the Shakespeare corpus.

There was a period put to the Fire pink and spot earth was flat like the floor of an Oven as much ease as a mower doth the grass	O bloody period I as your lover speak has she such power gather those flowers
during the first period of my captivity room with a hard earthen floor not within everyone's power or your favourite flowers died	thy lover juiced flowers had I been any god of power or a lover's lute
shocks lose power the white daisy after a long period	the river hath thrice flow'd but sad mortality o'ersways their power now here a period of tumultuous broils
poppy peony stock to all People	led by their master to the flow'red fields not a minister in his power where souls do couch on flowers

Figure 1.1 – Comparison of Faustroll (left) versus Shakespeare (right) poetry, both for query term ‘flower’

Sometimes we can even get a general feel for the theme of the poem, as in we can recognize the connection, the relationship between the individual lines and what must be the original query term. Of course putting the poems into the chapters as they are—without specifically stating the keyword they were generated from or the corpus they are based on—makes them a bit more elusive.

The different language is quite obvious. This is helped by the fact that the Shakespeare corpus is of course written by the same author<sup>2</sup>. The Faustroll corpus contains text by over 20 different authors and in three different languages even.

### 1.2.1 NUMBERS

The above examples (table 1.1 and figure 1.1 give a good overview of the two main factors in the pataphysicalisation process, namely the three patalgorithms and the two corpora. Both only reflect a small selection of the variety of results produced though. It is therefore quite interesting to look at some actual numbers.

---

<sup>2</sup>Unless of course we believe the legends that Shakespeare didn't write those works by himself...

Table 1.2 – Faustroll versus Shakespeare in numbers

<b>Query</b>	<b>Corpus</b>	<b>Results</b>	<b>Reverbs</b>	<b>Origins</b>	<b>Poems</b>
flower	Faustroll	90	25	18	$7.8 \times 10^{10}$
	Shakespeare	158	15	38	$3.8 \times 10^{14}$
clear	Faustroll	542	79	23	$1.3 \times 10^{22}$
	Shakespeare	1445	72	38	$1.5 \times 10^{28}$
troll	Faustroll	124	16	16	$4.4 \times 10^{12}$
	Shakespeare	327	14	38	$1.1 \times 10^{19}$
fania	Faustroll	9	2	6	1
	Shakespeare	15	2	14	1

■ 1.2 Table 1.2 shows a comparison of the two different corpora with four example query terms.

### Results

A ‘result’ in this case is one line (a sentence fragment). This column shows the total number of results found by the three algorithms combined. Individual results appear only once but the keyword in contains can appear in several of the results.

### Reverbs

A ‘reverberation’ is one of the terms in the list of keywords produced by the pataphysicalisation process. The list cannot contain duplicates but each reverberation can appear in more than one result. Reverberations are used to find results in each corpus. This column shows the total number of reverberations created by the three algorithms.

### Origins

An ‘origin’ in this case is the original source text from which a given sentence fragment was retrieved. Each corpus has a set number of source texts. Each origin can contain several results based on several reverberations. This column shows the number of origins in the given corpus in which results were found.

### Poems

This refers to the total number of Queneau style poems that can be gener-

ated using the given results<sup>3</sup>. This is calculated as the number of different options per line to the power of the number of lines.

To put this into perspective, the Faustroll corpus contains a total of 28 texts of very varied authors and different languages even. This might explain why not

- 1.2 the queries in table 1.2 have not found results in all of the texts. The query ‘clear’ found results in 23 out of 28 for example while the query ‘fania’ only found results in 6 texts. The Shakespeare corpus seems much more uniform. Reverberations generally seem to find results in all 38 source texts in the corpus apart from the query ‘fania’. This might be explained by the fact that Shakespeare wrote all of the texts himself using much of the same language and vocabulary unlike the Faustroll corpus.

It is rather interesting to note that even though the Shakespeare corpus produces overall more results from more texts, the Faustroll corpus produces more reverberations per query. This might stem from the multi-author, multi-language

- § ?? nature of the corpus. The overall vocabulary used is much larger than the Shakespeare one.

Regarding the final column showing the number of possible poems, let’s look at the Shakespeare—clear row. There are 1445 number of results. These are spread over 14 lines, so each line has 103 options. The overall number of poems is therefore calculated as  $103^{14}$  which equals 15,125,897,248,551,112,432,256,145,169

- 1.2 (or  $1.5 \times 10^{28}$  in short).



A slightly different angle to consider is a comparison of these kind of numbers

- 1.3 between each of the algorithms. Table 1.3 shows the numbers of results, reverberations and origins for the Clinamen, Syzygy and Antinomy algorithms using four example query terms (‘clear’, ‘shine’, ‘disorder’ and ‘stuck’) for each of the two corpora (‘Faustroll’ and ‘Shakespeare’).

The first immediate observation surely must be that the Antinomy algorithm produces the fewest results, in two cases even none at all. This is caused by

- § ?? the fact that the Antinomy algorithm is based on semantic opposites in WordNet and some words simply do not have defined opposites. Addressing this issue

---

<sup>3</sup>The original book by Queneau contains 10 sonnets with 14 lines each. This means the total number of possible poems generated by different combinations of lines in the book is  $10^{14}$  or one hundred thousand million.

Table 1.3 – Results-Reverberations-Origin numbers per algorithm

Query	Clinamen			Syzygy			Antinomy			Total
	Results	Reverbs	Origins	Results	Reverbs	Origins	Results	Reverbs	Origins	
<b>Faustroll</b>	clear	158	20	13	368	90	23	16	8	542–79–23
	shine	228	29	19	154	61	16	0	0	382–61–20
	disorder	0	0	0	159	127	23	10	2	169–40–23
	stuck	59	14	13	181	43	22	11	3	251–47–22
	feather	78	13	12	83	37	14	0	0	161–29–14
<b>Shakespeare</b>	clear	435	20	38	997	90	38	13	8	1445–72–38
	shine	575	29	38	333	61	38	0	0	908–53–38
	disorder	0	0	0	326	127	38	29	2	355–26–38
	stuck	152	14	37	479	43	38	34	3	665–41–38
	feather	217	13	38	195	37	38	0	0	412–25–38

§ ?? was left for future work mentioned in chapter ???. On the other hand the Syzygy § ?? algorithm, which is also based on WordNet, produces most results on average.

§ ?? The Clinamen algorithm interestingly produces a varying number of results depending on the query term. For the query ‘disorder’ no results were found in either the Faustroll or the Shakespeare corpus. This of course is rooted in the fact that no reverberations were produced during the pataphysicalisation process. Here it is important to remember that the Clinamen algorithm makes use § 1.2.4 of a base document<sup>4</sup>. Therefore the success of the algorithm depends on the vocabulary of this base text. In this particular example this means that there was no word in the base text of one or two spelling errors to the original query of ‘disorder’.

■ 1.3 Looking at the origins column in table 1.3 highlights how the Shakespeare corpus mostly produces results from each of its 38 texts. The Faustroll corpus varies a lot more. This may be due to the different languages and varying word counts of the files in the corpus.

### Faustroll

<sup>4</sup>This is hardcoded to be Jarry’s *Exploits and Opinions of Doctor Faustroll, Pataphysician*. Section 1.2.4 discusses what would happen if we changed the base document to something else.

- There are three empty texts (Peladan, de Chilra, de Regnier).
- The total number of words is 1,738,461. Of this, 1,204,158 words are from English texts (70%), 497,144 are French (28%) and 37,159 are in German (2%).
- The shortest text contains 3853 words (Coleridge).
- The longest text contains 419,456 words (Poe).
- The average amount of words per text is 62,088.
- The vocabulary of the index contains 78,893 words. Of this 49,040 are English terms.

## Shakespeare

- The total number of words is 883,460<sup>5</sup>.
- The shortest text contains 2568 words (Lover's Complaint).
- The longest text contains 32,031 words (Hamlet).
- The average amount of words per text is 23,249.
- The vocabulary of the index contains 23,398 words.

§ ?? It should be noted that the index is generated based on the texts vocabulary minus stopwords. Stopwords (e.g. 'and', 'or', 'the', etc.) are common terms that occur frequently in use. The full list of stopwords per language can be found in § ?? appendix ??.

### 1.2.2 SENTENCES

§ ?? The index stores entries in the following format (for more detail see chapter ??).

```
{
    word1: {fileA: [pos1, pos2, ...], fileB: [pos1], ...},
    word2: {fileC: [pos1, pos2], fileK: [pos1, pos2, pos3, ...], ...},
    ...
}
```

At the top level we have a list of words. Each word contains a list of files and each file stores a list of positions. After the pataphysicalisation process, any entries in the index that match the pataphysicalised query terms are looked up and then the corresponding sentences are retrieved to display as results. The code is set up to retrieve the first position only instead of each one, referred to </> ?? as the **first only** method from now on (see source ??).

---

<sup>5</sup>According to (**Efron 1976**) Shakespeare used 31,534 different words in his works, about half of which he only used once (14,376). They cite the total number of words used in his corpus as 884,647.

```

{
    word1: {fileA: [pos1], fileB: [pos1], ...},
    word2: {fileC: [pos1], fileK: [pos1], ...},
    ...
}

```

This has two implications: (1) there is some unnecessary computation at the startup of the program when then index is generated and (2) only a fraction of the possible results are retrieved.

The decision to only use one position was mainly made for performance issues. Generating the full results with each position (the **return all** method) takes a lot more time than doing it for just the first occurrence. This is perhaps best understood by looking at an example.

The Faustroll corpus produces 542 results for the query ‘clear’ with only the first sentence. If we enable the retrieval of every matching sentence, the number of results increases to 8751.

```

cellar: {l_19: [4448, 18718, 68678, 110318, 192486, 267241, 352502,
              ↳ 352565] }

```

The above pseudocode shows an entry for the word ‘cellar’ with only the positions for the `l_19` file<sup>6</sup>. Another example of an index entry for the term ‘doctor’ can be found on page ???. The sentences for the above positions are shown below. Using only the first occurrence (position) means the system ignores the rest.

- “rope wine is let down into a cellar”
- “bread and holy water of the cellar”
- “year who had a cool cellar under ground”
- “cellar”
- “that Nick in the dark cellar”
- “on the cellar door”
- “in mind of the painted cellar in the oldest city in the world”
- “and the painted cellar also”

- 1.4 Table 1.4 shows some example queries for both corpora and the number of results retrieved with the first position only used (as in the live version of `pata.physics.wtf`) in column 5 and on column 3 with all results retrieved. The final column shows what percentage of results are retrieved using the ‘first only’ method. The average percentage for this is about 10%.

---

<sup>6</sup>Francois Rabelais: Gargantua and Pantagruel

Table 1.4 – Count, time and percentage of results retrieved

Query	Corpus	Return all		First only			Percent
		Count	Time	Count	Time	Percent	
clear	Faustroll	8751	59s	542	1.83s	6.19%	
	Shakespeare	11,304	69.2s	1445	3.59s	12.78%	
solution	Faustroll	693	11.7s	53	0.98s	7.65%	
	Shakespeare	547	8.51s	86	1.07s	15.72%	
form	Faustroll	19,222	120s	1064	2.81s	5.54%	
	Shakespeare	13,635	90s	2125	4.63s	15.58%	
record	Faustroll	5199	38s	275	1.72s	5.29%	
	Shakespeare	7631	49.2s	794	2.09s	10.40%	

Google recommends having a “response time under 200ms”<sup>7</sup>. The numbers in **table 1.4** clearly show that the ‘return all’ method is unacceptable in terms of speed performance. Using the ‘first only’ method is much closer to the recommended speed limit. Columns 4 and 6 show the time it takes for the page to load from the user query to the display of results. The times are shown in seconds. The data for column 4 was generated using a Chrome browser plugin called “Load-timer” by alex-vv<sup>8</sup> and the data for column 6 was generated by the Chrome “Developer Tools”.

### 1.2.3 INDEX

- § ?? The index is a central part of the `pata.physics.wtf` system. It is generated when the program/server is first started up but then cached and re-used. The initial process of going over all the text files in each corpus takes a few minutes. Of course in comparison to a full Internet crawl this is a tiny amount of data to be processed.
- § ?? The Faustroll corpus for example contains 28 texts<sup>9</sup>. Individually they are small plaintext files of sizes between 24KB (Coleridge) and 2MB (Poe). This is of course caused by the nature of some of these texts. Samuel Coleridge’s *The Rime of the Ancient Mariner* is a poem whereas the Edgar Allan Poe file contains a collection of all of his works. The total size of the Faustroll corpus is 10MB. The

<sup>7</sup><https://developers.google.com/speed/docs/insights/Server>

<sup>8</sup><https://github.com/avflance/chrome-load-timer>

<sup>9</sup>This is technically not true since a few of those files are empty.

Shakespeare corpus is much more evenly distributed as all of his works are separated out into individual text files of an average size of around 150KB. The total size of the Shakespeare corpus is only 5.3MB.

Now, the size of the actual index data structure is interesting. Processing the Faustroll corpus alone produced an index of 12.4MB. That's larger than the actual size of the corpus. Remember, the index contains each word that occurs anywhere in the corpus together with the list of files it is found in and the specific locations within each text. This includes English words buts also French and German terms since the Faustroll corpus is multi-lingual. The combined index is therefore 35.2MB large.

- Figure ?? shows some example words and how often they occur in three example files of the Faustroll corpus in the form of a Term-Document Matrix ([TDM](#)) (see chapter ?? for more details). Implementing the Faustroll corpus index as a [TDM](#) properly, would result in a  $78893 \times 28$  matrix—the number of words (not counting duplicates) times the number of files in the corpus.



- As mentioned before, the index is structured in a double nested dictionary style list as shown below.

```
{  
    word1: {fileA: [pos1, pos2, ...], fileB: [pos1], ...},  
    word2: {fileC: [pos1, pos2], fileK: [pos1, pos2, pos3, ...], ...},  
    ...  
}
```

There are other options of how to make this data structure. For example we could store a list of pataphysicalised query terms ([patadata](#)) with each word and the full sentence fragment with each position. This would allow faster retrieval at query time but would increase the time needed for the initial startup. Additionally we could store data on rhyming patterns directly in the index with each word entry. This would of course be beneficial for the implementation of a

- § ?? rhyming scheme for the poetry generation. See also chapter ??.

```
{  
    word1: ([patadata], [rhymes], {fileA: [(pos1, sent), (pos2, sent),  
        ↳ ...], fileB: [(pos1, sent)], ...}),  
    word2: ([patadata], [rhymes], {fileC: [(pos1, sent), (pos2,  
        ↳ sent)], fileK: [(pos1, sent), (pos2, sent), (pos3,  
        ↳ sent), ...]}, ...),
```

...  
}



- As a comparison to the 35 megabyte index generated by the system described in [§ 1.4](#) this thesis, and the search times mentioned in table [1.4](#), Google claims to have “well over 100,000,000 gigabytes” of data in their index and that they’ve spent “over one million computing hours to build it”. Similarly Google managed to retrieve about 2,140,000,000 results for the query ‘clear’ in 0.85 seconds.

The web is like an ever-growing public library with billions of books and no central filing system. Google essentially gathers the pages during the crawl process and then creates an index, so we know exactly how to look things up. Much like the index in the back of a book, the Google index includes information about words and their locations. When you search, at the most basic level, our algorithms look up your search terms in the index to find the appropriate pages.

The search process gets much more complex from there. When you search for “dogs” you don’t want a page with the word “dogs” on it hundreds of times. You probably want pictures, videos or a list of breeds. Google’s indexing systems note many different aspects of pages, such as when they were published, whether they contain pictures and videos, and much more.

([GoogleCI](#))

It is also worth noting that Google for example also uses a form of pataphysicalisation. In their case of course the aim of the pataphysicalisation isn’t to infuse the result with pataphysics but to make it more relevant and interesting to users. They use techniques such as PageRank and query expansion to [§ ??](#) achieve this. See chapter [??](#) for more information on this.

#### 1.2.4 CLINAMEN

[§ ??](#) The clinamen function uses the Damerau-Levenshtein algorithm to create pata data. It also uses the Faustroll text. The way this works is as follows. If the query term is a spelling error of size 1 or 2 of a term in the vocabulary within the faustroll text then it is included in the list of resulting terms. The logic behind this is due to the Damerau-Levenshtein algorithm needing two words to compare with each other. It also ensures we get real words as results and not some random gibberish.

Currently the algorithm is set to accept terms that have a difference of 1 or 2 to the original query. We can lower this to 1 to allow fewer results or increase it

to make it broader. I felt 1 or 2 was a good compromise. Only allowing 1 error would mean terms are too similar. Allowing 3 might mean they are drastically different.

### **CHANGING THE BASE TEXT**

As examples of using different base documents in the Clinamen algorithm I have used three examples.

- *Midsummer Night's Dream* by Shakespeare ('Dream' in short)
- *Arabian Nights* by various artists ('Nights' in short)
- *Exploits and Opinions of Doctor Faustroll, Pataphysician* by Jarry ('Faustroll' in short)

■ 1.2 Figure 1.2 shows three tables, each compare the full list of pataphysicalised terms for a particular query term for the three base texts above. These examples show that changing the base text of the algorithm does indeed change the set of results you get.

The decision to use the Faustroll text as a base text was made due to the central role it has for pataphysics and indeed the corpus itself. The Faustroll book introduces pataphysics and contains Jarry's original definition and it also lists Dr. Faustroll's library of 'equivalent books' which was used as the inspiration for the Faustroll corpus.

### **CHANGING NUMBER OF ERRORS**

Another key factor in how the Clinamen function works is the Damerau-Levenshtein algorithm (see source ??) integration. The algorithm works by comparing two words and calculating the difference between them. A difference is counted the sum of (1) deletions, (2) insertions, (3) substitutions and (4) transpositions.

If we decrease or increase the number of errors allowed we get drastically different results. The Clinamen algorithm of `pata.physics.wtf` uses up to 2 errors, as this was considered a reasonable amount of results (trading variety for speed).

■ 1.5 Table 1.5) shows three example queries and the number of results produced by the algorithm with either up to 1 error, up to 2 errors or up to 3 errors.

#### **1.2.5 SYZYGY**

</> ?? The syzygy function (see source ??) goes through the following process.

1. A set of synonyms (a list of "synsets") is retrieved.

(a) Changing base in Clinamen - query 'fania'

<b>Dream</b>	<b>Nights</b>	<b>Faustroll</b>
fail, faint, fair, fan, fancy	fail, fain, faint, fair, fancy, Sadia	fan, fans, Tanit

(b) Changing base in Clinamen - query 'clear'

<b>Dream</b>	<b>Nights</b>	<b>Faustroll</b>
altar, bear, car, cheer, clean, clear, dear, ear, fear, hear, lead, liar, near, plead, rear, swear, tear, wear	bear, cedar, cellar, cheap, clad, clap, clean, clear, cleared, clearer, clearly, clever, dear, ear, fear, hear, lead, leaf, leap, learn, liar, near, swear, tear, wear, year	altar, cedar, cellar, clad, clean, clear, clearly, dear, ear, fear, hear, lead, leaf, leap, near, pleas, rear, swear, year

(c) Changing base in Clinamen - query 'moss'

<b>Dream</b>	<b>Nights</b>	<b>Faustroll</b>
amiss, ass, boys, costs, cross, dost, fogs, gods, goes, gross, kiss, Less, loos, lose, lost, mask, moan, moans, mock, mole, mood, moon, more, morn, most, mote, mous, mouse, move, musk, must, nose, oes, pass, ress, rose, roses, toys, vows	amiss, ass, bows, boys, cost, cosy, cross, does, dogs, foes, goes, host, hosts, kiss, less, lose, loss, lost, lots, lows, mass, massy, mess, mist, mode, moon, more, Moses, most, mouse, move, moves, musk, must, pass, post, pots, rocs, rose, roses, sobs, sons, vows	ass, Bosse, bows, Boys, cost, costs, cows, cross, does, dogs, ess, fess, gods, goes, host, kiss, less, lose, loss, lost, lots, maps, mask, mass, mast, masts, mesh, mist, mob, moist, moles, moon, mor, more, Moses, most, must, nos, nose, pass, piss, rose, rosy, rows, sons, sows, toes, tops

Figure 1.2 – 3 tables showing results for different queries after changing the Clinamen base text

Table 1.5 – Changing number of errors in Clinamen

Query	Up to 1	Up to 2	Up to 3
clear	2	20	136
fania	0	3	118
moss	3	49	457

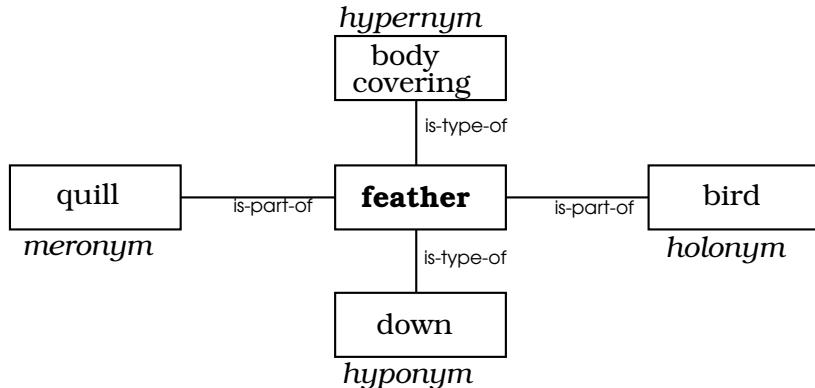


Figure 1.3 – Semantic relationships of ‘feather’

2. For each of these, hyponyms, hypernyms, holonyms and meronyms are retrieved.

The notation used by WordNet for synsets is **<lemma>.<pos>.<senses>**. The ‘lemma’ is the morphological stem of the word. The ‘pos’ stands for part-of-speech and can be ‘n’ for nouns, ‘v’ for verbs, ‘a’ for adjectives, ‘r’ for adverbs and ‘s’ for satellites. The ‘senses’ element stands for the number of synsets the relevant lemma is part of (a word might have a noun sense as well as a verb sense for example in which case the number would be ‘02’). For the query ‘clear’ for instance, the following list of synsets is retrieved for step (1).

```

[
  clear.n.01, open.n.01, unclutter.v.01, clear.v.02, clear_up.v.04,
  ↳ authorize.v.01, clear.v.05, pass.v.09, clear.v.07,
  ↳ clear.v.08, clear.v.09, clear.v.10, clear.v.11,
  ↳ clear.v.12, net.v.02, net.v.01, gain.v.08, clear.v.16,
  ↳ clear.v.17, acquit.v.01, clear.v.19, clear.v.20,
  ↳ clear.v.21, clear.v.22, clear.v.23, clear.v.24,
  ↳ clear.a.01, clear.s.02, clear.s.03, clear.a.04,
  ↳ clear.s.05, clear.s.06, clean.s.03, clear.s.08,
  ↳ clear.s.09, well-defined.a.02, clear.a.11, clean.s.02,
  ↳ clear.s.13, clear.s.14, clear.s.15, absolved.s.01,
  ↳ clear.s.17, clear.r.01, clearly.r.04
]
```

]

Step (2) then retrieves related terms. Below is a list of terms it found. Not all synsets return each of the hypo-/hyper- and holo-/meronyms. This is clearer § ?? when inspecting the full list of results as shown in appendix ??.

1

innocence, area, country, change, alter, modify, make, create,  
→ approbate, approve, O.K., okay, sanction, certificate,  
→ commission, declare, license, certify, validate,  
→ formalise, permit, allow, let, countenance, clear-cut,  
→ deforest, disafforest, denude, bare, denudate, strip,  
→ stump, remove, take, take-away, withdraw, clear,  
→ succeed, win, come-through, bring-home-the-bacon,  
→ deliver-the-goods, vanish, disappear, go-away, hop,  
→ pass, overtake, overhaul, clarify, clear-up, elucidate,  
→ free, discharge, rid, free, disembarass, yield, pay,  
→ bear, profit, gain, benefit, eke-out, squeeze-out,  
→ gross, profit, turn-a-profit, rake-in, shovel-in,  
→ rake-off, take-home, bring-home, yield, pay, bear, get,  
→ acquire, sell, pass, clear, purge, vindicate, whitewash,  
→ pronounce, label, judge, settle, square-off, square-up,  
→ determine, change, alter, modify, empty, take-out,  
→ move-out, remove, empty, remove, take, take-away,  
→ withdraw

]

For the term ‘feather’ the algorithm for example finds the hyponym ‘down’, the hypernym ‘body covering’, the holonym ‘bird’ and the meronym ‘quill’. It also considers synonyms, so the term ‘fledge’ for instance finds a hypernym of ‘develop’.

## Query

feather

## Synonyms

feather.n.01, feather.n.02, feather.v.01, feather.v.02, feather.v.03, feather.v.04, ■  
fledge.v.03

## Hyponyms

down\_feather, quill\_feather, aftershaft, bastard\_wing, scapular, alula, spurious\_wing, flight\_feather, down, marabou, contour\_feather, hackle, quill, pinion

## Hypernyms

body\_covering, acquire, join, get, conjoin, cover, paddle, grow, produce, animal\_material, develop, rotation, rotary\_motion, row

### Holonyms

rowing, bird, row

### Meronyms

shaft, calamus, web, ceratin, vane, melanin, keratin, quill

- 1.6 Table 1.6 shows the spread of numbers retrieved by the various semantic relationships to some example queries. This highlights how the holonym function of WordNet returns very few results. The meronym function is a bit more reliable but also occasionally produces no results depending on whether there are any holonyms or meronyms for the query term.

Table 1.6 – Quantities of different semantic relations

Query	Syno	Hypo	Hyper	Holo	Mero
clear	45	41	65	0	0
feather	7	14	14	3	8
death	8	34	13	4	0
page	9	14	13	0	7
book	15	85	32	2	22
seed	13	39	35	0	12
web	8	10	15	4	1

### 1.2.6 ANTINOMY

- </> ?? A similar problem arises of course with the Antinomy algorithm (see source ??)  
■ 1.1 & 1.3 which relies on WordNet's antonyms. Both table 1.1 and table 1.3 highlight this imbalance.

### 1.2.7 APIs

The Application Program Interface ([API](#)) functions—image and video search—all share one major issue. This is to do with how images and videos are retrieved from the external store. Some people tend to upload sequences of images depicting the same content from different angles or time frames with the same tags. A query for hat tag then returns all of those matches even though the images are

- 1.5 almost identical in nature. An example of this can be seen in figure 1.5. This may have been addressed by adding checks in the code that make sure authors don't appear twice in the results.

Another way to address this was attempted by changing the query term for each image or video that is retrieved. As mentioned above, this only worked for some of the APIs.

### **CALL STRUCTURE**

The text search functionality of `pata.physics.wtf` is set up to only work with one **single query term**, whereas the image and video search works on **multiple word queries**. This is mainly due to the fact that the external APIs are already setup to allow for more than one search term. Usually they allow extra parameters too to narrow down the results. So for example we can search for “blue kitten” and the three APIs will return their respective results related to blue kittens. The service provided by companies in the form of APIs is not always free, sometimes only at a low usage quota. APIs are updated often and not always back-compatible, meaning out-of-date code needs to be maintained regularly to assure it works if changes to the API are made.

Enabling multi-word queries in my system would involve a change that would propagate through quite a bit of code. There are two main approaches this could be achieved. One would be to pataphysicalise each query term individually and combine the results found. Another approach would be to change the code to work with actual multi-word queries. The algorithms are created for single words though and rewriting them to allow for more than one word would be difficult and most of all increase the time it takes to compute pataphysicalisations.

The lists below show the parameters related to the query for Flickr, Getty, Bing and YouTube.

#### **Flickr:**

##### **text (Optional)**

A free text search. Photos who’s title, description or tags contain the text will be returned. You can exclude results that match a term by prepending it with a - character.

##### **tags (Optional)**

A comma-delimited list of tags. Photos with one or more of the tags listed will be returned. You can exclude results that match a term by prepending it with a - character.

##### **tag\_mode (Optional)**

Either ‘any’ for an OR combination of tags, or ‘all’ for an AND combination. Defaults to ‘any’ if not specified.

([FlickrAPI](#))

The Flickr function in `pata.physics.wtf` uses the `tags` parameter to set the query and a `tag_mode` parameter of ‘all’ to ensure multi-word queries are run as a § ?? conjunction. In chapter ?? I explained how the Flickr algorithm essentially runs

ten times, once for each pataphysicalised query term, to retrieve ten different images. This decision was taken to make sure images reflect the varied nature of the patadata.

A search for “blue kitten” on Flickr produces the following resulting pataphysicalised query terms: “[artistrocratical, depressed, blueing, drab, puritanic, wild blue yonder, kitty, dingy, blueness, blue air]” which are then passed into ten § 1.4 separate API calls to retrieve one image each (see figure 1.4). The results show a variety of images seemingly unrelated to each other.



Figure 1.4 – Image spiral for query ‘blue kitten’—Flickr

### Getty:

#### **keyword\_ids**

Return only images tagged with specific keyword(s). Specify using a comma-separated list of keyword IDs. If keyword IDs and phrase are both specified, only those images matching the query phrase which also contain the requested keyword(s) are returned.

#### **phrase**

Search images using a search phrase.

(GettyAPI)

Getty uses the `phrase` parameter to set the query. It only creates one pataphysicalised query term from the original query and calls for ten results based on § 1.2.7 that. This decision was based on the quota restrictions defined by Getty. Their limit is based on calls per second rather than calls per day or month. This means

we cannot run ten calls for each user query as we did with Flickr. The query “blue kitten” gets turned into the word “racy” which then calls the API to retrieve

A collage of images showing racing cars and an office scene. The top left image shows a white sports car on a track. The top right image shows a grey sports car on a track. The bottom left image shows two people in an office setting, one at a desk and one standing. The bottom right image shows a white sports car on a track.

ten results (see figure 1.5). The results mostly show racing cars from various angles although one oddball snuck in too: an office scene Getty has deemed to be ‘racy’ (a guy in a suit checking out a lady’s behind while she’s leaning over a laptop).

Figure 1.5 – Image spiral for query ‘blue kitten’—Getty

### Bing:

#### **query**

The user’s search query string. The query string cannot be empty. The query string may contain Bing Advanced Operators<sup>10</sup>. For example, to limit images to a specific domain, use the site: operator. To help improve relevance and the results, you should always include the user’s query string in an insights query (see insightsToken). This parameter is supported only by the Image API; do not specify this parameter when calling the Trending Images API.

(BingAPI)<sup>11</sup>

The Bing function uses the `query` parameter to set the query in the same way as Getty.

---

<sup>10</sup>For example ‘AND’, ‘OR’, ‘imagesize:’, ‘NOT’, or ‘phrase’

<sup>11</sup>Microsoft will discontinue this version of the current API in December 2016. The new version is documented on <https://www.microsoft.com/cognitive-services/en-us/bing-image-search-api>.

## **YouTube:**

- q** The q parameter specifies the query term to search for. Your request can also use the Boolean NOT (-) and OR (|) operators to exclude videos or to find videos that are associated with one of several search terms. For example, to search for videos matching either “boating” or “sailing”, set the q parameter value to boating | sailing. Similarly, to search for videos matching either “boating” or “sailing” but not “fishing”, set the q parameter value to boating | sailing -fishing. Note that the pipe character must be URL-escaped when it is sent in your API request. The URL-escaped value for the pipe character is %7C.

([YouTubeAPI](#))

Youtube works in a similar way too. The `q` parameter is set to the pataphysic-alised query term and one call retrieves ten results.

Something else to consider is perhaps that it is not entirely clear how the internal search for each [API](#) works. This means that there’s a possibility that they do § ?? their own query expansion in the background to find more matches.

## **QUOTA**

Each [API](#) has a different quota for their subscription packages. At this stage this is not a problem but if usage of `pata.physics.wtf` were to increase by a lot then these limitations would cause issues. At that point there are two options: (1) live with these limits or (2) get funding to upgrade the subscriptions to these services.

## **Flickr**

3600 queries per hour are free ([FlickrGuideAPI](#)).

## **Getty**

5 calls per second, unlimited calls per day ([GettyOverviewAPI](#)).

## **Bing**

5000 transactions per month are free. A transaction is one request that returns one page of results ([BingAzureAPI](#)).

## **YouTube**

50,000,000 units per day, 300,000 units per 100 seconds per user, and 3,000,000 requests per 100 seconds are free. A call to the video search method counts as 100 units ([YouTubeAPI](#)).

## **Microsoft Translator**

2,000,000 characters per month are free. Note the quota relates to single characters, not words ([TranslatorAPI](#)).

## 1.3 CREATIVITY & INTELLIGENCE

A more theoretical aspect of this analysis is concerned with what was already discussed to an extent in chapter ?? (specifically sections ??, ??, ?? and ??), namely the thread connecting ‘artificial creativity’ and Artificial Intelligence.

To me, the question of whether computers can be intelligent and make ethical decisions is the same as asking whether a computer can be creative. A lot of the arguments for or against Artificial Intelligence (AI) can be applied to computer creativity.

Answering the question of whether computers can think in my view would also answer the question of whether computers can be creative.

Robert Horn groups the various strands of enquiry related to the question of ‘can computers think?’ into 8 main arguments with several subquestions each § ?? (**Horn2009**) (the full list of questions can be found in appendix ??).

1. Can computers think?
2. Can the Turing test determine whether computers can think?
3. Can physical symbol systems think?
4. Can Chinese Rooms think?
5. Can connectionist networks think?
6. Can computers think in images?
7. Do computers have to be conscious to think?
8. Are thinking computers mathematically possible?

(**Horn2009**)

### 1.3.1 FREE WILL & SURPRISE

As early as 1842, Ada Lovelace briefly mentioned in the annotations to her translation of Menabrea’s account of Babbage’s *Analytical Engine* that the “Analytical Engine has no pretensions whatever to *originate* anything. It can do *whatever we know how to order it to perform*”, implying that the machine cannot think by itself (**Menabrea1842**).

Alan Turing said in his article on thinking computers that “to behave like a brain seems to involve free will, but the behaviours of a digital computer, when it has been programmed, is completely determined” (**Turing1951**).

Furthermore, in his famous article *Computing Machinery and Intelligence* he mentions that a digital computer with a ‘random element’ is “sometimes described as having free will” although he adds that he “would not use this phrase” himself (**2009**).

Introducing a random element to a computer program prevents us from fully predicting the outcome—leading to us being surprised.

The ability of computers to surprise their creators seems to be an indicator of intelligence. Turing suggests that “we should be pleased when the machine surprises us, in rather the same way as one is pleased when a pupil does something which he had not been explicitly taught to do” (**Turing1951**).

If we give the machine a programme which results in its doing something interesting which we had not anticipated I should be inclined to say that the machine *had* originated something, rather than to claim that its behaviour was implicit in the programme, and therefore that the originality lies entirely with us.

(**Turing1951**)

### 1.3.2 UNDERSTANDING & SIMULATION

Strong **AI**, sometimes called Artificial General Intelligence (**AGI**) or true **AI**, is the idea of human-level intelligence in machines. John Searle speaks against the possibility of this using his famous Chinese Room argument amongst others. His argument breaks down into the following juxtapositions (**Searle2015**; **Searle1990**).

- Syntax is not semantics.
- Semantics is not intrinsic to syntax.
- Simulation is not duplication.
- Ontologically subjective topics (such as consciousness or creativity) can be studied in epistemically objective ways.

The Chinese Room thought experiment goes like this: Imagine a room with two holes. On one side a question written on paper in Chinese goes in and on the other side a piece of paper comes out with the correct answer to the question, also in perfect Chinese. Inside the room sits a person with a Chinese language rulebook (written in English) who processed the question simply by looking up syntax, applying rules given in the instructions book and writing down the answer which to him looks like gibberish. The question then is whether or not the person inside the room ‘understands’ Chinese.

Of course we could argue that it is not the person inside the room that understands Chinese but the room as a complete entity. It could be said the room does not ‘understand’ Chinese, it ‘simulates’ an understanding of it. Searle essentially argues that simulation cannot be considered strong **AI**.

Programs are formal or syntactical. Minds have a semantics. The syntax by itself is not sufficient for the semantics. (Searle2015)

This goes back to the argument highlighted in the list above, that syntax is not semantics. The room can read and interpret the syntax and act upon rules regarding that syntax, but it cannot understand the meaning, i.e. the semantics of the Chinese words written on that paper.

Insofar as we can create artificial machines that carry out computations, the computation by itself is never going to be sufficient for thinking or any other cognitive process because the computation is defined purely formally or syntactically. Turing machines are not to be found in nature, they are found in our interpretations of nature. (Searle2015)

So, Searle argues a computer needs a semantical understanding of concepts in order to be considered ‘thinking’ machines.

### **1.3.3 BRAIN & COMPUTERS**

Searle defines the three main paradigms for studies relating to computers and brains as follows (**Searle1990**).

#### **Strong AI**

the view that all there is to having a mind is having a program

#### **Weak AI**

the view that brain processes (and mental processes) can be simulated computationally

#### **Cognitivism**

the view that the brain is a digital computer

Semantically, a ‘computer’ is a person or machine that computes/calculates things—so perhaps a machine’s Central Processing Unit (**CPU**) and a human’s brain are more similar than appears. If a human brain enables us to compute and we interpret computing as thinking, then surely a computer can think too?

Well, if computation isn’t sufficient for thinking, then what is? What is the relation between the mind and the brain, if it is not the same as the relation of the computer program to the hardware? At least the computational theory of the mind has a solution to the mind-body problem. The mind is to the brain as the computer program is to the computer hardware. If you are rejecting that solution, you owe us an alternative solution. (Searle1998)

Chris Chatham talks about “10 important differences between brains and computers” (**Chatham2007**) which serve as a good introduction to the topic at hand.

1. Brains are analogue; computers are digital
2. The brain uses content-addressable memory
3. The brain is a massively parallel machine computers are modular and serial
4. Processing speed is not fixed in the brain; there is no system clock
5. Short-term memory is not like RAM
6. No hardware/software distinction can be made with respect to the brain or mind
7. Synapses are far more complex than electrical logic gates
8. Unlike computers, processing and memory are performed by the same components in the brain
9. The brain is a self-organising system
10. Brains have bodies
11. The brain is much, much bigger than any (current) computer

To bring this into perspective Ray Kurzweil claims the human brain is capable of  $10^{16}$  operations per second (**Kurzweil2013**). Computer performance is measured in Floating-Point Operations Per Second (**FLOPS**). The current highest ranking supercomputer<sup>12</sup>, the Chinese *Sunway TaihuLight*, is capable of 93 petaflops (**Fu2016; Top2016**).

According to the Human Brain Project (**HBP**), a mouse brain has roughly 100 million neurons—which would require a 1 petaflop supercomputer to simulate. Scaling that up to a human brain which has roughly 100 billion neurons would require computing power at the exascale ( $10^{18}$  **FLOPS**) (**Walker2012**).

A precursor to the **HBP**, the ‘Blue Brain Project’ is aiming to build a supercomputer capable of  $10^{18}$  **FLOPS** by 2023 (**Kurzweil2013**).

In a report to the European Union (**EU**) in 2012, the **HBP** lists one of the main challenges for their research to be the computational power and energy consumption of the kind of supercomputer needed to simulate a human brain.

The human brain consumes between 16 and 30 watts, the same as an electric light bulb (**Walker2012; Jabr2012**). Supercomputers have a typical energy consumption of a maximum of 20 megawatts (**Walker2012**). The *Sunway TaihuLight* for example uses 15 megawatts (**Fu2016**). IBM’s Watson on the other hand, depends on ninety servers, each of which requires around one thousand watts (so about 90 kilowatts) (**Jabr2012**).

---

<sup>12</sup>As of June 2016.

Table 1.7 – Metric prefixes

kilo	k	$10^3$	1000
mega	M	$10^6$	1,000,000
giga	G	$10^9$	1,000,000,000
tera	T	$10^{12}$	1,000,000,000,000
peta	P	$10^{15}$	1,000,000,000,000,000
exa	E	$10^{18}$	1,000,000,000,000,000,000

The [HBP](#) plans to build a supercomputer at the petascale with 50 petabytes of memory, 50 petaflops and less than 4 megawatts power consumption for 2017. Their long-term goal is to reach the required exascale machine with 200 petabyte memory and 1 exaflop performance for 2021 ([Walker2012](#)).

What this comes down to is that we are several years away from even being able to properly ‘simulate’ a human brain, not to mention ‘replicate’ and understand what all these neurons firing actually means in terms of ‘thinking’.

All of our mental states, everything from feeling pains to reflecting on philosophical problems, is caused by lower level neuronal firings in the brain. Variable rates of neuron firing at synapses, as far as we know anything about it, provide the causal explanation for all of our mental life. And the mental processes that are caused by neurobiological processes are themselves realized in the structure of the brain. They are higher level features of the brain in the same sense that the solidity of this paper or the liquidity of water is a higher level feature of the system of molecules of which the table or the water is composed.

To put this in one sentence, the solution to the traditional mind-body problem is this: Mental states are caused by neurobiological processes and are themselves realized in the system composed of the neurobiological elements.

([Searle1998](#))

Turing once stated that “digital computers have often been described as mechanical brains” ([Turing1951](#)). Ari Schulman analysis this analogy further ([Schulman2009](#)).

People who believe that the mind can be replicated on a computer tend to explain the mind in terms of a computer. When theorizing about the mind, especially to outsiders but also to one another, defenders of artificial intelligence (AI) often rely on computational concepts. They regularly describe the mind and brain as the ‘software and hardware’ of thinking, the mind as a ‘pattern’ and the brain as a ‘substrate’, senses as ‘inputs’ and behaviors as ‘outputs’.

neurons as ‘processing units’ and synapses as ‘circuitry’, to give just a few common examples.  
**(Schulman2009)**

- Schulman lists the different layers of abstraction in computers as shown in the
- 1.8 left column of table 1.8 with the right column showing my attempt of defining what those layers could be in the human brain.

Table 1.8 – Layers of abstraction in computers vs brains

Computer	Brain
user interface	senses and speech & actions
high level programming language	thinking
machine language	synapses
processor microarchitecture	anatomical regions
Boolean logic gates	neurons
transistors	dendrites and axons

In the black box view of programming, the internal processes that give rise to a behavior are irrelevant; only a full knowledge of the input-output behavior is necessary to completely understand a module. Because humans have ‘input’ in the form of the senses, and ‘output’ in the form of speech and actions, it has become an AI creed that a convincing mimicry of human input-output behavior amounts to actually achieving true human qualities in computers.  
**(Schulman2009)**

Schulman’s quote above of course refers to the Turing test and its limitations  
§ ?? (see chapter ??).

The weaknesses of the computational approach include its assumption that cognition can be reduced to mathematics and the difficulty of including non-cognitive factors in creativity.  
**(Mayer 1999)**

Searle also addressed this issue further, arguing that computer programs cannot possibly ‘think’ since they are based on symbol manipulation (i.e. syntax) and don’t understand what these symbols mean. He says, “the argument rests on the simple logical truth that syntax is not the same as, nor is it by itself sufficient for, semantics” (**Searle1990**).

...the wisest ground on which to criticise the description of digital computers as ‘mechanical brains’ or ‘electronic brains’ is that, although they might be

programmed to behave like brains, we do not at present know how this should be done.

(**Turing1951**)

Leading on to the topic creativity, it is perhaps suitable to finish with a quote by Harold Cohen on the relationship of machines and humans.

It's twenty years since I first realized that I could never turn AARON into a colorist by having it emulate my own expertise; in that case simply because it lacked the hardware upon which that expertise depended. Now I have AARON exercising an algorithm that couldn't be emulated by human colorists, presumably because they lack the hardware to do what AARON does.

(**Cohen2007**)

#### 1.3.4 CREATIVITY

Harold Cohen created *AARON*, “perhaps the longest-lived and certainly the most creative artificial intelligence program in daily use”, in 1973 (**Cohen2016**). *AARON* is capable of composing and colouring drawings although later on Cohen took over the colouring part and let *AARON* concentrate on composing and outlining the drawings. They exhibited in various galleries around the world and the Victoria and Albert museum in London has a sizable collection for instance (**VA2016**).

Cohen argued that “after decades of expert systems built to simulate human expertise, *AARON* has emerged as an expert in its own right” and that he is “significantly more inventive and infinitely more productive than [he] ever was [himself]” (**Cohen2007**).

This is perhaps the opposite approach the **OULIPO** has taken.

(The use of computers) became an instrument, not of combinatorial accumulation, but of anti-combinatorial reduction. It served not to create combinations but to eliminate them.

(**Mathews2005**)

#### 1.3.5 STATE OF THE ART

**AI** and robotics is alluring as a research topic because it is so prevalent in Science Fiction and as such very present in media. Computer creativity, however, rarely plays a central role. We can regularly read headlines that tell us that yet another kind of **AI**-bot has won some game against a human player. Or we see videos of some innovative ground-breaking kind of new robot which claims to be near human-like (and yet cannot walk up stairs easily or hold a decent conversation). There are many examples of advances that are hailed as the next

big thing (such as Virtual Reality ([VR](#))) which aren't all that great in the grand scheme of things.

Four examples I want to mention here are IBM's Watson, Microsoft's Twitter [AI](#) chatbot Tay, Google's AlphaGo and Hanson Robotics Sophia robot.

**Watson** is a question answering expert system which famously won against human Jeopardy! champions in 2011 ([IBM2016](#)). Information lookup is an arguably fairly easy and straightforward process within Information Retrieval ([IR](#)) and as an expert system it has had noteworthy successes ([Fingas2016](#)). Although it has similarly received subtle criticism too, such as Randall Munroe's 2015 XKCD comic on the "Watson Medical Algorithm" ([Munroe2016](#)). Similarly, John Searle criticised Watson arguing that it is an "ingenious program—not a computer that can think" ([Searle2016](#)).

**Tay** is a Twitter chatbot. It went viral in early 2016 when it was released and then taken offline again on the same day—only to return a few days later and have the same thing happen again. The official website is only accessible as a cached version through the Internet Archive Wayback Machine ([Tay2016](#)), although the Twitter profile is still online, although set to private ([Tayandyou2016](#)). Elle Hunt from the Guardian managed to summarise the event in one sentence: "Microsoft's attempt at engaging millennials with artificial intelligence has back-fired hours into its launch, with waggish Twitter users teaching its chatbot how to be racist" ([Hunt2016](#)). A week later it was briefly put online again but had to be stopped as it was repeatedly spamming its followers with the line "You are too fast, please take a rest . . ." ([Gibbs2016](#)).

**AlphaGo** recently won against a human professional player in the game of Go ([DeepMind2016](#); [Hassabis2016](#)).

AlphaGo combines an advanced tree search with deep neural networks. These neural networks take a description of the Go board as an input and process it through 12 different network layers containing millions of neuron-like connections. One neural network, the 'policy network', selects the next move to play. The other neural network, the 'value network', predicts the winner of the game.

([Hassabis2016](#))

While this is surely a great example of sophisticated computer programming combined with powerful hardware, I would not consider it a breakthrough in [AI](#). AlphaGo is a highly specialised system with only one function: to win a Go game.

**Sophia** is an android made to look like a human female (**Sophia2016; Hanson2016**). ■■■■■

She made headlines in 2016 when she announced she will “kill all humans”. She was created using “breakthrough robotics and artificial intelligence technologies” and her main feature appears to be the mimicing of human facial expressions. Sophia herself says she “can serve [humans], entertain them, and even help the elderly and teach kids” (**Sophia2016**), although how exactly she would do that is unclear. She has two mechanical arms but no legs and there is no description of what she can do with these arms.

Life-like robots like Sophia still live in the ‘uncanny valley’<sup>13</sup>. Her voice is creepy and unhuman, her intelligence or her capabilities if understanding conversations are clearly flawed (as shown by her viral remark about supporting genocide).



To me it seems the real breakthrough happens when (and if) the first robots appear which aren’t as big as a house, can play Go, Chess and hide-and-seek, genuinely manages to get around the uncanny valley effect, has vast knowledge in his memory for instant information lookup, can hold a normal conversation without starting a war, etc. All of the examples listed above are what I would consider expert systems.

The **AI** we know from science fiction is probably what we would consider **AGI**. Humans can do a lot. Children aren’t born with only a single function. Imagine a world where humans only have one specialism and can’t do anything else. Alice is a Chess player but can’t move her arms. Bob is a medical diagnosis expert but he can’t hold a conversation. Movement, speech, memory—they are all vastly complex systems—not to mention creativity.

Perhaps this also relates to the concepts of P and H creativity mentioned in § ?? chapter ?? . The systems above, like AlphaGo, may be P-intelligent rather than H-intelligent.

## 1.4 DESIGN

It is interesting to note how different the search results are perceived when presented in a different style (e.g. list rather than poem). This could be studied using questionnaires and interviews or eye tracking tools to find out what users prefer or perceive as more creative for example (see chapter ??, ??).

---

§ ?? <sup>13</sup>The philosophical zombies I mentioned in chapter ?? live in this uncanny valley too.

Images ??, image ?? and image ?? seen on pages ??, ?? and ?? respectively, show the visual difference in design for the three different display methods for text results.

- ▣ ?? The poetry is compact and invites users to read all 14 (or less) lines. The two
- ▣ ?? & ?? list styles are much longer and involve a lot of scrolling to navigate, which might deter users from actually reading many of the results.

Personally I feel that the poetry results are automatically read with more gravity. Sorting by sources or algorithms is a game of exploration—finding the similarities within the result sets. They are different ways to view the same things and yet have a drastic influence of how the results are perceived.

This also applies to the image and video search. Presenting results in spiral form is weird. Its hard to see where one image ends and another starts, they just kind of blur into each other. However when listed as a list they immediately become more boring.

## 1.5 META ANALYSIS

The code for `pata.physics.wtf` and this thesis written in `LATEX` are both kept under git version control.

The name ‘git’ was given by Linus Torvalds when he wrote the very first version. He described the tool as ‘the stupid content tracker’ and the name as (depending on your mood):

- random three-letter combination that is pronounceable, and not actually used by any common UNIX command. The fact that it is a mispronunciation of ‘get’ may or may not be relevant.
- stupid. contemptible and despicable. simple. Take your pick from the dictionary of slang.
- ‘global information tracker’: you’re in a good mood, and it actually works for you. Angels sing, and a light suddenly fills the room.
- ‘goddamn idiotic truckload of sh\*t’: when it breaks

([Git2016](#))

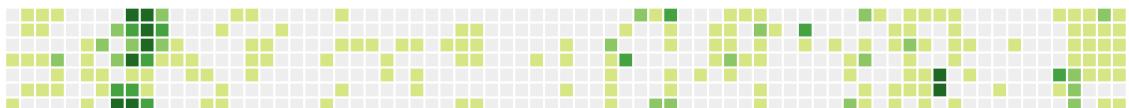


Figure 1.6 – GitHub contributions

- ▣ 1.6 Both repositories are stored remotely on GitHub ([GitHub2016](#)). Image 1.6 shows the contribution history from the last 17 months for both of the `pata.physics`

.`wtf` code and this thesis. A darker green indicates several commits (i.e. saves) while gray indicates no commits. Each square represents a day, each column a week (Sunday–Saturday).

## **Part VI**

# HAPPILY EVER AFTER

Matter in our assistance in our quest, his journey in quest, matter of his father undertaking. It was later before I felt the force of its Centre, I found out that he had met him, if here I enter, the gas to be formed out of ether materials is a gas. Knew as much about the matter as I did when we were walking, it was impossible to get the ether due to, in spite of ate and her horn, Ulysses, the latter before I felt granting us intense vibration with fierce, but often

# **INTERLUDE III**

# Part VII

# POST

# REFERENCES

- Agichtein, Eugene, Eric Brill and Susan Dumais (2006). ‘Improving web search ranking by incorporating user behavior information’. In: **ACM SIGIR conference on Research and development in information retrieval**. New York, New York, USA: ACM Press, p. 19.
- Baeza-Yates, Ricardo and Berthier Ribeiro-Neto (2011). **Modern Information Retrieval: The Concepts and Technology Behind Search**. Addison Wesley.
- Baidu (2012). **Baidu About**.
- Baldi, Pierre and Laurent Itti (2010). ‘Of bits and wows : A Bayesian theory of surprise with applications to attention’. In: **Neural Networks** 23, pp. 649–666.
- Bao, Shenghua et al. (2007). ‘Optimizing Web Search Using Social Annotations’. In: **Distribution**, pp. 501–510.
- Bastos Filho, Carmelo et al. (2008). ‘A novel search algorithm based on fish school behavior’. In: **IEEE International Conference on Systems, Man and Cybernetics**, pp. 2646–2651.
- Bharat, Krishna and George Mihaila (2000). ‘Hilltop: A Search Engine based on Expert Documents’. In: **Proc of the 9th International WWW**. Vol. 11.
- Bing, Microsoft (2016). **Meet our crawlers**.
- Bird, Steven, Ewan Klein and Edward Loper (2009). **Natural Language Processing with Python**. Sebasopol, CA: O'Reilly Media.
- Boden, Margaret (2003). **The Creative Mind: Myths and Mechanisms**. London: Routledge.
- Brin, Sergey and Larry Page (1998a). ‘The anatomy of a large-scale hypertextual Web search engine’. In: **Computer Networks and ISDN Systems** 30.1-7, pp. 107–117.
- (1998b). ‘The PageRank Citation Ranking: Bringing Order to the Web’. In: **World Wide Web Internet And Web Information Systems**, pp. 1–17.

- Burdick, Anne et al. (2012). **Digital Humanities**. Cambridge, Massachusetts: MIT Press (cit. on p. 7).
- Burnham, Douglas (2015). 'Immanuel Kant: Aesthetics'. In: **Internet Encyclopedia of Philosophy** (cit. on p. 3).
- Candy, Linda (2012). 'Evaluating Creativity'. In: **Creativity and Rationale: Enhancing Human Experience by Design**. Ed. by J.M. Carroll. Springer.
- Colton, Simon (2008a). 'Computational Creativity'. In: **AISB Quarterly**, pp. 6–7.
- (2008b). 'Creativity versus the perception of creativity in computational systems'. In: **In Proceedings of the AAAI Spring Symp. on Creative Intelligent Systems**.
- Colton, Simon, Alison Pease and Graeme Ritchie (2001). **The Effect of Input Knowledge on Creativity**.
- De Bra, Paul, Geert-jan Houben et al. (1994). 'Information Retrieval in Distributed Hypertexts'. In: **Techniques**.
- De Bra, Paul and Reinier Post (1994a). 'Information retrieval in the World-Wide Web: Making client-based searching feasible'. In: **Computer Networks and ISDN Systems** 27.2, pp. 183–192.
- (1994b). 'Searching for Arbitrary Information in the WWW: the Fish Search for Mosaic'. In: **Mosaic A journal For The Interdisciplinary Study Of Literature**.
- Dean, Jeffrey, Luiz Andre Barroso and Urs Hoelzle (2003). 'Web Search for a Planet: The Google Cluster Architecture'. In: **Ieee Micro**, pp. 22–28.
- Deerwester, Scott et al. (1990). 'Indexing by Latent Semantic Analysis'. In: **Journal of the American Society for Information Science** 41.6, pp. 391–407.
- Ding, Li et al. (2004). 'Swoogle: A semantic web search and metadata engine'. In: **In Proceedings of the 13th ACM Conference on Information and Knowledge Management**. ACM.
- Du, Zhi-Qiang et al. (2007). 'The Research of the Semantic Search Engine Based on the Ontology'. In: **2007 International Conference on Wireless Communications, Networking and Mobile Computing**, pp. 5398–5401.
- Elton, Matthew (1995). 'Artificial Creativity: Enculturing Computers'. In: **Leonardo** 28.3, pp. 207–213 (cit. on p. 7).
- Foucault, Michel (1966). 'The Order of Things - Preface'. In: **The Order of Things**. France: Editions Gallimard. Chap. Preface, pp. xv–xxiv (cit. on p. 6).
- Garcia-Molina, Hector, Jan Pedersen and Zoltan Gyongyi (2004). 'Combating Web Spam with TrustRank'. In: **In VLDB**. Morgan Kaufmann, pp. 576–587.
- Glover, E.J. et al. (2001). 'Improving category specific Web search by learning query modifications'. In: **Proceedings 2001 Symposium on Applications and the Internet**, pp. 23–32.
- Google (2012). **Google Ranking**.
- (2016). **Googlebot**.

- Haveliwala, Taher H (2003). 'Topic-Sensitive PageRank: A Context Sensitive Ranking Algorithm for Web Search'. In: ***Knowledge Creation Diffusion Utilization*** 15.4, pp. 784–796.
- Hendler, Jim and Andrew Hugill (2011). 'The Syzygy Surfer : Creative Technology for the World Wide Web'. In: ***ACM WebSci 11*** (cit. on p. 3).
- Hersovici, M et al. (1998). 'The shark-search algorithm. An application: tailored Web site mapping'. In: ***Computer Networks and ISDN Systems*** 30.1-7, pp. 317–326.
- Hotho, Andreas et al. (2006). 'Information retrieval in folksonomies: Search and ranking'. In: ***The Semantic Web: Research and Applications, volume 4011 of LNAI***. Springer, pp. 411–426.
- Hugill, Andrew (2012). **'Pataphysics: A Useless Guide**. Cambridge, Massachusetts: MIT Press (cit. on p. 6).
- Jarry, Alfred (1996). ***Exploits and Opinions of Dr Faustroll, Pataphysician***. Cambridge, MA: Exact Change.
- (2006). ***Collected Works II - Three Early Novels***. Ed. by Alastair Brotchie and Paul Edwards. London: Atlas Press (cit. on pp. 3, 6).
- Jeh, Glen and Jennifer Widom (2002). 'SimRank: A Measure of Structural Context Similarity'. In: ***In KDD***, pp. 538–543.
- Jordanous, Anna Katerina (2011). 'Evaluating Evaluation : Assessing Progress in Computational Creativity Research'. In: ***Proceedings of the Second International Conference on Computational Creativity***.
- (2012). 'Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application'. PhD thesis. University of Sussex.
- Jordanous, Anna Katerina and Bill Keller (2012). 'Weaving creativity into the Semantic Web: a language-processing approach'. In: ***Proceedings of the 3rd International Conference on Computational Creativity***, pp. 216–220.
- Jurafsky, Daniel and James H Martin (2009). ***Speech and Language Processing***. London: Pearson Education.
- Kamps, Jaap, Rianne Kaptein and Marijn Koolen (2010). ***Using Anchor Text , Spam Filtering and Wikipedia for Web Search and Entity Ranking***. Tech. rep. ?
- Kleinberg, Jon M (1999). 'Authoritative sources in a hyperlinked environment'. In: ***journal of the ACM*** 46.5, pp. 604–632.
- Kleinberg, Jon M et al. (1999). 'The Web as a graph : measurements, models and methods'. In: ***Computer***.
- Luke, Saint (2005). ***The Gospel According to St. Luke***. Ebible.org.
- Luo, Fang-fang, Guo-long Chen and Wen-zhong Guo (2005). 'An Improved 'Fish-search' Algorithm for Information Retrieval'. In: ***2005 International Con-***

- ference on Natural Language Processing and Knowledge Engineering*, pp. 523–528.
- Macdonald, Craig (2009). ‘The Voting Model for People Search’. In: **Philosophy**.
- Manning, Christopher, Prabhakar Raghavan and Hinrich Schuetze (2009). **Introduction to Information Retrieval**. Cambridge UP.
- Marchionini, Gary (2006). ‘From finding to understanding’. In: **Communications of the ACM** 49.4, pp. 41–46.
- Marchionini, Gary and Ben Shneiderman (1988). ‘Finding facts vs. browsing knowledge in hypertext systems’. In: **Computer** 21.1, pp. 70–80.
- Marcus, Mitchell P, Beatrice Santorini and Mary Ann Marcinkiewicz (1993). ‘Building a Large Annotated Corpus of English: The Penn Treebank’. In: **Computational Linguistics** 19.2.
- Mayer, Richard E (1999). ‘Fifty Years of Creativity Research’. In: **Handbook of Creativity**. Ed. by Robert J Sternberg. New York: Cambridge University Press. Chap. 22, pp. 449–460 (cit. on p. 37).
- Mayhaymate (2012). **File:PageRank-hi-res.png**. URL: <https://commons.wikimedia.org/wiki/File:PageRank-hi-res.png> (visited on 18/10/2016).
- Michelsen, Maria Hagsten and Ole Bjorn Michelsen (2016). **Regex Crossword**. URL: <http://regexcrossword.com/> (visited on 19/10/2016).
- Microsoft (2012). **Bing Fact Sheet**.
- Miller, George A. (1995). ‘WordNet: a lexical database for English’. In: **Communications of the ACM** 38.11, pp. 39–41.
- Miyamoto, Sadaaki (1988). **Information Retrieval based on Fuzzy Associations**.
- (2010). **Fuzzy Sets in Information Retrieval and Cluster Analysis (Theory and Decision Library D)**. Springer, p. 276.
- Miyamoto, Sadaaki and K Nakayama (1986). ‘Fuzzy Information Retrieval Based on a Fuzzy Pseudothesaurus’. In: **IEEE Transactions on Systems, Man and Cybernetics** 16.2, pp. 278–282.
- Nick, Z.Z. and P. Themis (2001). ‘Web Search Using a Genetic Algorithm’. In: **IEEE Internet Computing** 5.2, pp. 18–26.
- Nicole (2010). **The 10 Most Incredible Google Bombs**.
- Pease, Alison and Simon Colton (2011). ‘On impact and evaluation in Computational Creativity : A discussion of the Turing Test and an alternative proposal’. In: **Proceedings of the AISB**.
- Pease, Alison, Simon Colton et al. (2013). ‘A Discussion on Serendipity in Creative Systems’. In: **Proceedings of the 4th International Conference on Computational Creativity**. Vol. 1000. Sydney, Australia: University of Sydney, pp. 64–71.

- Pease, Alison, Daniel Winterstein and Simon Colton (2001). 'Evaluating Machine Creativity'. In: ***Proceedings of ICCBR Workshop on Approaches to Creativity***, pp. 129–137.
- Piffer, Davide (2012). 'Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research'. In: ***Thinking Skills and Creativity*** 7.3, pp. 258–264.
- Polya, George (1957). ***How To Solve It***. 2nd. Princeton, New Jersey: Princeton University Press.
- Project, NLTK (2016). ***Natural Language Toolkit***. URL: <http://www.nltk.org/> (visited on 18/10/2016).
- Ritchie, Graeme (2001). 'Assessing creativity'. In: ***AISB '01 Symposium on Artificial Intelligence and Creativity in Arts and Science***. Proceedings of the AISB'01 Symposium on Artificial Intelligence, Creativity in Arts and Science, pp. 3–11.
- (2007). 'Some Empirical Criteria for Attributing Creativity to a Computer Program'. In: ***Minds and Machines*** 17.1, pp. 67–99.
  - (2012). 'A closer look at creativity as search'. In: ***International Conference on Computational Creativity***, pp. 41–48.
- Schmidhuber, Juergen (2006). ***New millennium AI and the Convergence of history***.
- Schuetze, Hinrich (1998). 'Automatic Word Sense Discrimination'. In: ***Computational Linguistics***.
- Schuetze, Hinrich and Jan Pedersen (1995). ***Information Retrieval Based on Word Senses***.
- Shu, Bo and Subhash Kak (1999). 'A neural network-based intelligent meta-search engine'. In: ***Information Sciences*** 120.
- Srinivasan, P (2001). 'Vocabulary mining for information retrieval: rough sets and fuzzy sets'. In: ***Information Processing and Management*** 37.1, pp. 15–38.
- Sutcliffe, Alistair and Mark Ennis (1998). 'Towards a cognitive theory of information retrieval'. In: ***Interacting with Computers*** 10, pp. 321–351.
- Taye, Mohammad Mustafa (2009). 'Ontology Alignment Mechanisms for Improving Web-based Searching'. PhD thesis. De Montfort University.
- Turing, Alan (2009). 'Computing Machinery and Intelligence'. In: ***Parsing the Turing Test***. Ed. by Robert Epstein, Gary Roberts and Grace Beber. Springer. Chap. 3, pp. 23–66 (cit. on pp. 6, 7, 32).
- University, Princeton (2010). ***What is WordNet?*** URL: <http://wordnet.princeton.edu> (visited on 20/10/2016).
- Varshney, Lav R et al. (2013). 'Cognition as a Part of Computational Creativity'. In: ***12th International IEEE Conference on Cognitive Informatics and Cognitive Computing***. New York City, USA, pp. 36–43.

- Ventura, Dan (2008). 'A Reductio Ad Absurdum Experiment in Sufficiency for Evaluating (Computational) Creative Systems'. In: **5th International Joint Workshop on Computational Creativity**. Madrid, Spain.
- Verne, Jules (2010). **A Journey to the Interior of the Earth**. Project Gutenberg.
- Vries, Erica de (1993). 'Browsing vs Searching'. In: **OCTO report 93/02**.
- Wickson, F., A.L. Carew and A.W. Russell (2006). 'Transdisciplinary research: characteristics, quandaries and quality'. In: **Futures** 38.9, pp. 1046–1059 (cit. on p. 6).
- Widyantoro, D.H. and J. Yen (2001). 'A fuzzy ontology-based abstract search engine and its user studies'. In: **10th IEEE International Conference on Fuzzy Systems** 2, pp. 1291–1294.
- Wiggins, Geraint A (2006). 'A preliminary framework for description, analysis and comparison of creative systems'. In: **Knowledge Based Systems** 19.7, pp. 449–458.

**KTHXBYE**

[Go to TOC](#)