

List of Todos

<input type="checkbox"/> double check hyperref links in pdf toc	1
<input type="checkbox"/> finish introduction	4
<input type="checkbox"/> complete assumptions	4
<input type="checkbox"/> answer research questions in conclusion	9
<input type="checkbox"/> ref	28
<input type="checkbox"/> Finish section here. references and all	34
<input type="checkbox"/> finish intro to creativity and computers chapter	36
<input type="checkbox"/> put summaries at back of chapter or front? styling?	38
<input type="checkbox"/> general introduction about human creativity models	40
<input type="checkbox"/> Mel Rhodes vs Ross Mooney??	41
<input type="checkbox"/> DOB of authors?	43
<input type="checkbox"/> redo diagram	43
<input type="checkbox"/> ref	44
<input type="checkbox"/> ref	44
<input type="checkbox"/> ref	47
<input type="checkbox"/> references	53
<input type="checkbox"/> rewrite and format	56
<input type="checkbox"/> cross ref	58
<input type="checkbox"/> update all graphics with inkscape	69

■ references for this whole chapter?	69
■ example TDM for faustroll sentence?	71
■ cross references with hyperlink hypertarget	71
■ change font again because of ligatures?	73
■ decide on which method for highlighting words — italic or apostrophe . .	73
■ describe NLTK and the core functionality	80
■ reformat	94
■ ref	100
■ comparison table Poincare, Wallas, Polya, CC?	113
■ redraw figure	121
■ style inline code	126
■ run code on laptop and get snippets of all variable contents, e.g. faustroll, froll_dict,	126
■ give examples of different results if using different base documents! . .	126
■ add section about which pieces of code are not written by me	126
■ fix page numbers	126
■ folder structure	128
■ add audio? update this section depending on what i do	128
■ Explain difference in Text, Image and Video	133
■ find ref for dameraulevenshtein in baeza-yates book?	133
■ explain reasoning behind algorithms like this for all:	137
■ rewrite getnym function to automatically get all three without the ifs . .	137
■ get proper ref for sonnet style	142
■ get new screenshots for prototype 1	145
■ don't mention James?	145

■ discuss problems with algorithms, pros and cons...	147
■ ref	152
■ put cross refs for each piece of code here	167
■ finish list here	174

Institute of Creative Technologies
De Montfort University

FANIA RACZINSKI

WEB COLLIDER

**A pataphysical methodology for
applying creativity to exploratory
search**

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: 1st November 2015

PRE☺

Abstract

Web Collider — Fania Raczinski

A pataphysical methodology for applying creativity to exploratory search

Creativity, Pataphysics and Computers

Obscure French Philosophy

This paper looks at defining, analysing and practicing how creativity can be applied to search tools. It defines creativity with respect to search and discusses how these concepts could be applied in software engineering using principles from the pseudo-philosophy of pataphysics. The aim of the proposed tool is to generate surprising, novel, humorous and provocative search results instead of purely relevant ones, in order to inspire a more creative interaction between a user, their information need and the application. A proof-of-concept prototype is described to justify the ideas presented before issues and future work are discussed.

Contents

Todo list	1
PREFACE	i
Abstract	ii
Contents	iii
List of Figures	vii
List of Tables	ix
List of Code Snippets	x
List of Acronyms	xi
I HELLØ WORLD	2
1 Introduction	3
1.1 Problem / Motivation / Context	4
1.1.1 Related Work	6
1.1.2 Research Questions	8
1.2 Methodology	9
1.3 Deliverables / Outcomes	9
1.3.1 Contribution to Knowledge	10
1.3.2 Publications	10
1.4 The Hitchhiker's Guide to this Thesis	11
2 Methodology	12
2.1 Intradisciplinary Research	14
2.1.1 Computer Science Research	15
2.1.2 Humanities Research	17
2.1.3 Arts Research	17
2.2 Transdisciplinary Research	17
2.2.1 Practice Based Research	18
2.3 Research Approach	18

II ΤΗΘΟΛΣ ΘΕ ΤΗΣ ΤΡΔΣ	19
3 Pataphysics	20
3.1 Conscious	23
3.2 Self-conscious	28
3.2.1 Symbology	31
3.2.2 Antimony	31
3.2.3 Anomaly	32
3.2.4 Sysygy	32
3.2.5 Clinamen	33
3.2.6 Absolute	33
3.3 Unconscious	34
3.3.1 Oulipo	34
3.3.2 Borges	35
4 Creativity	36
4.1 In Humans	40
4.1.1 Mel Rhodes and Ross Mooney	40
4.1.2 Arthur Koestler	41
4.1.3 Henri Poincaré, Graham Wallas and George Pólya	42
4.1.4 James Kaufman and Ron Beghetto	43
4.1.5 Margaret Boden	45
4.1.6 Robert J. Sternberg, James C. Kaufman and Timothy Leary	47
4.2 In Computers	48
4.2.1 Bipin Indurkhya	49
4.2.2 Partridge and Rowe	50
4.2.3 David Gelernter	51
4.2.4 Marvin Minsky	52
4.2.5 Matthew Elton	52
4.3 In Academia	53
4.3.1 Computational Creativity	55
4.3.2 Creative Computing	56
4.3.3 Speculative Computing	59
4.3.4 Digital Humanities	61
4.3.5 Computer Ethics	68
5 Technology	69
5.1 Information Retrieval	70
5.1.1 IR Models	72
5.1.2 Ranking	75
5.1.3 Query Expansion and Relevance Feedback	78
5.2 Natural Language Processing	80
5.2.1 Damerau-Levenshtein	80
5.3 Linguistics / WordNet	91
5.4 Algorithm Formalisation	92
6 Evaluation	94
6.1 IR Evaluation	98
6.2 Approaches	99
6.3 Output Minus Input	101

6.4	Measurable Attributes	101
6.5	Linda Candy	103
III THE CORE: TΣCHNΘ-LΘGIC		110
7 Theoretical Foundations		111
7.1	Searching vs. Browsing	112
7.2	Creativity	113
7.3	Pataphysics	114
7.3.1	and Creativity	114
7.3.2	and Computers	118
7.4	Patalgorithms	121
7.4.1	Pataphysicalisation	121
7.4.2	Patadata	123
7.4.3	Ranking — Pranking?	124
7.5	User experience	125
8 Practical Implementation		126
8.1	Corpus	128
8.2	Setup	130
8.3	Text	133
8.3.1	Clinamen	133
8.3.2	Syzygy	136
8.3.3	Antinomy	137
8.4	Image & Video	138
8.4.1	REST & API	139
8.5	Design	141
8.6	Prototypes	144
8.6.1	Prototype 2	144
8.6.2	Prototype 1	145
9 Impact and Applications — Case Study		147
9.1	Patakosmos	149
9.2	Soeren and the other guy	149
9.3	Digital Opera	149
9.3.1	Use	149
9.3.2	Result	149
9.3.3	Interview	150
IV MΣTΛ-LΘGICΛLYSIS		151
10 Interpretation		152
10.1	Measurable Attributes	154
10.2	Problems with Evaluation	156
10.3	5 P's: product, process, people, place and purpose	158
11 Patacritical Analysis		160
11.1	Problems Encountered	160
11.2	Design Aspects	160

11.3 Search Results	160
V H\PPILY \EV\ER \VFT\ER?	161
12 Aspirations	162
12.1 Code	162
12.2 Interface	163
12.3 Algorithms	163
12.4 Architecture	163
12.5 Research	163
13 Observations	164
13.1 Achievements	165
13.2 Implications	165
13.3 Recommendations	165
POSTFACE	166
A Code	167
A.1 Setup	167
A.2 Damerau Levenshtein	168
A.3 WordNet	169
A.3.1 Antinomy	169
A.3.2 Syzygy	170
A.4 Random Quotes	173
B Publications	174
Bibliography	209

List of Figures

2.1	Epistemology	13
3.1	Men in Black	23
3.2	figures1–3	24
3.3	Ubu	25
3.4	Faustroll	28
3.5	Crocodile	31
3.6	Gidouille	31
3.7	Green Candle	31
4.1	4 Aspects of Creativity	41
4.2	The 4 C Model	43
4.3	The 4 C Model2	44
4.4	Digital Humanities	63
5.1	Search Engine Architecture	70
5.2	Various wordcounts	72
5.3	Vector Model	74
5.4	PageRank algorithm	76
5.5	Grammers	89
6.1	Precision and Recall	99
6.2	Multi-dimensional Model of Creativity and Evaluation	106
6.3	Text for Table of Contents	107
7.1	Kaufman vs Boden	114
7.2	Pata central	122
7.3	Pata central	122
7.4	Pataphysicalisation	123
7.5	Feedback button	125
8.1	Toulouse-Lautrec’s “Jane Avril”	131
8.2	Bonnard’s “Revue Blanche”	131
8.3	Aubrey Beardsley’s “Docteur Faustroll”	131

8.4	Oberthuer's "Saint Cado"	131
8.5	Poetry results screenshot	141
8.6	Poetry results mobile screenshot	142
8.7	Fibonacci Spiral	143
8.8	Prototype 2 screenshot	144
8.9	Prototype 1 screenshot	146

List of Tables

4.1	Creative Triptych	42
4.2	Leary's four types of creativity	48
4.3	Leary's Social Labels	49
5.1	MaxEnt Example table	86
7.1	Creative Computing vs Digital Humanities vs Computational Creativity	115
7.2	4 Step Model vs 4 P Model vs Problem Solving	115
7.3	Creativity vs Pataphysics	119
10.1	Creativity attributes	155
13.1	Comparison of algorithms	165

List of Source Code

5.1 Pseudo-code for computing vector scores	75
8.1 Adding text files to the corpus library.	130
8.2 ‘setupcorpus’ function to process the corpus and create the index.	132
8.3 Clinamen function	134
8.4 ‘get_results’ function to get all sentences for a list of words.	134
8.5 ‘pp_sent’ function to retrieve a sentence from a file.	135
8.6 Syzygy function.	136
8.7 Antinomy function.	137
8.8 Function to pataphysicalise image and video query terms.	138
8.9 Translation function.	139
8.10 Using the Microsoft Bing API to retrieve images.	140
8.11 Code for rendering Queneau style poems.	143
8.12 Code for randomising poems.	145

Acronyms

AI Artificial Intelligence. [52, 55](#)

API Application Program Interface. [127, 138, 139](#)

CC Creative Computing. [45, 55, 56, 57, 58, 59](#)

DH Digital Humanities. [56, 58, 61](#)

DMU

De Montfort University. [174](#)

HTTP

Hypertext Transfer Protocol. [139](#)

ICCC

International Conference on Computational Creativity. [55, 56](#)

IDF Inverse Document Frequency. [71, 72](#)

IJCrC

International Journal of Creative Computing. [56](#)

IN Information Need. [111, 112](#)

IOCT

Institute of Creative Technologies. [174](#)

IR Information Retrieval. [70, 72, 73, 75, 79](#)

LMS

Leicester Media School. [174](#)

NLP

Natural Language Processing. [81, 130](#)

NLTK

Natural Language Tool Kit. [80, 130, 136](#)

REST

Representational State Transfer. [139](#)

TDM

Term-Document Matrix. [71](#)

TF Term Frequency. [71](#), [72](#)

double check hyperref links in pdf toc

Part I

HΣLLΘ WΘRLD

Chapter 1

Introduction

1.1	Problem / Motivation / Context	4
1.1.1	Related Work	6
1.1.2	Research Questions	8
1.2	Methodology	9
1.3	Deliverables / Outcomes	9
1.3.1	Contribution to Knowledge	10
1.3.2	Publications	10
1.4	The Hitchhiker's Guide to this Thesis.	11

Part of this chapter has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

§¶

HELLO WORLD

This thesis describes a general study into computer creativity and in particular a pataphysical methodology applied to creative information retrieval.

A pataphysical methodology for applying creativity to exploratory search.

A semantic search tool based on a pataphysical algorithm using a 'patadata' ontology

1. Semantic Web technologies
2. IR techniques
3. Pataphysics

4. Ranking algorithms
5. Data — metadatada — patadata
6. Ontologies, RDF, OWL, etc

finish introduction

In the following chapters, I will examine how pataphysics and creativity map onto one another, give an outline of the field of information retrieval...

1.1 Problem / Motivation / Context

Assumptions: computers can be creative.

complete assumptions

From relevant to creative.

Why pataphysics? Pataphysics is highly subjective and particular and is as such very suitable for this kind of transformation from relevant to creative. Pataphysics can provide some useful techniques that are very suitable for creative computing.

Purposive without purpose (Kant)

The aim of this project, in simple words, is to design a tool for creative searching on the Web. It tries to create a fresh way of searching and navigating through information and content on the Internet, to bring back some of the inspirational chance encounters that were so characteristic for libraries.

Current information retrieval systems might be used for creative purposes, however, they do not directly provide creative results to their users, instead they focus on precise and relevant results only. Therefore we argue that a new style of system is required. It is clear that the fundamental problem in this is that standard algorithms are not suited for these problems, with them considering a document to be groupings of words in traditional IR systems, and that an entire document falls under the same classifications in semantic IR systems.

The tool will challenge the way we understand data and metadata and suggest new meanings between them in accordance with the underlying philosophy. The way we think about information is redefined. It will create new standards or rather get rid of all previous standards and classifications. It will create new

links and connections between pieces of information that are not simply based on keyword similarities but rather a more poetic sense of unity. It will be a refreshing new view compared to the structured and standardised thinking of computer scientists. It will challenge the binary logic that dominates the world of programmers and concentrate on what they would view as an illogical and unstructured system going against all standards.

Jorge Luis Borges 'Chinese Encyclopaedia' ([Borges 2000](#)) illustrates this idea very well. The encyclopaedia lists the following results under the category of 'animal'.

1. those that belong to the Emperor,
2. embalmed ones,
3. those that are trained,
4. suckling pigs,
5. mermaids,
6. fabulous ones,
7. stray dogs,
8. those included in the present classification,
9. those that tremble as if they were mad,
10. innumerable ones,
11. those drawn with a very fine camelhair brush,
12. others,
13. those that have just broken a flower vase,
14. those that from a long way off look like flies.

Although these are all perfectly valid results, it is clear that they form a more creative, even poetic, view of what an animal might be than the Oxford English Dictionary's prosaic: "a living organism which feeds on organic matter".

We have introduced the motivation and concept for a creative Web search tool and discussed some of the major challenges a project like this faces. With Web search being a major research and learning tool nowadays, it is imperative to think about how such a tool could be (ab) used. Ethical issues that arise through the provision of unexpected results, and the misunderstandings this could lead to, will be discussed in future work. Nevertheless, we believe that creative Web search can facilitate inspirational learning through an exploratory search journey and we hope our tool will provide just that.

1.1.1 Related Work

The Syzygy Surfer

The research presented here is built on the initial ideas of Jim Hendler and Andrew Hugill's "Syzygy Surfer" (Hendler and Hugill 2011; Hendler and Hugill 2013). They first suggested the use of three pataphysical principles, namely clinamen, syzygy and anomaly, to create a new type of Web search engine, reminiscent of the experience of "surfing the Web". This is in contrast to current search engines which value relevant results over creative ones. "Surfing" used to be a creative interaction between a user and the web of information on the Internet, but the regular use of modern search engines has changed our expectations of this sort of knowledge acquisition. It has drifted away from a learning process by exploring the Web to a straightforward process of information retrieval similar to looking up a word in a dictionary.

The ambiguity of experience is the hallmark of creativity, that is captured in the essence of pataphysics. Traversing the representations of this ambiguity using algorithms inspired by the syzygy, clinamen and anomaly of pataphysics, using a panalogical mechanism applied to metadata, should be able to humanize and even poeticize the experience of searching the Web.(Hendler and Hugill 2013)

Their inspirations come from Borges' "Chinese Encyclopaedia" (Borges 2000) (for the underlying poetic sense of unity), Jarry's Pataphysics (Jarry 1996) (for the concept of patadata – data beyond metadata) and panalogies (parallel analogies – to introduce ambiguity, since it allows various descriptions of the same object) as formulated by Singh (Singh 2005).

Yossarianlives

Traditional search like Google, Bing, Yahoo, or DuckDuckGo returns the most popular results, and gives you expected & cliché ideas. It's what everyone else already thinks about a topic.

Yossarian creative search returns diverse and unexpected results that share loose associations to your search. The results help you think about your topic in new ways and generate new ideas.

If you search "beauty" in Google you get pages and pages of white models. It returns a singular way of thinking about the topic.

Search "beauty" in Yossarian it returns disparate results with shared attributes, showing many different ways of thinking. Is beauty a

diamond (strong, rare, flawless), or is family beautiful (activity, togetherness) or is beauty architected (designed, planned, precision)?¹

Use traditional search for when you know what you are looking for.

Traditional search is for learning what the world already knows.

Use creative search for when you don't know what you are looking for.

Creative search is for helping you come up with new ideas.²

Augmented creativity

Seeing lateral connections is an uniquely human ability, and your individual ability to make these connections is based on your own experience. We build tools that help you see more by suggesting connections that you probably wouldn't have considered otherwise, increasing the diversity and frequency of your ideas, and priming you for creativity.

Lateral Discovery

When working to escape filter bubbles too often randomness and complete serendipity are seen as the answer for new discovery. We build tools that explore lateral association between content and believe this approach can lead to predictable moments of discovery that drive engagement, traffic, and purchase, etc.

Metaphorical Search

Search engines today have a problem in that they tell us what the world already knows, reinforcing existing knowledge. We build tools that seek to remedy this by returning results that are disparate, but metaphorically related to the query. These types of results are incredibly useful for any one who derives value from new ideas.

Creative Graph

Facebook created the Social Graph helping us understand how people and things are connected, Google created the Knowledge Graph to codify the world's information, and here at Yossarian Lives we are developing the Creative Graph, a radical new way to understand the conceptual relationships between things.³

¹<https://yossarianlives.com/>

²<https://yossarianlives.com/>

³<http://about.yossarianlives.com/index.html>

The Library of Babel

The Library of Babel is a place for scholars to do research, for artists and writers to seek inspiration, for anyone with curiosity or a sense of humor to reflect on the weirdness of existence — in short, it's just like any other library. If completed, it would contain every possible combination of 1,312,000 characters, including lower case letters, space, comma, and period. Thus, it would contain every book that ever has been written, and every book that ever could be — including every play, every song, every scientific paper, every legal decision, every constitution, every piece of scripture, and so on. At present it contains all possible pages of 3200 characters, about 104677 books.

Since I imagine the question will present itself in some visitors' minds (a certain amount of distrust of the virtual is inevitable) I'll head off any doubts: any text you find in any location of the library will be in the same place in perpetuity. We do not simply generate and store books as they are requested — in fact, the storage demands would make that impossible. Every possible permutation of letters is accessible at this very moment in one of the library's books, only awaiting its discovery. We encourage those who find strange concatenations among the variations of letters to write about their discoveries in the forum, so future generations may benefit from their research.⁴

1.1.2 Research Questions

- How can we make a search tool that is inspirational rather than informational?
- How can we get search results that are unexpected and yet make sense?
- How can we rank search results but still be true to Pataphysics philosophy?
- How can we represent and structure data to reflect its context, meaning and subjectivity?
- How can we present search results in a creative and pataphysical way?
- How does Pataphysics relate to creative computing?
- How can we use Pataphysics as inspiration for search ranking?
- How can we write a specifically creative algorithm?
- How can Semantic Web technologies help with the representation of patadata?
- What does it mean for search results to be creative/relevant?
- Can computers be creative?
- What does it actually mean to be creative even for a human being, etc

⁴<https://libraryofbabel.info/>

- Is pataphysics creative?
- What is a relevant search result?
- Is creativity irrelevant?

answer research questions in conclusion

1.2 Methodology

This project combines research in science and art. It is an interdisciplinary research project.

This project has roots in disciplines such as Computer Science and Humanities.

Information Retrieval

: Software Engineering, Semantic Web

Pataphysics

: Literature, Philosophy, Ontology

Creativity

: Cognitive Science, Artificial Intelligence

In regards to my project:

- A concept implementation method is used with a descriptive-other approach
- A qualitative investigation into if and why the proposed search results are useful will be done
- Following experimental methodologies, to evaluate the proposed new solution to the problem of creative search

Epistemology

: Subjective/Argumentative

Methodology

: Experimental, Interpretative, Qualitative

Methods

: Concept implementation, (Heuristic) Evaluation

1.3 Deliverables / Outcomes

- Design a tool for creative searching on the Web

- Design pataphysics inspired algorithms to model creativity in this tool
- Produce a proof-of-concept prototype
- Propose a framework for evaluating and interpreting creative search results

<http://pata.fania.eu>

!!!

1.3.1 Contribution to Knowledge

Compared to the mainstream search engines available publicly, my project's approach to searching will be very different. Google, for example, ranks web pages based on the number and quality of incoming hyperlinks (Google 2012), while I will be using the concept of patadata (data-metadata-patadata) and semantic web technologies.

There have been interdisciplinary projects like Johanna Drucker's Speclab (Drucker 2009) which have similar inspirations, and literature on Pataphysics is found in several places (Boek 2002; Hugill 2012b) although the idea of using it as an inspiration for a search engine appears to be very new (Hendler and Hugill 2013).

A lot of literature exists on how to model creativity on a computer [5] [7], explaining several theories on how the mind works and how to simulate this on a machine. However none of these have any kind of pataphysical inspiration and aren't applicable to search engines, so we have presented a paper on creativity in search results just recently (Raczinski, Yang and Hugill 2013).

1.3.2 Publications

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. (Sawle, Raczinski and Yang 2011)

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. (Hugill, Yang et al. 2013)

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. (Raczinski, Yang and Hugill 2013)

1.4 The Hitchhiker's Guide to this Thesis

PREFACE

Part I

IN THE BEGINNING...

Chapter 1

Introduction

Chapter 2

Methodology

Part II

IN A GALAXY FAR FAR AWAY...

Chapter 3

Pataphysics

Chapter 4

Creativity and Computers

Chapter 5

IR and NLP

Part III

THE CORE: TECHNO-LOGIC

Chapter 6

Theoretical Foundations

Chapter 7

Practical Implementation

Chapter 8

Impact and Applications — Case Study

Part IV

INTECHNOIL-LOGICALYSIS

Chapter 9

Interpretation / Evaluation

Chapter 10

Patacritical Analysis

Part V

HAPPY END

Chapter 11

Aspirations

Chapter 12

Observations

POSTFACE

Chapter 2

Methodology

2.1	Intradisciplinary Research	14
2.1.1	Computer Science Research	15
2.1.2	Humanities Research	17
2.1.3	Arts Research	17
2.2	Transdisciplinary Research	17
2.2.1	Practice Based Research	18
2.3	Research Approach	18

Description and justification of methodology...

“Only those who attempt the absurd achieve the impossible.” (attributed to M.C. Escher)

Ernest Edmonds, Linda Candy, 2010. “Relating Theory, Practice and Evaluation in Practitioner Research” LEONARDO, Vol 43, No 5, pp470–476.

Epistemology

: “A broad and high-level outline of the reasoning process by which a school of thought performs its empirical and logical work.” Wikipedia

Methodology

: “Less high level than epistemology is methodology. It refers to a more specific manner in which to do empirical and logical work. The same epistemology can have several methodologies.” Wikipedia

Method

: A methodology can consist of several methods. Wikipedia

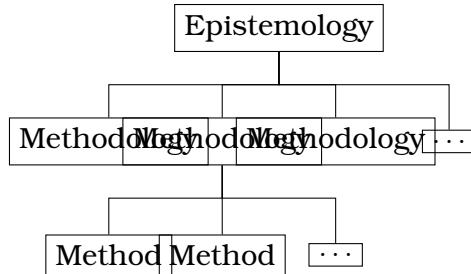


Figure 2.1: Epistemology breakdown chart

Epistemology

“is the branch of philosophy concerned with the nature and scope (limitations) of knowledge. It addresses mainly the following questions: What is knowledge? How is knowledge acquired? To what extent is it possible for a given subject or entity to be known?” [2]

“A broad and high-level outline of the reasoning process by which a school of thought performs its empirical and logical work.” [1]

Methodology

“is usually a guideline system for solving a problem, with specific components such as phases, tasks, methods, techniques and tools. It can be defined also as follows: 1. ‘the analysis of the principles of methods, rules, and postulates employed by a discipline’ 2. ‘the systematic study of methods that are, can be, or have been applied within a discipline’ 3. ‘the study or description of methods’” [2]

“Less high level than epistemology is methodology. It refers to a more specific manner in which to do empirical and logical work. The same epistemology can have several methodologies.” [1]

Research Strategy is a procedure for achieving a particular intermediary research objective—such as sampling, data collection, or data analysis. We may therefore speak of sampling strategies or data analysis strategies. [2]

Research Approach refers to an integrated set of research principles and general procedural guidelines. Examples of research approaches include experiments, surveys, correlational studies, ethnographic research, and phenomeno-

logical inquiry. [2]

Qualitative researchers aim to gather an in-depth understanding of human behavior and the reasons that govern such behavior. The qualitative method investigates the why and how of decision making, not just what, where, when. Hence, smaller but focused samples are more often needed than large samples. [2]

Quantitative research refers to the systematic empirical investigation of social phenomena via statistical, mathematical or computational techniques. The objective of quantitative research is to develop and employ mathematical models, theories and/or hypotheses pertaining to phenomena. Quantitative data is any data that is in numerical form such as statistics, percentages, etc. [2]

2.1 Intradisciplinary Research

Different disciplines prefer different research methodologies. It makes sense that research in medicine, chemistry, literature or mathematics all use different methods. What could a mathematician achieve in a white laboratory coat and test tubes in his hand, and similarly, what could a chemist achieve with pen, paper and a calculator?

“methodological pluralism is acceptable but what is not acceptable is philosophical pluralism”??

What would be traditional RM in those fields?

Why can I not mix and match them?

What do I do now/instead?

Can inter/multi/trans-disciplinary research be NOT collaborative but done by a single person?

“When you describe your methods it is necessary to state how you have addressed the research questions and/or hypotheses. The methods should be described in enough detail for the study to be replicated, or at least repeated in a similar way in another situation. Every stage should be explained and justified with clear reasons for the choice of your particular methods and materials.”¹

¹<http://bit.ly/1Edj84y>

2.1.1 Computer Science Research

In their rather old but still insightful analysis of over 600 papers, published between 1995 and 1999, Ramesh et al ([Ramesh, Glass and Vessey 2004](#)) have shown that -by far- the most common approach to research in computer science was “formulative” (as opposed to “descriptive” and “evaluative”) in particular in regards to “processes, methods and algorithms”.

Research Approach in CS [4]

Descriptive: (9.88%)

- Descriptive-system (4.14%)
- Descriptive-other (5.10%)
- Review of literature (0.64%)

Evaluative: (10.98%)

- Evaluative-deductive (1.11%)
- Evaluative-interpretive (-)
- Evaluative-critical (-)
- Evaluative-other (9.87%)

Formulative: (79.15%)

- Formulative-framework (2.39%)
- Formulative-guidelines/standards (0.64%)
- Formulative-model (5.73%)
- Formulative-process, method, algorithm (52.55%)
- Formulative-classification/taxonomy (0.80%)
- Formulative-concept (17.04%)

Research Method in CS [4]

- Action research (-)
- Conceptual analysis (15.13%)
- Conceptual analysis/mathematical (73.41%)
- Concept implementation (2.87%)
- Case study (0.16%)
- Data analysis (0.16%)
- Discourse analysis (-)
- Ethnography (-)
- Field experiment (-)
- Field study (0.16%)
- Grounded theory (-)
- Hermeneutics (-)
- Instrument development (-)

- Laboratory experiment (human subjects) (1.75%)
- Literature review / analysis (0.32%)
- Laboratory experiment (software) (1.91%)
- Meta-analysis (-)
- Mathematical proof (2.39%)
- Protocol analysis (-)
- Phenomenology (-)
- Simulation (1.75%)
- Descriptive/exploratory survey (-)

Formal

Formal methodologies are mostly used to prove facts about algorithms and system. Formal specification of a software component in order to allow the automatic verification of an implementation of that component, the time or space complexity of an algorithm, or on the correctness or the quality of the solutions generated by the algorithm. [5]

Experimental

Experimental methodologies are broadly used in CS to evaluate new solutions for problems. Experimental evaluation is often divided into two phases. In an exploratory phase the researcher is taking measurements that will help identify what are the questions that should be asked about the system under evaluation. Then an evaluation phase will attempt to answer these questions. A well-designed experiment will start with a list of the questions that the experiment is expected to answer. [5]

Build

A “build” research methodology consists of building an artifact — either a physical artifact or a software system — to demonstrate that it is possible. To be considered research, the construction of the artifact must be new or it must include new features that have not been demonstrated before in other artifacts. [5]

Process

A process methodology is used to understand the processes used to accomplish tasks in Computing Science. This methodology is mostly used in the areas of Software Engineering and Man-Machine Interface which deal with the way humans build and use computer systems. The study of processes may also be used to understand cognition in the field of Artificial Intelligence. [5]

Model

The model methodology is centered on defining an abstract model for a real system. This model will be much less complex than the system that it models, and therefore will allow the researcher to better understand the system and to use the model to perform experiments that could not be performed in the system itself because of cost or accessibility. The model methodology is often used in combination with the other four methodologies. Experiments based on a model are called simulations. When a formal description of the model is created to verify the functionality or correctness of a system, the task is called model checking. [5]

[14]: Four quadrant model: 1. What do we want to achieve? 2. Where does the data come from? 3. What do we do with the data? 4. Have we achieved our goal? Iterative process, can repeat etc.

2.1.2 Humanities Research

2.1.3 Arts Research

2.2 Transdisciplinary Research

Multidisciplinarity

: “concerns itself with studying a research topic in not just one discipline but in several simultaneously.”

Interdisciplinarity

: “has a different goal than multidisciplinarity. It concerns the transfer of methods from one discipline to another.”

Transdisciplinarity

: “concerns that which is at once between the disciplines, across the different disciplines, and beyond all disciplines.”

(Nicolescu 2010)

Problem Focus: (solve complex, multi-dimensional, particular problems)

“TD research therefore starts with a problem that is ‘in the world and actual’ as opposed to ‘in my head and conceptual.’” “This inherent feature of ‘creating change’ highlights the relevance of using the term ‘consequential’ to characterise TD research approaches and problems.” [9]

2.2.1 Practice Based Research

2.3 Research Approach

Part II

T $\ominus\Theta$ LS \ominus F TH Σ TR \forall D Σ

Chapter 3

'Pataphysics

3.1	Conscious	23
3.2	Self-conscious	28
3.2.1	Symbology	31
3.2.2	Antimony	31
3.2.3	Anomaly	32
3.2.4	Sysygy	32
3.2.5	Clinamen	33
3.2.6	Absolute	33
3.3	Unconscious	34
3.3.1	Oulipo	34
3.3.2	Borges	35

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

“To understand 'pataphysics is to fail to understand 'pataphysics.”
(Hugill 2012b)

1

¹Although note how the perplexing apostrophe that sometimes appears before the word 'pataphysics undermines too literal an interpretation of this construction. Jarry only ever used the apostrophe on a single occasion, specifying that he did so “in order to avoid a simple pun”. What that pun might be has never been fully explained.

It is probably impossible to define pataphysics in one sentence. There is no definition that does justice to what pataphysics really means and no single definition that is truer than any other. In fact, the college of pataphysics in France itself has published a book ([Alastair Brotchie, Chapman et al. 2003](#)) with over 100 definitions that they all call equally valid. This chapter therefore begins with a selection of definitions from that book (quoting their original sources).

“Pataphysics . . . is the science of that which is superinduced upon metaphysics, whether within or beyond the latter’s limitations, extending as far beyond metaphysics as the latter extends beyond physics. [...] Pataphysics will be, above all, the science of the particular, despite the common opinion that the only science is that of the general. Pataphysics will examine the laws governing exceptions, and will explain the universe supplementary to this one. [...] DEFINITION: Pataphysics is the science of imaginary solutions, which symbolically attributes the properties of objects, described by their virtuality, to their lineaments.” (Alfred Jarry, “Exploits and Opinions of Dr Faustroll, Pataphysician” written in 1897–8 and published posthumously in 1911) ([Jarry 1996](#))

“‘Pataphysics is patient; ‘Pataphysics is benign; ‘Pataphysics envies nothing, is never distracted, never puffed up, it has neither aspirations nor seeks not its own, it is even-tempered, and thinks not evil; it mocks not iniquity: it is enraptured with scientific truth; it supports everything, believes everything, has faith in everything and upholds everything that is.” (“Épanorthose sur le Clinamen moral”, Cahiers du Collège de ‘Pataphysique, 21, 22 Sable 83 (29 December 1955 vulg.)) ([Alastair Brotchie, Chapman et al. 2003](#))

“‘Pataphysics passes easily from one state of apparent definition to another. Thus it can present itself under the aspect of a gas, a liquid or a solid. (Patafluens 2001, Istituto Patafisico Vitellianese, Viadana, 2002)” ([Alastair Brotchie, Chapman et al. 2003](#))

“‘Pataphysics, ‘the science of the particular’, does not, therefore, study the rules governing the general recurrence of a periodic incident (the expected case) so much as study the games governing the special occurrence of a sporadic accident (the excepted case). [...] Jarry performs humorously on behalf of literature what Nietzsche performs seriously on behalf of philosophy. Both thinkers in effect attempt to dream up a ‘gay science’, whose *joie de vivre* thrives wherever the tyranny of truth has increased our esteem for

the lie and wherever the tyranny of reason has increased our esteem for the mad." (Christian Bök, 'Pataphysics, The Poetics of an Imaginary Science, Northwestern University Press, ill., 2002) (Boek 2002)

La pataphysique est la fin des fins.	'Pataphysics is the end of ends.
La pataphysique est la fin des faims.	'Pataphysics is the end of hunger.
La pataphysique est la faim des fins.	'Pataphysics is the hunger for ends.
La pataphysique est le fin du fin.	'Pataphysics is the finest of the fine.

(The first motto is that of the official Collège notepaper. Its three variants have appeared in elsewhere in Collège publications. — Collège de 'Pataphysique) (Alastair Brotchie, Chapman et al. 2003)

The branch of philosophy that deals with an imaginary realm additional to metaphysics. (Oxford Dictionary)

Epiphany – “to express the bursting forth or the revelation of pataphysics” Dr Sandomir (Hugill 2012b, p.174)

I divided my research about pataphysics into four approaches. The first: learn about its inventor — Alfred Jarry. The second: read his work. The third: read what others have to say about pataphysics. The fourth: read other literature that could be classed as pataphysical. Eventually I ended up seeing pataphysics in everything, there was no escaping it anymore. I had turned a Pataphysician.

Personally, when I try to explain pataphysics to laymen, I use the last scene of the movie 'Men In Black' as an example. The scene zooms out further and further, from a close up of Will Smith, to his car, to a shot of the city from above, to a shot of the earth, the galaxy, the universe and finally it is revealed that the universe is in a marble that is being toyed with by an alien. This is a good example of different layers of abstraction – Will Smith in his car represents the physical layer, the universe the metaphysical layer and the alien marble the pataphysical layer. The outro scene can be seen on YouTube².

²<http://youtu.be/1QP11-TKaEE>



Figure 3.1: Men in Black Screenshots of Ending Sequence

3.1 Conscious

Jarry was “attempting to transcend his own existence.” ([Hugill 2012a](#))

“It is certainly true that making life ‘as beautiful as literature’ was one of his goals.” ([Hugill 2012a](#))

Alfred Jarry was born in Laval, Mayenne, France in 1873 and died in Paris in 1907, at the age of 34. He was known as a poet, dramatist, novelist and journalist but also as a graphic artist. His hobbies included entomology, fishing, cycling, fencing, shooting and drinking.

He went to school in Rennes, where his physics teachers Félix-Frédéric Hébert left such a big impression on Jarry that he would later be his inspiration for Père Ubu. He passed his baccalauréat with 17 and moved to Paris to attend the lycée Henri IV in preparation to apply for admission to the École Normale Supérieure but eventually gave up on the entrance exam after several unsuccessful attempts. He met another teacher at the lycée, this time a philosophy teacher called Henri Bergson, who inspired him greatly. He published his first collection of poems in 1893, aged 20, the year his mother died. One of his classmates there described him as follows.

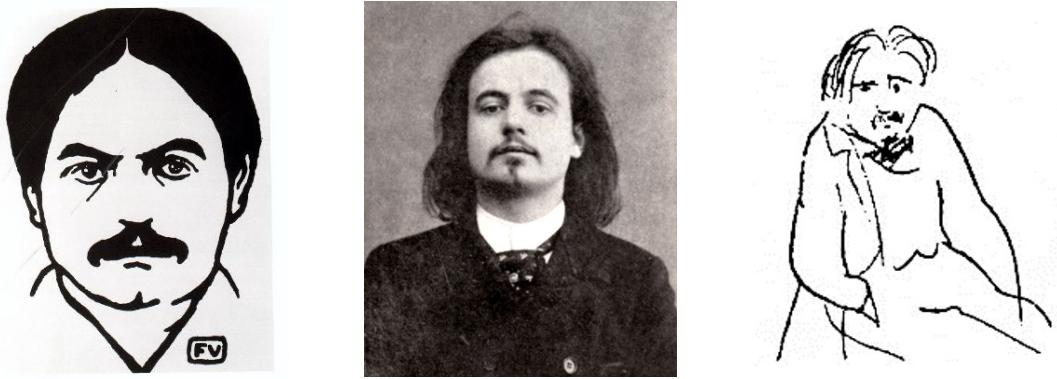


Figure 3.2: FelixVallotton + jarry + jpicaso

[...] I found Jarry's mental processes disturbing. When he let himself go he seemed in thrall to a torrent of words outside his control. It was no longer a person speaking, but a machine controlled by a demon. His staccato voice, metallic and nasal, his abrupt puppet-like gestures, his fixed expression and uncontrolled flood of language, his grotesque and brilliant turns of phrases, ended up provoking a feeling of disquiet. He was informed, intelligent, and discriminating; he was good person, secretly kind, perhaps even shy beneath it all [...] but his originality resembled nothing short of a mental anomaly. (Jarry's classmate at the lycée Henri IV: Gandilhon Gens-d'Armes "Alfred Jarry au lycée Henri IV" Les Marges, XXIII, 91 (15 Jan 1922) as quoted in ([Alastair Brotchie 2011](#)))

He was at the centre of the avant-garde movement in Paris around that time, at the centre of the Tuesday meetings of the Mercure de France (a literary magazine run by Alfred Valette and his wife Rachilde, who soon became a sort of substitute family to Jarry who was roughly 15 years younger than them). Being rather misogynist at times and homosexually inclined, Rachilde was one of his very few female friends.

The following year, 1895, he briefly joined the army in the 101st Infantry, after having dodged it by being an enrolled student at the lycée. He followed rules there pedantically but hated the loss of his individualism. According to Brotchie, he "chose subservience, but subservience taken to the point of parody: the pata-physical solution to the problem of obedience" ([Alastair Brotchie 2011](#)). Probably the only thing he enjoyed there was the fencing and shooting training. He looked funny in the uniform that was too big for him being so small (5'3") so he was eventually excused from parades and after a few months he was allowed to leave to Paris frequently. He was discharged in December 1895 on medical grounds: gallstones. It is not unlikely that he faked the illness by drinking picric acid.



Figure 3.3: Woodprint of Ubu by Alfred Jarry

His father had died just two months earlier and had left him a small inheritance, which he spent mostly on publishing his very own magazine dedicated to symbolist wood carvings, the *Perhinderion*. He had previously co-edited the magazine *L'Ymagier* with Remy de Gourmont between 1893 and 1894. He joined Aurélien Lugné-Poë as his secretary (his only ever real job) at the Théâtre de l'Œuvre after his discharge at the army, where he would pour his utmost attention to putting his *Ubu* play on the stage. He also played a small role in the production of *Peer Gynt* at the *Œuvre* earlier in 1896. The printed version of *Ubu Roi* appeared in *Le Livre d'Art* in the middle of the year with Jarry's carved woodcut image of Ubu that became so popular. The première took place on 10th December that year and caused an outrage in the audience after the first word: "merdre" (sometimes translated as "pshit"). Jarry had previously arranged for certain friends to counter any reaction of the general audience and to prevent under all circumstances for the play to reach its conclusion. The performance went according to plan. The uproar after the first word was uttered was immense, the performance had to be interrupted at times to calm the audience and it finished in shouts of praise, protest and insults. There were no further performances but the event was considered historic even at the time and is now widely seen as the first "modern" play (Alastair Brotchie 2011, p.168-169). And as Dave Walsh puts it: "Movements such as Dadaism, Surrealism, Futurism, Expressionism Cubism, Theatre of the Absurd — all owe debts to [Jarry's] works." (Walsh 2001)

Although Ubu's mannerism of speech was originally imitating Jarry's, as suggested by Lugné-Poë (Alastair Brotchie 2011, p.155), Jarry continued to adapt Ubu's mannerisms.

"Those who knew him said that his nauseating appearance hid a youth who was stubborn yet shy, proud and little full of himself, but good-natured and ingenuous behind his cynicism, one who was fiercely independent and rigorously honest." (Henri de Régnier, as quoted in ([Alastair Brotchie 2011](#), p.181))

"Alfred Jarry had a very particular way of speaking to that was disconcerting to those who heard it for the first time. He said 'we', when referring to himself, and substituted verbs for nouns, in imitation of ancient Greek. Example: 'celui qui soufflé' (that which blows) for the wind, and 'celui qui se traîne' (that which crawls along) for the train, even if it was an express! This made conversation somewhat complicated, not least because of the rapidity of his delivery." (Rachilde, as quoted in ([Alastair Brotchie 2011](#), p.181))

"Alfred Jarry was a man of letters to an unprecedented extent. His smallest actions, his childish pranks, everything he did was literature. His whole life was shaped by literature, and only by literature." (Apollinaire, as quoted in ([Alastair Brotchie 2011](#), p.307))

Jarry spent the next few years writing. He had spent all his inheritance on the publication of his magazine and the production of Ubu Roi. It is during this time that he moved to his infamous tiny flat on the second-and-a-half floor. Jarry could just about stand upright but any guests had to crouch. He had no electricity or gas and no means of cooking ([Alastair Brotchie 2011](#), p.195). In December 1897 he formed a marionette theatre with his friend Claude Terasse: the Théâtre de Pantins and they performed Ubu Roi in January 1898 without riots in the audience.

Jarry then gradually withdrew from the literary circles in Paris and spent more time in a little shack on the banks of the Seine near the village of Le Coudray. He started writing a regular review column for the Revue Blanche in 1900, the income of which he certainly needed much. There was a brief revival of the Ubu marionette play in the Cabaret des Quat'z'Arts in 1901.

Around 1904 he began drinking ether, the absinthe not strong enough anymore. In the winter of 1905 he was very ill, the cold and poverty not helping. In 1906, his friends became more and more concerned about his deteriorating health and eventually Valette and Saltas sent him to his sister Charlotte. He then spent some time in Paris and some in Laval at his sister's place over the next year. Jarry then died in November of 1907 of meningeal tuberculosis. His last request was for a toothpick.

“He believes that the decomposing brain goes on working after death and it is its dreams that are Paradise.” (Jarry 1906 in a letter to Rachilde ([Alastair Brotchie and Chapman 2007](#)) — ‘he’ refers to himself)

Studying Jarry’s life gives certain insights into the man who created pataphysics and why he might have done so. Alastair Brotchie has written the probably most concise and recent biography of Alfred Jarry in English language and most of the information summarised here comes from this book ([Alastair Brotchie 2011](#)). Roger Shattuck gives a very nice introduction about Jarry in relation to the time and place he lived in, in his book “The Banquet Years” ([Shattuck 1959](#)). However, he does not focus on Jarry alone but rather on the time period and four personalities (Alfred Jarry, Henri Rousseau, Erik Satie and Guillaume Apollinaire) he chose as representatives of the era.

His Writing

Jarry has written a good amount of texts in his short life and he didn’t confine himself to a single category either. He wrote poems, novels, short stories, essays, art reviews, theatre reviews and plays and also produced translations of a few texts into French. Many of his texts were completely fictional, some had autobiographical aspects and some scientific and most of them had a sarcastic sense of humour. Trying to summarise Jarry’s style of writing or attempting to interpret his whole body of work seems impossible though.

“Jarry was an acknowledged classical scholar, had already worked as a reviewer of art and drama, had edited two art magazines, was up to date with modern scientific theory, especially physics, read widely in mathematics and psychology, and had an extensive basic knowledge of philosophy.” ([Alastair Brotchie 2011](#))

James A Cutshall says that “instead of Jarry the man and the meaning of his literary endeavours becoming clearer with the passage of time, both have become increasingly indistinct” ([Cutshall 1988](#), p.246). The intention of his thesis was to show the seriousness implied behind the humour in many of Jarry’s novels, in order to give the author the merit he deserved. Cutshall wrote about Jarry’s novels rather than simply seeing him as the playwright of the Ubu plays. He surveyed existing criticism about Jarry’s texts and provided his own view on them. He immortalised Jarry by saying “whether or not this is the sort of ‘éternité’ sought by the heroes of Jarry’s novels, it is certainly that which their author somewhat belatedly has found” ([Cutshall 1988](#), p.248).



Figure 3.4: Faustroll illustration by Steve Morrison

3

Cutshall was not the only one who has written about certain less-known texts by Jarry. Marieke Dubbelboer's thesis "Ubusing Culture" is also interesting in this regard since it concentrates completely on the "Almanachs du Père Ubu" (published in 1898 and 1901) (Dubbelboer 2009). She was looking for keys to Jarry's poetics in those texts, which she says "seemed to defy labelling or literary norms" (Dubbelboer 2009, p.10). She claims the Almanacs to be quite radical and exemplary of his innovative poetics moving away from Symbolism and towards the Avant-Garde. In general she says his work "can be characterized as playful, elusive, paradoxical and provocative" (Dubbelboer 2009, p.197) and his two Almanacs are the essence of his non-conformist attitude. They were written at a time of change for Jarry, when he withdrew from his usual circles in Paris and he published in new magazines, which links his change in writing according to Dubbelboer.

A list of his works can be found in the appendix [A](#).

ref

3.2 Self-conscious

We will need to understand the essence of pataphysics to understand how it relates to creative computing.

³<http://bit.ly/1Q0VZW9>

Jarry first defined pataphysics in his book “Exploits and Opinions of Dr Faustroll, Pataphysician” written in 1898 and published posthumously in 1911 ([Jarry 1996](#)). But the concept appeared as early as in 1893 in his prose text Guignol that won him a prize in the newspaper L’Echo de Paris and it appears in many of his writings. He originally intended to write a whole book called “Elements of Pataphysics” but only part of this appeared in Faustroll.

Zoe Corbyn gives a very simple short introduction for beginners of the topic in an article in the Guardian ([Corbyn 2005](#)) in 2005. She describes it like this:

Correct definitions are equivalent to wrong ones; all religions are on a par as imaginary and equally important; chalk really is cheese. It’s an escape from reality — reminding us of just how idiotic the rules that dog our everyday existence are. ([Jarry 1996](#))

Jean Baudrillard has a few other definitions for pataphysics in his text ([Baudrillard 2007](#)). According to him pataphysics is “the highest temptation of the spirit, the nail in the tire, the philosophy of the gaseous state, the science or the unique imaginary solution to the absence of problems” to name just a few.

Another rather strange interpretation of pataphysics is Asger Jorn’s. He calls pataphysics a religion in the making ([Jorn 1961](#)). He claims that since “natural religion is the spiritual confirmation of material existence”, “metaphysical religion represents the establishment of an ever deepening rift between material and spiritual life.” He refers to the idea of equivalence in pataphysics and the absolute and links them to religion. He says “the great merit of Pataphysics is to have confirmed that there is no metaphysical justification for forcing everybody to believe in the same absurdity”.

Cruickshank ([Cruickshank nd](#)) wrote a rather funny article on anti-matter. He links the creation of anti-matter atoms at CERN around 1996 with Jarry, saying that he had “beaten them to the punch” with his pataphysics.

Christian Bök ([Boek 2002](#)) tries to draw science and poetry together using pataphysics as the string that binds them. He compares Jarry and Nietzsche, saying Jarry performs humorously on behalf of literature what Nietzsche performs seriously on behalf of philosophy; both try to create an antiphilosophy ([Boek 2002](#), p.9). He also claims that science and poetry have a similar history, undergoing the same four phases of distinct change but also that they have not evolved in sync with each other ([Boek 2002](#), p.15).

Animalistic phase

: signs exist long before being known, they are written by nature

Mechanistic phase

: signs exist by being known, they are written by culture

Organismic phase

: signs evolve by being known, they are written across events by culture

Cyborganic phase

: signs evolve beyond being known, they are written as events by culture

Pataphysics is a surreal perspective that has had an extensive, yet forgotten, influence upon the canon history of radical poetics. [...] Not only does this avant-garde pseudoscience valorise whatever is exceptional and paralogical; it also sets the parameters for the contemporary relationships between science and poetry. ([Boek 2002](#), p.27)

Bök also compares Jarry and Nietzsche in regards to perspectivism ([Boek 2002](#), p.31). For Nietzsche reality is the effect of a dream world in which “there are many kinds of truths, and consequently there is no truth”. And similarly for Jarry, reality is an aspect of eternity in which “there are only hallucinations, or perceptions” and every “perception is a hallucination which is true”. Both argue that no view is absolute as well and pataphysics argues that every viewpoint is dissolute, including its own because no view can offer a norm. Even Jarry’s eternity is nowhere and somewhere at the same time.

In Faustroll, Bök says, “Jarry parodies the discourse of such scientific luminaries, who attempt to demonstrate the utility of science through the dramaturgic performance of a mechanical experiment” ([Boek 2002](#), p.29).

According to the Collège de 'Pataphysique, it is convention to use the apostrophe at the beginning of the word ('Pataphysics) only in reference to Jarry's texts, to the science of imaginary solutions as such. Used as an adjective or in a more unconscious way it is written without the apostrophe. Jarry himself just indicated that the word is preceded by the apostrophe to avoid a pun.

- Vian, B. (2006). 'Pataphysics? What's That? (S. Chapman, Trans.). London: Atlas Press.([Vian 2006](#))
- Daumal, R. (2012). Pataphysical Essays. (T. Vosteen, Trans.). Cambridge, Massachusetts: Wakefield Press.([Daumal 2012](#))
- Brotchie, A. (Ed.). (1995). A True History of the College of 'Pataphysics — 1. (P. Edwards, Trans.). London: Atlas Press.([Alistair Brotchie 1995](#))

3.2.1 Symbology

Probably the most famous symbol of pataphysics is the grand gidouille, the big spiral on Ubu's fat belly. Not simply because it is a feature of Jarry's most popular creation but also because it represents one of the concepts of pataphysics itself: the antimony. The spiral can be interpreted as two spirals in one, the outer and the inner spiral. They represent the duality of pataphysics, the mutually incompatible in perfect harmony. The Collège de 'Pataphysique has adopted the spiral for its membership badges, in various colours and sizes for the different ranks of the college.

Another symbol of pataphysics is the green candle which refers to one of Jarry's last endeavours, published posthumously, a vast collection of his journalistic essays ([Hugill 2012a](#)). Some animals also symbolise pataphysics. The crocodile, the current vice-curator of the college is a crocodile named Lutembi ([Hugill 2012a](#)). Owls are another symbol; Jarry kept stuffed and live owls ([Alastair Brotchie 2011, p.46](#))[13 p46] in his flat. The chameleon is another, having the ability to change colour and looking in two directions at the same time.



Figure 3.5: Crocodile from the CoP website



Figure 3.6: The Grand Gidouille



Figure 3.7: The green candle

We argue that pataphysics can facilitate creative computing. A pataphysical grammar consists of exceptions, syzygies, anomalies, clinamen, antinomies, contradictions, equivalents and imaginaries. Such constraints can transform the ways in which we may navigate and transform our conceptual space. Pataphysical concepts are likely to cause surprise and could therefore be considered unconventional and provocative.

3.2.2 Antimony

The antimony is the mutually incompatible. It appears everywhere in Jarry's writings. It represents the duality of things, the echo or symmetry, the good

and the evil at the same time. Examples are the plus minus, the faust-troll, the haldern-ablou, the yes-but, the ha-ha and the paradox.

The ‘Ha Ha’, the only words Bosse-da-Nage ever utters in Faustroll, “is the idea of duality, of echo, of distance, of symmetry, of greatness and duration, of the two principles of good and evil.” ([Hugill 2012a](#)) Referring to the yes-but statement Hugill says “this may be taken as a standard pataphysical response to any proposition (including this one).” And most obviously the antimony can be seen in all the contradictions that pataphysics is so fond of.

The antinomy, in a pataphysical sense, is the mutually incompatible or paradox. Mutually contradictory opposites can and do co-exist in the pataphysical universe.

3.2.3 Anomaly

The anomaly is the exception. And exceptions are important in pataphysics. But then again everything is equal, so in a pataphysical world no exceptions would exist at all, or rather, everything would be equally exceptional. The anomaly disrupts and surprises. Hugill mentioned a great example of a collection of anomalies: the sourcebook project by William Corliss, who collects scientific papers that are anomalous. Bök says it is “the repressed part of a rule which ensure that the rule does not work” ([Boek 2002](#), p.38).

3.2.4 Sysygy

The syzygy surprises and confuses. It originally comes from astronomy and denotes the alignment of three celestial bodies in a straight line. In a pataphysical context it is the pun. It usually describes a conjunction of things, something unexpected and surprising. Serendipity is a simple chance encounter but the syzygy has a more scientific purpose. Bök mentions Jarry saying that the fall of a body towards a centre is the same as the ascension of a vacuum towards a periphery ([Boek 2002](#), p.42).

A syzygy both surprises and confuses. The concept originally comes from the field of astronomy where it denotes the alignment of three celestial bodies. In a pataphysical context it usually describes a conjunction of things, something unexpected and surprising. Unlike serendipity, a simple chance encounter, the syzygy has a more scientific purpose. A typical instance is the pun, which Jarry called the syzygy of words ([Jarry 1996](#)). Next to being intentionally funny, puns demonstrate a clever use (or abuse) of grammar, syntax, pronunciation and/or semantics, often taken to a quite scientific level, such that without understand-

ing of what is said and what is the intended meaning, the humour of the pun might be lost.

3.2.5 Clinamen

The clinamen is the unpredictable swerve that Bök calls “the smallest possible aberration that can make the greatest possible difference” (Boek 2002, p.43). He links it to Lucretius idea of an atom serving in its streamlined flow to create matter and to Epicurus’ parenklisis. But he also points out similarities to ideas like the Situationists’ ‘détournement’, the reuse of pre-existing aesthetic elements and Hugill links it to the Dadaists’ ready-mades and Oulipo’s verbal games. An obvious example is Jarry’s ‘merdre’, a swerve of the French word for shit (merde).

The concept of the clinamen can be understood as an unpredictable swerve which Bök called the smallest possible aberration that can make the greatest possible difference (Boek 2002). One of the most famous examples of a clinamen is Jarry’s merdre (the first word in his Ubu plays). He squeezed an extra ‘r’ into the French word merde (meaning shit) and translates into something like pshit.

3.2.6 Absolute

The absolute is a reference to a transcended reality. Jarry talks about ‘ethernity’ in Faustroll (Jarry 1996, p.104).

Others

Other concepts that are pataphysical or can be linked to it in a sense are alchemy and quantum mechanics. Alchemy because of its laws or equivalence and the union of opposites (Hugill 2012a) and quantum mechanics because of principles of uncertainty, indeterminacy and the idea of the multiverse of course.

Because string theory is speculation based on ideas that are themselves speculative (i.e., theories of general relativity and quantum mechanics), string theory is not in fact physics, but 'pataphysics.

Likewise, string theory and quantum calculations are, increasingly, not descriptive of an actual reality, but are simply mathematical pataphors. (P. Lopez)⁴

⁴<http://www.urbandictionary.com/define.php?term=pataphysics>

3.3 Unconscious

3.3.1 Oulipo

Finish section here. references and all

Potential literature is “the search for new forms and structures that may be used by writers in any way they see fit.” Raymond Queneau (p2)

The Oulipo’s goal is to discover new structures and to furnish for each structure a small number of examples. François Le Lionnais (p3)

“a formal quest” Warren Motte (p3)

Erecting the aesthetic of formal constraint, then, the Oulipo simultaneously devalues inspiration. (p10)

Three levels in the hierarchy of constraints:

1 Minimal level: constraints on the language in which the text is written

2 Intermediate level: constraints on genre and certain literary norms

3 Maximum level: consciously preelaborated and voluntarily imposed systems of artifice

Oulipo is in the maximum level. François Le Lionnais (p11)

Oulipian systems of formal constraint are often based on the alphabet. François Le Lionnais (p13)

The nature of Oulipoan constraint is mathematical. François Le Lionnais

The Oulipo is anti-chance Claude Berge (p17)

What is the objective of our work? To propose new ‘structures’ to writers, mathematical in nature, or to invent new artificial or mechanical procedures that will contribute to literary activity: props for inspiration as it were, or rather, in a way, aids for creativity. Raymond Queneau (p51)

Aleatoricism is the incorporation of chance into the process of creation, especially the creation of art or media. The word derives from the Latin word alea, the rolling of dice. It should not be confused with either improvisation or indeterminacy.[1] Wikipedia

3.3.2 Borges

You could argue that by reading other literature that is pataphysical (whether or not it was intended) one can learn something about pataphysics. Reading Borges ([Borges 1964](#); [Borges 1999](#); [Borges and Guerrero 1957](#); [Borges and Dembo 2010](#); [Borges 2010](#); [Borges 2000](#)) is a good example. His text “The analytical language of John Wilkins” ([Borges 2000](#)) contains a brilliant example of pataphysical thinking and coincidentally a good example of the kinds of search results my search tool should hopefully produce.

Referring to a certain Chinese dictionary entitled “The Celestial Emporium of Benevolent Knowledge” he claims that animals can be divided into:

1. those belonging to the Emperor
2. those that are embalmed
3. those that are tame
4. pigs
5. sirens
6. imaginary animals
7. wild dogs
8. those included in this classification
9. those that are crazy-acting
10. those that are uncountable
11. those painted with the finest brush made of camel hair
12. miscellaneous
13. those which have just broken a vase
14. those which, from a distance, look like flies

This kind of categorisation has also been discussed by Foucault in his book ‘The Order of Things’ ([Foucault 1966](#)).

Chapter 4

Creativity

4.1	In Humans	40
4.1.1	Mel Rhodes and Ross Mooney	40
4.1.2	Arthur Koestler	41
4.1.3	Henri Poincaré, Graham Wallas and George Pólya	42
4.1.4	James Kaufman and Ron Beghetto	43
4.1.5	Margaret Boden	45
4.1.6	Robert J. Sternberg, James C. Kaufman and Timothy Leary . .	47
4.2	In Computers	48
4.2.1	Bipin Indurkha	49
4.2.2	Partridge and Rowe	50
4.2.3	David Gelernter	51
4.2.4	Marvin Minsky	52
4.2.5	Matthew Elton	52
4.3	In Academia	53
4.3.1	Computational Creativity	55
4.3.2	Creative Computing	56
4.3.3	Speculative Computing	59
4.3.4	Digital Humanities	61
4.3.5	Computer Ethics	68

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

finish intro to creativity and computers chapter

Educational psychologist Richard Mayer identified several different approaches to human creativity research and each approach has its own methodologies and goals. (Mayer 1999, p.453)

Psychometric

(creativity as a mental trait): quantitative measurement, controlled environments, ability based analysis

Psychological

(creativity as cognitive processing): controlled environments, quantitative measurements, cognitive task analysis

Biographical

(creativity as a life story): authentic environments, qualitative descriptions, quantitative measurements

Biological

(creativity as a physiological trait): physiological measures

Computational

(creativity as a mental computation): formal modelling

Contextual

(creativity as a context-based activity): social, cultural and evolutionary context

There are many debates across the involved disciplines. Specifically, Mayer identified five big questions of human creativity research: (Mayer 1999, p.450-451)

1. Is creativity a property of people, products, or processes?
2. Is creativity a personal or social phenomenon?
3. Is creativity common or rare?
4. Is creativity domain-general or domain-specific?
5. Is creativity quantitative or qualitative?

An important challenge for the next 50 years of creativity research is to develop a clearer definition of creativity and to use a combination of research methodologies that will move the field from speculation to specification. (Mayer 1999, p.459)

“In research communities, approaches to the study of creativity differ in three main respects: 1) the type of research design, whether

experimental, psychometric, observational etc. 2) the focus of the research, whether on human attributes cognitive processes or features of creative outcomes, and 3) the type of information that is used for the basis of evidence, by which is meant whether the time frame is present (real-time observation) or past (historical data) and whether the situation is artificial (laboratory) or natural (real world settings)." (Candy 2012, p.3)

Creativity is a human quality — a process. Computers are essentially a product process of human creativity. product

Models of human creativity don't necessarily apply to computer creativity.

Creativity can be studied at various **levels** (neurological, cognitive, and holistic/systemic), from different **perspectives** (subjective and objective) and **characteristics** (combinational, exploratory and transformative).

This chapter introduces relevant models of human and computer creativity and describes the disciplines of computational creativity and creative computing.

These two simple statements already point to one of the main problems with evaluating creative computer software: do we evaluate the process or the product? See § 10.

put summaries at back of chapter or front? styling?

Summary:

- novelty/typicality/acceptability/variety/imagination/originality
- quality/value/appreciation/appropriateness/usefulness/relevance (/surprising?)
- efficiency/skill
- subjective/P/little-c
- objective/H/Big-C
- combinational, exploratory and transformative
- product/process
- The 4 P's
- Unified theory
- Associative and bisociative thinking
- Creative triptych (humour, discovery, art)
- 4 step model
- Problem solving
- P-creativity and H-creativity

4.1 In Humans

general introduction about human creativity models

Let us define creativity as *the ability to use original ideas to create something new and surprising of value*. We generally speak of creative ideas rather than products, since creative products merely provide evidence of a creative process that has already taken place.

“Creativity is the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context” (Plucker et al., 2004, p. 90) ([Jordanous and Keller 2012](#))

4.1.1 Mel Rhodes and Ross Mooney

Mel Rhodes (1916–1976), who has a background in education and psychology, identified four common themes of creativity in 1961, which he termed “the four P’s of creativity” ([Rhodes 1961](#)):

Persons

personality, intellect, temperament, physique, traits, habits, attitudes, self-concept, value systems, defence mechanisms and behaviour.

Process

motivation, perception, learning, thinking and communication.

Press

relationship between human beings and their environment

Products

a thought which has been communicated to other people in the form of words, paint, clay, metal, stone, fabric, or other material.

Rhodes highlights the importance of a holistic view on creativity through these four areas of study, which he hoped would become the basis of a unified theory of creativity.

Where, what, who and how – those are the questions we need to ask regarding creativity.

Ross Mooney identified four aspects of creativity in 1963 (as cited in ([Sternberg 1999](#))) which are essentially the same.

1. The creative environment

2. The creative person
3. The creative process
4. The creative product

Mel Rhodes vs Ross Mooney??

I believe these four aspects represent four conditions of creativity in a way, since all of them influence the outcome. Sometimes the creative environment and the creative person are merged into one, simply because people are often a product of their environment. The person will always be influenced by its surroundings, its culture, family, etc. and the environment alone cannot influence anything if not through a person. Another interpretation of course is that the creative environment is the context in which the creative product exists. Depending on which way we look at it, the first and second can be interchanged. This project focuses on the creative product, making sure the essence of the search tool is creative – the algorithms that define the main functionality of the search tool.

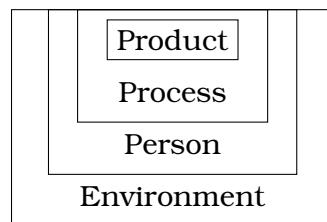


Figure 4.1: 4 Aspects of Creativity

Figure 4.1 shows how these aspects relate to each other. The environment influences all others and the person creates the product in a process.

4.1.2 Arthur Koestler

Arthur Koestler (1905–1983) published his study on creativity entitled “The Act of Creation” in 1964 ([Koestler 1964](#)). The book still carries influence today. His main contribution to the field is probably the concept of **bisociation**, a term he coined for the idea of two “self-consistent but habitually incompatible frames of reference” intersecting to give rise to new creative idea ([Koestler 1964](#), p.35). It is interesting however to look at some of his other views on creativity as well.

He splits creativity into three domains, a triptych, without sharp boundaries: humour, discovery and art (see table 4.1). All creative acts traverse the three domains of this triptych from left to right, that is, the emotional climate of the creator changes “from an absurd through an abstract to a tragic or lyric view of

existence” during the process (Koestler 1964, p.27). Central to all three domains is the “discovery of hidden similarities”, or bisociation. Koestler differentiates between associative thinking and bisociative thinking. He links those broadly to habit and originality, respectively. More specifically, associative thinking is conscious, logical, habitual, rigid, repetitive and conservative and bisociative thinking is unconscious, intuitive, original, flexible, novel and destructive/constructive.

Humour	\Rightarrow	Discovery	\Rightarrow	Art
Laugh		Understand		Marvel
Riddle		Problem		Allusion
Debunking		Discovering		Revealing
Coincidence		Trigger		Fate
Aggressive		Neutral		Sympathetic

Table 4.1: Koestler: Creative Triptych

4.1.3 Henri Poincaré, Graham Wallas and George Pólya

Henri Poincaré (1854–1912) (Poincare 2001) and Graham Wallas (1858–1932) (Wallas 1926) have defined a popular model (Boden 2003; Koestler 1964; Partridge and Rowe 1994) of the creative process (it was suggested by Poincaré ((Poincare 2001) book: “science and method”, chapter III:“mathematical discovery”, pages 387–400) and formulated by Wallas).

1. Preparation – focusing the mind on the problem
2. Incubation – unconscious internalising
3. Illumination – eureka moment from unconsciousness to consciousness
4. Verification – conscious evaluation of the idea and elaboration...

Weisberg criticises the stages of incubation and illumination (referred to by Partridge and Rowe 1994), saying that the creative process is really just simple problem solving, and that incubation is what he calls “creative worrying”.

“First, we have to **understand** the problem; we have to see clearly what is required. Second, we have to see how the various items are connected, how the unknown is linked to the data, in order to obtain the idea of the solution, to make a **plan**. Third, we **carry**

out our plan. Fourth, we **look back** at the completed solution, we review and discuss it." (Polya 1957, p.5-6, his emphasis)

4.1.4 James Kaufman and Ron Beghetto

DOB of authors?

James C. Kaufman (1974-) and Ronald A. Beghetto (DOB?)... (See [Kaufman and Beghetto 2009](#)).

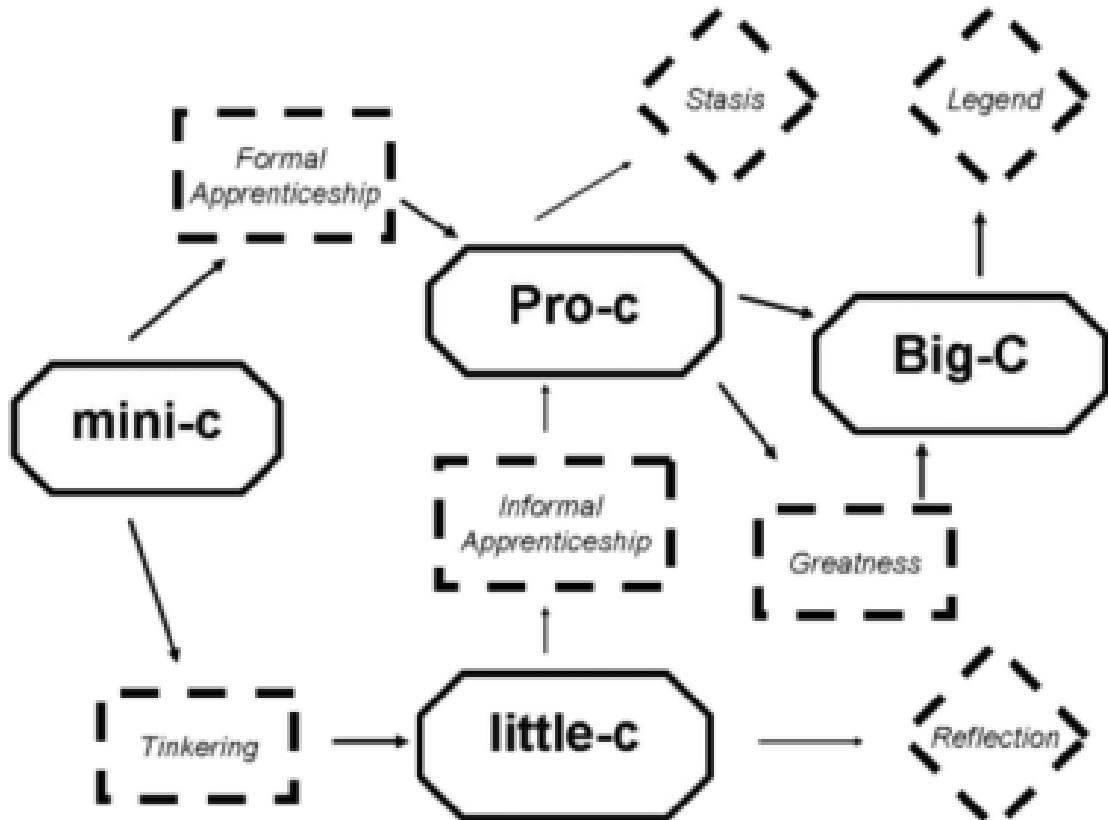


Figure 4.2: The 4 C Model

redo diagram

Big-C

Eminent Accomplishments. Big-C creativity consists of clear-cut, eminent creative contributions. Big-C creativity often requires a degree of time. Indeed, most theoretical conceptions of Big-C nearly require a posthumous evaluation.

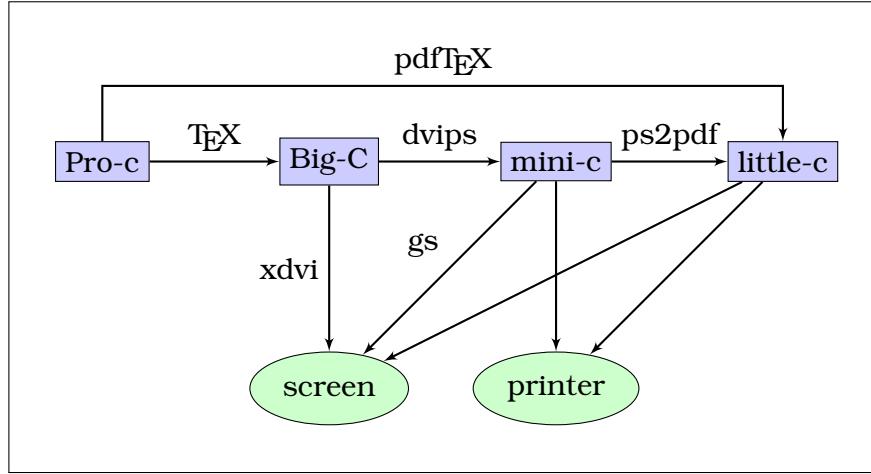


Figure 4.3: The 4 C Model2

Pro-c

Professional Expertise. Pro-c represents the developmental and effortful progression beyond little-c. The concept of Pro-c is consistent with the expertise acquisition approach of creativity.

[ref](#)

Ericsson1996, Ericsson2007 Propulsion Theory of Creative Contributions ([Sternberg 1999](#); [Sternberg 2006](#)): Replication, redefinition, forward incrementation, advance forward incrementation. Redirection, Reconstruction, reinitiation, integration.

Little-c

Everyday Innovation. More focused on everyday activities, such as those creative actions in which the non-expert may participate each day.

Mini-c

Transformative Learning. Encompasses the creativity inherent in the learning process. “Mini-c is defined as the novel and personally meaningful interpretation of experiences, actions, and events.” ([Beghetto and Kaufman 2007](#)) Central to the definition of mini-c creativity is the dynamic, interpretive process of constructing personal knowledge and understanding within a particular sociocultural context. “a transformation or reorganization of incoming information and mental structures based on the individual’s characteristics and existing knowledge” **[p.63]Moran2003**

[ref](#)

Moreover, mini-c stresses that mental constructions that have not (yet) been expressed in a tangible way can still be considered highly creative. Mini-c highlights the intrapersonal, and more process focused aspects of creativity.

Applies to all: openness to new experiences, active observation, and willingness to be surprised and explore the unknown.

4.1.5 Margaret Boden

Professor Margaret Boden (1936-) is a prominent figure in the fields of [Creative Computing \(CC\)](#) and computational creativity. She has a background in medical sciences, psychology and philosophy and currently works as a cognitive scientist in computer science and artificial intelligence. Her main interest is in how the human mind works and how computer models of the mind and specific thinking processes can help us understand both better. She has provided two important contributions to the field. The first is her description of three distinct forms of creativity and the second is her important distinction between two senses of creativity ([Boden 2003](#)).

[Creativity is] the ability to come up with ideas or artefacts that are **new, surprising and valuable**. ([Boden 2003](#)) (her emphasis).

She identified three distinct forms or cognitive processes of how creativity can happen. These are combinational, exploratory and transformational creativity, which can happen at the same time. ([Boden 2003](#))[17, 21].

Combinational creativity

making unfamiliar combinations of familiar ideas; juxtaposition of dissimilar; bisociation; deconceptualisation

Exploratory creativity

exploration of conceptual spaces; noticing new things in old spaces

Transformative creativity

transformation of space; making new thoughts possible by altering the rules of old conceptual space

Central to these three forms is the idea of a **conceptual space**. For any idea, its conceptual space describes the characteristics and constraints that define it in its most fundamental way. The conceptual space of a tea cup would contain information like: it is a container that can hold a hot fluid, it should hold about a half a pint of fluid and it might or might not be built in such a way as to not burn the hand that carries it. The specific colour of the cup or what material it is made of for example are not contained in its conceptual space.

Combinational creativity is the most common form of the three and is concerned with the unusual juxtaposition of common ideas. This aspect is highlighted

in her definition of creativity, which requires novelty and surprise. The main idea is that any particular combination of ideas has to be unusual, causing surprise, but not (necessarily) the individual ideas themselves. She safeguards against purely random combination by including the usefulness of the result as a requirement in the definition. Exploratory creativity requires a person (or computer program) to fully explore the conceptual space of an idea and find unusual or interesting aspects of it. This form of creativity is about pushing an idea to its limits. Transformational creativity takes this exploration one step further. Once the limits of an idea have been identified, they can be transformed. This means that we can step out of the normal conceptual space of an idea, create a new one, alter or ignore the given constraints, add new ones, etcetera.

Boden argues that creative ideas are surprising because they go against expectations ([Boden 2003](#)). She also believes that constraints support creativity and are even essential for it to happen.

Constraints map out a territory of structural possibilities which can then be explored, and perhaps transformed to give another one.
([Boden 2003](#))

These three forms of creativity can be then interpreted on two levels. Any idea should be viewed and evaluated at the appropriate level. Consider the following scenario. A child and a professional architect both build a corbelled arch out of material available to them. Who is being creative here? The level of expertise is clearly different between the two. The child has no experience and is experimenting with the possibilities and limitations of the building blocks (exploring their conceptual space) while the architect has studied the technique for years and is simply applying knowledge he has learned from others (familiar use of a familiar idea). Clearly the child is being more creative in this example. Boden proposed to view and judge the creativity of these two persons separately by differentiating between two levels of creativity, a personal one and a historical one. **Psychological creativity** (P-creativity) is a personal kind of creativity that is novel in respect to an individual and **historical creativity** (H-creativity) is fundamentally novel in respect to the whole of human history. The child in the earlier scenario was P-creative but the architect was neither, he was simply applying his trained skills.

P-creativity involves coming up with a surprising, valuable idea that's new to the person who comes up with it. It doesn't matter how many people have had that idea before. But if a new idea

is H-creative, that means that (so far as we know) no one else has had it before: it has arisen for the first time in human history. (Boden 2003)

Boden suggests that it is helpful to regard aspects such as novelty, quality and process as dimensions of creativity. Instead of asking 'is x creative?' (assuming a boolean judgement) or 'how creative is x?' (assuming a linear judgement) we should ask 'where does x lie in creativity space?' (assuming an n-dimensional space for n criteria where we can measure each dimension). (Pease, Winterstein and Colton 2001, p.8)

Boden argues that process does matter, stating that a program is creative only if it produces items in the right way — by transforming the boundaries of a conceptual space. This, she claims, can only be done if the program contains reflexive descriptions which mark its own procedures and is capable of varying them. The program should contain a meta-level which assesses methods of transforming a space and considers when and how to apply them. (Pease, Winterstein and Colton 2001, p.8)

4.1.6 Robert J. Sternberg, James C. Kaufman and Timothy Leary

Sternberg and Kaufman identify a set of personality traits that are associated with creative people in their "Handbook of Creativity" (Sternberg 1999; Sternberg 1999). These are independence of judgement, self-confidence, and attraction to complexity, aesthetic orientation, and tolerance for ambiguity, openness to experience, psychoticism, risk taking, androgyny, perfectionism, persistence, resilience, and self-efficacy. It is easy to find common characteristics among creative people but that doesn't mean that these automatically make a person or a product they make creative.

Timothy Leary took this idea of common characteristics a bit further and suggested there are four types of creative personalities ([25] as cited in [27]).

ref

From his ideas we can draw the conclusion that a creative person needs to be able to make novel combinations from novel ideas.

Reproductive Blocked

(no novel combinations, no direct experience)

Reproductive Creator

(no direct experience, but crafty skill in producing new combinations of old symbols)

Creative Creator

(new experience presented in novel performances)

Creative Blocked

(new direct experience expressed in conventional modes)

Tables 4.2 and 4.3 are in Leary's words.

Reproductive Blocked	Reproductive Creator	Creative Creator	Creative Blocked
The routine, well-socialised person who experiences only in terms of what he has been taught and who produces only what has been produced before.	The innovating performer who experiences only in terms of the available categories but has learned to manipulate these categories in novel combinations.	The person who experiences directly outside the limits of ego and labels, and who has learned to develop new models of communications, or who can manipulate familiar categories in novel combinations or who can let natural modes develop under his nurture.	The person who experiences uniquely and sensitively outside of game concepts (either by choice or helplessly by inability) but who is unable to communicate or uninterested in communicating these experiences outside the conventional manner.
Reproductive Performer	Creative Performer		Reproductive Performer
Reproductive Experience		Creative Experience	

Table 4.2: Leary's four types of creativity

4.2 In Computers

In this section I am summarising a few models that try to implement creative thinking models in computers. It is really just a survey of different concepts and

Reproductive Blocked	Reproductive Creator	Creative Creator	Creative Blocked
Unimaginative, incompetent hack.	Reliable nihilist, insensitive, unsuccessful innovator whose shock value changes to morbid curiosity as fads of performance change.	The mad creative genius, the undiscovered far-out crackpot creator who is recognised by later generations as a creative giant.	Psychotic, religious crank, eccentric who uses conventional forms for expressing mystical convictions.
Competent, responsible, reliable worker.	Bold initiator who wins game recognitions but whose fame crumbles as fads of performance change.	The truly creative giant recognised by his own age and the ages to come.	Solid, reliable person with a “deep streak”.
Reproductive Performer	Creative Performer		Reproductive Performer
Reproductive Experience		Creative Experience	

Table 4.3: Leary's social labels to describe the types of creativity

views and does not immediately apply to my specific research on creative search tools unfortunately.

4.2.1 Bipin Indurkhyā

Indurkhyā argues that there are two main cognitive mechanisms of creativity: namely juxtaposition of dissimilar and deconceptualization. He says that we are constrained by associations of our concept networks that we inherit and learn in our lifetime, but that computers do not have those conceptual associations and have therefore an advantage when it comes to creative thinking ([Indurkhyā 1997](#)). He suggests a computer model using two layers that interact with each other: a perceptual and a conceptual layer.

- Juxtaposition of dissimilar
- Deconceptualization

4.2.2 Partridge and Rowe

Partridge and Rowe have written a good survey of computational models of creativity in their book “Computers and Creativity” (Partridge and Rowe 1994) although it is now probably quite out of date (the book was published in 1994). They mention the computer as an unbiased medium for executing creative programs (Partridge and Rowe 1994, p.26). Some of the computational methodologies they discuss are as follows, many taken from classical artificial intelligence research.

- Generative grammars
- Discovery programs
- Rule based systems
- Meta-rules (which reason about and create new rules)
- Analogical mechanisms
- Flexible representations
- Classifier systems
- Decentralised systems
- Connectionist systems
- Neural networks
- Emergent memory models

Classifier systems for example, consist of a set of rules and a message list.

1. Place input messages on current message list
2. Find all rules that can match messages
3. Each such rule generates a message for the new message list
4. Replace current message list with the new one
5. Process new list for any system output
6. Return to step 1

These can easily be combined with genetic algorithms to enable the system to learn an appropriate classifier set. This is called emergent behavior. Another approach is connectionism a.k.a. neural networks. They then go on to describe their emergent-memory model. They are applying the ideas of Poincare and Wallas and are heavily influence by Minsky’s theory of K-lines (Minsky 1980; Minsky 1988). They define the following characteristics for creative programs:

- flexible knowledge representation scheme
- representational imprecision

- multiple representations
- self-assessment
- full elaboration

4.2.3 David Gelernter

Gelernter introduces a theory of how the human mind works in ([Gelernter 1994](#)). His “spectrum model” is based on the idea of mental focus and relates well to creativity. According to him we have a thought spectrum. The higher the mental focus, the more awake we are, the more adult we are and modern, logical and rational, convergent, abstract and detailed. The less focused we are the younger or ancient or dreaming we are. Low focus thoughts are metaphoric, hallucinations, divergent, creative, inspirations, concrete, ambient and emotional. Emotions glue low focus thoughts together.

He gives a good example of his own computer program that is being trained by a set of simple pairs (or memories) in the form -mood: happy- for example. These sets of pairs form the experience of the system, the memory that the system can access. It's fetching all memory pairs that match a certain probe, then generalizes them and picks out a feature that is common to all and then uses that to probe further if necessary.

He models his spectrum concept in a way that if we want the system to operate at low focus, more memory pairs would be fetched and more generalised features are deducted and so on. He describes his FGP program (Fetch Generalise Project) as follows ([Gelernter 1994](#), p.132).

1. Fetch memory pairs in response to a probe (question)
2. Sandwich them together and peer through the bundle at once
3. Notice the common features that emerge strongly (generalise)
4. Pick out interesting emergent details and probe further if necessary

With low focus the system would not generalise as much and just pick out a particular memory, etc. The computer system he has built seems very limited. His memory pairs cannot describe everything. For example they can describe states but not actions.

This idea of accessing thoughts/memories is very closely related to searching. Searching an index in a search engine is similar to remembering, trying to find all memories related to the current thought for example.

4.2.4 Marvin Minsky

Minsky introduces the concepts of k-lines in his Society of Mind ([Minsky 1980](#); [Minsky 1988](#)). It is basically a theory of memory. He claims that the “function of a memory is to recreate a state of mind”. His theory of k-lines is as follows.

When you get an idea, or solve a problem, or have a memorable experience, you create what we shall call a K-line. This K-line gets connected to those mental agencies that were actively involved in the memorable mental event. When that K-line is later activated, it reactivates some of those mental agencies, creating a partial mental state resembling the original. ([Minsky 1980](#); [Minsky 1988](#))

This theory works quite well with Gelernter’s idea of memory. K-lines in this sense are nothing other than Gelernter’s memory pairs.

He and his student Push Singh have formalised the idea of a panalogy, which could be relevant for my project. The idea is that an idea can and should be conceptualised in many different ways. This could be seen as a fall-back mechanism for computational models, if one approach didn’t return the desired/expected results.

4.2.5 Matthew Elton

Elton explains the concept of “Artificial Creativity” which can be seen as a sub-area of [Artificial Intelligence \(AI\)](#). AI research isn’t human enough, he argues, it needs to include less abstract ideas like emotions, morals, aesthetic sensibility and creativity. He goes on to explain in detail how production, evaluation and etiology play a role in everything ([Elton 1995](#)).

Opposed to the traditional approach of AI to study some aspect of the human brain in a specific domain only, he argues that in order to understand creativity we need to look at more than that. Creativity arises from a process that is not isolated. The etiology (its history) is essential for something to be classed as creative. Generation (of artefacts or ideas) cannot count as creative if it doesn’t undergo evaluation in the process. In order to evaluate we need a sound knowledge of the relevant domain. “We want creative evaluation to be influenced by a longstanding history of interaction with entities (of whatever kind) in the world.” Computer systems can be seen in two perspectives: plastic and implastic (resettable). “All systems can be seen from the implastic perspective since ultimately all systems are built out of physical components that are (statically) well behaved, but for certain explanatory purposes some are best understood plastic-

ally.” Connectionist networks are an example of a plastic system. The brain is a plastic system too.

How do we get enough cultural information and background into the machine to train it? “There is no pure science of creativity, because it is paradigmatically idiographic — it can only be understood against the backdrop of a particular history.”

His comments on evaluation are inspirational. How do I make my system evaluate its results or productions (as opposed to me testing my system)?

4.3 In Academia

Two transdisciplinary fields of study have emerged from the variety of disciplines concerned. These are computational creativity and creative computing. The former lies at the cross section of artificial intelligence and cognitive science and the latter is mostly distinguished by its involvement in art. Creative computing focuses on the process of creativity rather than just the outcome as in computational creativity.

Summary

- Boden: Combinational, exploratory and transformative ([Boden 2003; Wiggins 2006](#)) (process)
- Boden: new, surprising, valuable ([Boden 2003](#)) (product)
- Colton: Skill + appreciation + imagination = creativity (or the appearance of) ([Colton 2008b](#)) (product+process)
- Wiggins: relevance + acceptability + quality ([Wiggins 2006](#)) (product)
- Ritchie: typicality + quality ([Ritchie 2001; Ritchie 2007](#)) (product)
- Pease: novelty + value ([Pease, Winterstein and Colton 2001](#)) (product+process)
- Ventura: efficiency + variety ([Ventura 2008](#)) (product+process)
- Jordanous: value (related concepts: usefulness, appropriateness, relevance) + novelty (related concepts: originality, newness) ([Jordanous and Keller 2012](#))

references

The concept of creative computing has existed for some time but has not yet managed to evolve into a recognised discipline within computer science. Computational creativity, on the other hand, has emerged as a field within artificial intelligence research¹ and overlaps with creative computing ideas to some extent.

¹<http://www.computationalcreativity.net/iccc2013/>

It is important to differentiate between the terms creative computing and computational creativity. Intuitively the former is about doing computations in a creative way, while the latter is about achieving creativity through computation. You can think of the latter falling into the artificial intelligence category (using formal computational methods to mimic creativity as a human trait, see also²) and the former being a more poetic endeavour of how the computing itself is done, no matter what the actual purpose of the program is.

As a good example of creative computing, consider the International Obfuscated C Code Contest³. The competition revolves around writing compilable/runnable code, while visually appearing as obfuscated as possible. They value unusuality, obscurity and creativity but expect contestants to follow the strict rules and constraints of the C programming language.

Examples of computational creativity are Simon Colton's Painting Fool⁴ or Harold Cohen's AARON⁵; both are computer programs that paint pictures. Kurzweil's Cybernetic Poet⁶ is a classic example of a program that produces poetry.

But how may we apply the insights into creativity described above in computing? One approach is described by Simon Colton ([Colton 2008a](#)), who suggests we should adopt human skill, appreciation and imagination.

Without skill, they would never produce anything. Without appreciation, they would produce things which looked awful. Without imagination, everything they produced would look the same. ([Colton 2008a](#))

He thinks that evaluating the worth of an idea or product is the biggest challenge facing computational creativity. Whereas in conventional problem solving success is defined as finding a solution, in a creative context more aesthetic considerations have to be taken into account. He suggests three ways for computer programs to generate creative artefacts:

1. Mimicking human skill
2. Mimicking human appreciation
3. Mimicking human imagination

²<http://www.computationalcreativity.net/iccc2013/>

³<http://www.ioccc.org/>

⁴<http://www.thepaintingfool.com/>

⁵<http://www.kurzweilcyberart.com/aaron/history.html>

⁶http://www.kurzweilcyberart.com/poetry/rkcp_overview.php

4.3.1 Computational Creativity

Computational creativity is a relatively new discipline and as such not well defined. Simon Colton, the creator of the Painting Fool, describes it as the discipline of generating artefacts of real value to someone (Colton 2008a). This is in contrast to classic artificial intelligence problem solving. He identifies that evaluating the worth of such an artefact as the biggest problem of computational creativity. In problem solving, success is when a solution to the problem has been found. In artefact generation a more aesthetic consideration has to be taken into account.

One could say that computational creativity is the attempt at giving computers the skills, appreciation and imagination needed to produce creative artefacts. Whether or not this makes the computer creative or the programmer is another question that I will not try to answer here.

Computational creativity has emerged from within AI research. Simon Colton and Geraint Wiggins argue AI falls within a problem solving paradigm: “an intelligent task, that we desire to automate, is formulated as a particular type of problem to be solved ” (Colton and Wiggins 2012, p.2), whereas “in Computational Creativity research, we prefer to work within an artefact generation paradigm, where the automation of an intelligent task is seen as an opportunity to produce something of cultural value. ” (Colton and Wiggins 2012, p.2, my emphasis)

The International Association for Computational Creativity (ACC) promotes the advancement of computational creativity which is defined as follows.

Computational Creativity is the art, science, philosophy and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative. ([International Conference on Computational Creativity \(ICCC\)14 website](#))

Computational creativity is multidisciplinary, bringing together researchers from artificial intelligence, cognitive psychology, philosophy, and the arts. Its role within computer science falls under the scientific paradigm (Hugill 2013, p.8), (see also A. H. Eden 2007), as opposed to CC in the technocratic paradigm. Its main goal is to model, simulate or replicate human creativity using a computer and it has the following three aims:

- to construct a program or computer capable of human-level creativity

- to better understand human creativity and to formulate an algorithmic perspective on creative behavior in humans
- to design programs that can enhance human creativity without necessarily being creative themselves

The ACC manages the annual [ICCC](#), whose recent call for papers (for [ICCC 2014](#)) gives a useful insight into their research agenda. It can be broken down as follows:

- Paradigms, metrics, frameworks, formalisms, methodologies, perspectives
- Computational creativity-support tools
- Creativity-oriented computing in education
- Domain-specific vs. generalised creativity
- Process vs. product
- Domain advancement vs. creativity advancement
- Black box vs. accountable systems

Simon Colton and Geraint Wiggins have also identified several directions for future research in the field: ([Colton and Wiggins 2012](#), p.5)

1. Continued integration of systems to increase their creative potential.
2. Usage of web resources as source material and conceptual inspiration for creative acts by computer.
3. Using crowd sourcing and collaborative creative technologies bringing together evaluation methodologies based on product, process, intentionality and the framing of creative acts by software.

This reminds of the 4 P's, and CC and Digital Humanities (DH) models

- Domain-specific vs. generalised creativity
- Process vs. product
- Domain advancement vs. creativity advancement
- Black box vs. accountable systems

4.3.2 Creative Computing

rewrite and format

In the recent first issue of the [International Journal of Creative Computing \(IJCrC\)](#) Hugill and Yang introduced [CC](#) as a new discipline ([Hugill and Yang 2013](#)) with an overarching theme of “unite and conquer” ([Yang 2013](#), p.1, his emphasis). Its broad aim is to “reconcile the objective precision of computer systems (mathesis) with the subjective ambiguity of human creativity (aesthesia).” ([Hugill and Yang 2013](#), p.5). Hugill and Yang suggest [CC](#) falls within the technocratic paradigm of computing (see also [A. H. Eden 2007](#), p.8), i.e. the discipline is closest related to software engineering, rather than mathematics or natural sciences. They identify five main topics for [CC](#) research ([Hugill and Yang 2013](#), p.15-17):

Challenges

transdisciplinarity, cross-compatibility, continuity and adaptivity

Types

creative development of a product, development of a [CC](#) product and development of tool for creativity support

Mechanisms

Boden’s combinational, exploratory and transformational creativity

Methods

development of suitable transdisciplinary [CC](#) research methodologies

Standards

resist standardisation, novel, continuous user interaction, creative mechanisms

The main challenge is for technology to become “more adaptive, smarter and better engineered to cope with frequent changes of direction, inconsistencies, irrelevancies, messiness and all the other vagaries that characterise the creative process” ([Hugill and Yang 2013](#), p.5). In part, these issues are due to the transdisciplinary nature of the field and factors such as common semantics, standards, requirements and expectations are typical challenges. Hugill and Yang therefore argue that creative software should be flexible and able to adapt to ever changing requirements, it should be evaluated and re-written continuously and it should be cross-compatible.

The different **types** of [CC](#) highlight the different aspects researcher and practitioners focus on during their work. These are

Process

creative development of a computing product,

Product

development of a Creative Computing product and

Community

development of computing environment to support creativity.

The creative computing process should consist of combinational, exploratory and transformational activities (in the sense of Margaret Boden's theory, as discussed in

cross ref

).

Broadly speaking, you could say that approaches to CC are therefore either bottom-up (1) or top-down (2).

The third type of CC in a way reflects what Hugill and Yang call the “local and global levels”, which represent the two types of creativity identified by Boden (P- and H-creativity, see above). It is concerned with developing environments, tools and methods and the management of these.

This includes cross-compatibility, which directly represents the solution to the personal/local and historical/global issues mentioned by Boden and Hugill and Yang.

Similar to the four step model of the creative process by Poincaré and Wallas (Poincare 2001; Wallas 1926) and the four step model of problem solving by Pólya (Polya 1957), they propose a four step model for the creative computing process. They do this by comparing the acts of artistic creation and software engineering in some detail. They found that the two processes follow essentially the same levels of abstraction (from the abstract to the concrete). The four steps are (Hugill and Yang 2013, p.15):

1. Motivation (digitised thinking)
2. Ideation (design sketch)
3. Implementation (creative system)
4. Operation (effect of system/revision)

This reminds of the 4 P's, and CC and DH models??

Given the transdisciplinary nature of CC, Hugill and Yang suggest that existing research methodologies are unsuitable and new ones have to be developed. The following is an example of a possible CC research methodology they propose as a starting point (Hugill and Yang 2013, p.17):

1. Review literature across disciplines
2. Identify key creative activities
3. Analyse the processes of creation
4. Propose approaches to support these activities and processes
5. Design and implement software following this approach
6. Experiment with the resulting system and propose framework

Hugill and Yang propose four **standards** for CC ([Hugill and Yang 2013](#), p.17) namely, resist standardisation, perpetual novelty, continuous user interaction and combinational, exploratory and or transformational.

Summary

- Transdisciplinary
- Technocratic paradigm of computer science
- Mathesis + aesthesis
- Local + global
- Top-down + bottom-up
- Continuous life-cycle, cross-compatibility, adaptive software, interoperability

4.3.3 Speculative Computing

SpecLab ([Drucker 2009](#)) is a book by Johanna Drucker about her experiences as a researcher moving between disciplines and the projects she worked on as part of the Digital Humanities laboratory at the University of Virginia, USA. Several of those had pataphysical inspirations.

In his review, on the back cover of the book, John Unsworth says that Drucker “emphasizes the graphical over the textual, the generative over the descriptive, and aesthetic subjectivity over analytical objectivism.” Her main argument is that in the design of digital knowledge representation, subjectivity and aesthetics are an essential feature. She confronts logical computation with aesthetic principles with the idea that design is information.

Aesthesia is the theory of ambiguous and subjective knowledge, ideological and epistemological, while Mathesis is formal objective logic and they contrast each other. Knowledge is always interpretation and subjectivity is always in opposition to objectivity. Knowledge becomes synonymous with information and as such can be represented digitally as data and metadata.

Arguably, few other textual forms will have greater impact on the way we read, receive, search, access, use and engage with the

primary materials of humanities studies than the metadata structures that organize and present that knowledge in digital form. (Drucker 2009, p.9)

But how is this metadata analysed? How do we analyse this type of structured data? And most important of all she asks, what can be considered as data, what can be expressed in those quantitative terms or other standard parameters? Is data neutral, raw or does it have meaning? Here she also points out that many information structures have graphical analogies and can be understood as diagrams that organize the relations of elements within the whole.

Because “computational methods rooted in formal logic tend to be granted more authority [...] than methods grounded in subjective judgement”, she introduces the discipline of Speculative Computing as the solution to that problem. The concept can be understood as a criticism of mechanistic, logical approaches that distinguish between subject and object.

Speculative computing takes seriously the destabilization of all categories of entity, identity, object, subject, interactivity, process, or instrument. In short, it rejects mechanistic, instrumental, and formally logical approaches, replacing them with concepts of autopoiesis (contingent interdependency), quantum poetics and emergent systems, heteroglossia, indeterminacy and potentiality, intersubjectivity, and deformance. Digital Humanities is focused on texts, images, meanings, and means. Speculative Computing engages with interpretation and aesthetic provocation. (Drucker 2009, p.29)

Pataphysics governs exceptions and anomalies and she introduces a, what she calls, “patacritical” method of including those exceptions as rules — even if repeatability and reliability are compromised. Bugs and Glitches are privileged over functionality, and although that may not be as useful in all circumstances, they are “valuable to speculation in a substantive, not trivial, sense.” In an essay on speculative computing (Drucker and Nowviskie 2007) she says “Pataphysics celebrates the idiosyncratic and particular within the world of phenomena, thus providing a framework for an aesthetics of specificity within generative practice.” To break out of the formal logic and defined parameters of computer science we need speculative capabilities and Pataphysics. “The goal of pataphysical and speculative computing is to keep digital humanities from falling into mere technical application of standard practices.”

“Pataphysics inverts the scientific method, proceeding from and

sustaining exceptions and unique cases, while quantum methods insist on conditions of indeterminacy as that which is intervened in any interpretative act. Dynamic and productive with respect to the subject-object dialectic of perception and cognition, the quantum extensions of speculative aesthetics have implications for applied and theoretical dimensions of computational humanities.” (Drucker and Nowviskie 2007)

With this, Drucker introduces Speculative Aesthetics, which links interface design in which other speculative computing principles. She also refers to Kant and his idea of “purposiveness without purpose.” She says that the appreciation of design as it is (outside of utility) is the goal of speculative aesthetics.

4.3.4 Digital Humanities

Anne Burdick, Johanna Drucker, Peter Lunefeld, Todd Presner and Jeffrey Schnapp (referred to as ‘the authors’ in this section) have collaboratively written an authoritative manifesto for the field of DH (Burdick et al. 2012). Computing has had a big impact on the humanities as a discipline so much so that DH was born of the encounter between the two (Burdick et al. 2012, p.3). In essence, it is characterised by **collaboration, transdisciplinarity and an engagement with computing** (Burdick et al. 2012, p.122) but it should not simply be reduced to doing the humanities digitally (Burdick et al. 2012, p.101). It spans across many traditional areas of research, such as literature, philosophy, history, art, music, design and of course computer science.

Transliteracy⁷ therefore is fundamental (Thomas et al. 2007);

“The field of Digital Humanities may see the emergence of polymaths who can ‘do it all’: who can research, write, shoot, edit, code, model, design, network, and dialogue with users. (Burdick et al. 2012, p.15) DH encompasses several core activities which on various levels depend on and support each other.

Design

Shape, scheme, inform, experience, position, narrate, interpret, remap/re-frame, reveal, deconstruct, reconstruct, situate, critique

Curation, analysis, editing, modelling

Digitise, classify, describe, metadata, organise, navigate

⁷Sue Thomas et al. define transliteracy as “the ability to read, write and interact across a range of platforms, tools and media from signing and orality through handwriting, print, TV, radio and film, to digital social networks.” (Thomas et al. 2007)

Computation, processing

Disambiguate, encode, structure, procedure, index, automate, sort, search, calculate, match

Networks, infrastructure

Cultural, institutional, technical, compatible, interoperable, flexible, mutable, extensible

Versioning, prototyping, failures

Iterate, experiment, take-risks, redefine, beta-test

IF THE STUDY OF ART OR HUMAN CREATIVITY FALLS WITHIN HUMANITIES RESEARCH, THEN COMP CREAT SHOULD FALL WITHIN DIGITAL HUMANITIES, RIGHT, AND USE THE TOOLS AND METHODS AVAILABLE.

DESIGN

The authors suggest that “for digital humanists, design is a creative practice harnessing cultural, social, economic, and technological constraints in order to bring systems and objects into the world.” ([Burdick et al. 2012](#), p.13)

In generative mode, these designers shape structural logics, rhetorical schemata, information hierarchies, experiential qualities, cultural positioning, and narrative strategies. When working analytically, their task is to visually interpret, remap or reframe, reveal patterns, deconstruct, reconstruct, situate, and critique. ([Burdick et al. 2012](#), p.12)

CURATION, ANALYSIS, EDITING, MODELING

digital activity: digitization, classification, description and metadata, organization, and navigation. ([Burdick et al. 2012](#), p.17)

Involving archives, collections, repositories, and other aggregations of materials, CURATION is the selection and organization of materials in an interpretive framework, argument, or exhibit. ([Burdick et al. 2012](#), p.17)

The parsing of the cultural record in terms of questions of authenticity, origin, transmission, or production is one of the foundation stones of humanistic scholarship upon which all other interpretive work depends. But editing is also productive and generative, and it is the suite of rhetorical devices that make a work. Editing is the

creative, imaginative activity of making, and as such, design can be also seen as a kind of editing (Burdick et al. 2012, p.18)

MODELING highlights the notion of content models—shapes of argument expressed in information structures and their design. (Burdick et al. 2012, p.18)

COMPUTATION, PROCESSING

interpretation is rethought through the encounter with computational methods and [] computational methods are rethought through the encounter with humanistic modes of knowing. (Burdick et al. 2012, p.103)

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)

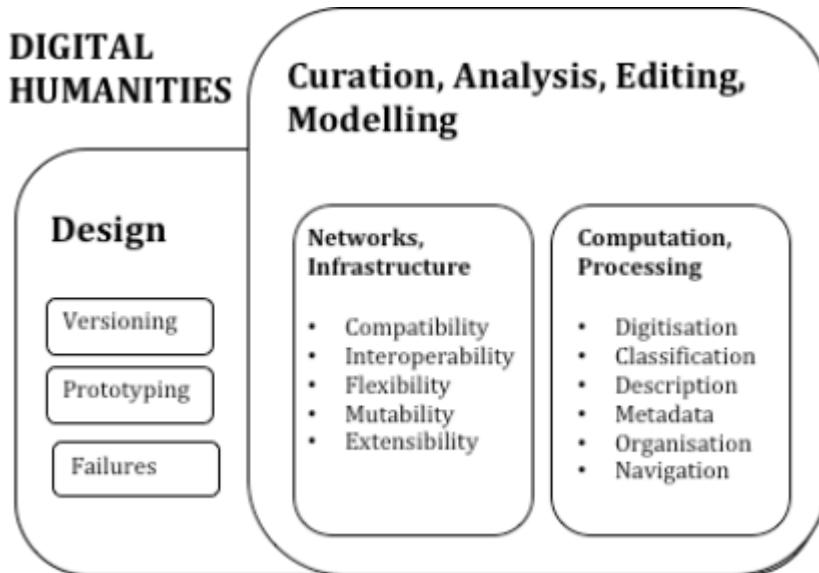


Figure 4.4: Digital Humanities model

NETWORKS, INFRASTRUCTURE

Designing and building digital projects depend on knowledge of these fundamentals and on a nuanced understanding of the networked environments in which the projects will develop and variously reside. (Burdick et al. 2012, p.17)

Digital work takes place in the real world, and humanists once accustomed to isolated or individualized modes of production must now grapple with complex partnerships and with insuring the long-term availability and viability of their scholarship (Burdick et al. 2012, p.21)

VERSIONING, PROTOTYPING, FAILURES

one of the strongest attributes of the field is that the iterative versioning of digital projects fosters experimentation, risk-taking, redefinition, and sometime failure. (Burdick et al. 2012, p.21)

SOUNDS LIKE SOFTWARE ENGINEERING

It is important that we do not short-circuit this experimental process in the rush to normalize practices, standardize methodologies, and define evaluative metrics. (Burdick et al. 2012, p.21)

argument for creative computing too

Field map of digital humanities: emerging methods and genres

(Burdick et al. 2012, p.29-60)

- enhanced critical curation
 - digital collections
 - multimedia critical editions
 - object-based argumentation
 - expanded publication
 - experiential and spatial
 - mixed physical and digital
- augmented editions and fluid textuality

- o structured mark-up
- o natural language processing
- o relational rhetoric
- o textual analysis
- o variants and versions
- o mutability
- scale: the law of large numbers
- o quantitative analysis
- o text-mining
- o machine reading
- o digital cultural record
- o algorithmic analysis
- distant/close, macro/micro, surface/depth
- o large-scale patterns
- o fine-grained analysis
- o close reading
- o distant reading
- o differential geographies
- cultural analytics, aggregation, and data-mining
- o parametrics
- o cultural mash-ups
- o computational processing
- o composite analysis
- o algorithm design
- visualization and data design
- o data visualization
- o mapping
- o information design
- o simulation environments
- o spatial argument
- o modelling knowledge
- o visual interpretation
- locative investigation and thick mapping
- o spatial humanities
- o digital cultural mapping
- o interconnected sites
- o experimental navigation
- o geographic information systems (GIS)
- o stacked data
- the animated archive
- o user communities

- o permeable walls
- o active engagement
- o bottom-up curation
- o multiplied access
- o participatory content creation
- distributed knowledge production and performative access
- o global networks
- o ambient data
- o collaborative authorship
- o interdisciplinary teams
- o use as performance
- o crowd-sourcing
- humanities gaming
- o user engagement
- o rule-based play
- o rich interaction
- o virtual learning environments
- o immersion and simulation
- o narrative complexity
- code, software, and platform studies
- o narrative structures
- o code as text
- o computational processes
- o software in a cultural context
- o encoding practices
- database documentaries
- o variable experience
- o user-activated
- o multimedia prose
- o modular and combinatoric
- o multilinear
- repurposable content and remix culture
- o participatory Web
- o read/write/rewrite
- o platform migration
- o sampling and collage
- o meta-medium
- o inter-textuality
- pervasive infrastructure
- o extensible frameworks
- o heterogeneous data streams

- polymorphous browsing
- cloud computing
- ubiquitous scholarship
- augmented reality
- web of things
- pervasive surveillance and tracking
- ubiquitous computing
- deterritorialization of humanistic practice

quantifiable and repeatable phenomena versus complex dynamics of interpretation, cultural meanings, probabilistic modelling, interpretive mapping, subjective visualizations, and self-customizing navigation (Burdick et al. 2012, p.103)

TOOLS

Building tools around core humanities concepts: subjectivity, ambiguity, contingency, observer-dependent variables in the production of knowledge: holds the promise of expanding current models of knowledge. As such, the next generation of digital experimenters could contribute to humanities theory by forging tools that quite literally embody humanities centred views regarding the world. (Burdick et al. 2012, p.104)

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)

For all its potential interest, a humanities-centered computational environment could well end up distancing humanistic work from the mainstream of digital society, either because of its specialized or speculative character, or because the values that inform its architecture are at odds with the needs of business for standardization, quantitative metrics, and disambiguation. (Burdick et al. 2012, p.105)

Summary

- Collaborative, Transdisciplinary and Computing

4.3.5 Computer Ethics

ETHICS: PROCESS< PRODUCT< PURPOSE

ROBOT ETHICS: similar to 4-p's of creativity (McBride 2013)

it has three actors: Robot engineer, client and user.

4 approaches:

- challenge the myth of autonomy
- Developing practice-based approaches (in context of its purpose and environment)
- Managing ethical variety
- A model for human centred robot ethics

Virtuous robot:

- Human-centred
- Man-machine interdependency
- Practice based (context)
- Ethical variety

Chapter 5

Technology

5.1	Information Retrieval	70
5.1.1	IR Models	72
5.1.2	Ranking	75
5.1.3	Query Expansion and Relevance Feedback	78
5.2	Natural Language Processing	80
5.2.1	Damerau-Levenshtein	80
5.3	Linguistics / WordNet	91
5.4	Algorithm Formalisation	92

Parts of this chapter have been published in peer-reviewed...

Knowledge needed to understand project:

- Search engines
- index
- corpus
- query — expansion etc
- results
- searching vs browsing
- Web programming
-

update all graphics with inkscape

references for this whole chapter?

5.1 Information Retrieval

Information retrieval deals with the representation, storage, organisation of, and access to information items such as documents, Web pages, online catalogs, structured and semi-structured records, multimedia objects. The representation and organisation of the information items should be such as to provide the users with easy access to information of their interest. (Baeza-Yates and Ribeiro-Neto 2011)

In simple terms, a typical search process can be described as follows. A user is looking for some information so she or he types a search term or a question into the text box of a search engine. The system analyses this query and retrieves any matches from the index, which is kept up to date by a Web crawler. A ranking algorithm then decides in what order to return the matching results and displays them for the user. In reality of course this process involves many more steps and level of detail, but it provides a sufficient enough overview. See figure 5.1.

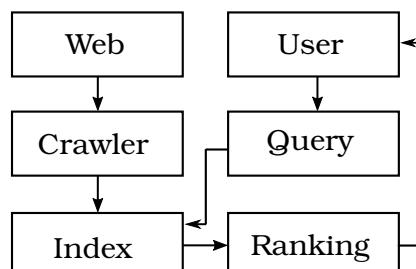


Figure 5.1: Abstract search engine architecture

Most big Web search engines like Google, Baidu or Bing focus on usefulness and relevance of their results.(Google 2012; Baidu 2012; Microsoft 2012) Google uses over 200 signals (Google 2012) that influence the ranking of Web pages including their original PageRank algorithm (Brin and Page 1998b; Brin and Page 1998a).

Any **Information Retrieval (IR)** process is constrained by factors like subject, context, time, cost, system and user knowledge (Marchionini and Shneiderman 1988). Such constraints should be taken into consideration in the development of any search tool. A Web crawler needs resources to crawl around the Web, language barriers may exist, the body of knowledge might not be suitable for

all queries, the system might not be able to cater for all types of queries (e.g. multi-word queries), or the user might not be able to understand the user interface, and many more. It is therefore imperative to eliminate certain constraining factors (for example by choosing a specific target audience or filtering the amount of information gathered by a crawler from Web pages).

Crawler The crawler, sometimes called spider, indexer or bot, is a program that processes and archives information about every available webpage it can find. It does this by looking at given ‘seed’ pages and searching them for hyperlinks. It then follows all of these links and repeats the process over and over. The Googlebot¹ and the Bingbot² are well-known examples.

Index An index is a list of keywords (called the dictionary or vocabulary) together with a list (called postings list) that indicates the documents in which the terms occurs. One way to practically implement this is to create a **Term-Document Matrix (TDM)**. In this case $f_{i,j}$ is the frequency of term k_i in document d_j .

$$\begin{matrix} & d_1 & d_2 \\ \begin{matrix} k_1 \\ k_2 \\ k_3 \end{matrix} & \begin{bmatrix} f_{1,1} & f_{1,2} \\ f_{2,1} & f_{2,2} \\ f_{3,1} & f_{3,2} \end{bmatrix} \end{matrix} \quad (5.1)$$

example TDM for faustroll sentence?

Total wordcount of files: Faustroll=131891, Gospel=139669, Voyage=497295.

cross references with hyperlink hypertarget

The dictionary is usually **preprocessed** to eliminate punctuation and stop-words (e.g. I, a, and, be, by, for, the, on, etc.) that would be useless in everyday text search engines. For specific domains it even makes sense to build a ‘controlled vocabulary’ which can be seen as a domain specific taxonomy and are very useful for query expansion.

Ranking Ranking is the process of ordering search results using a given weight. One simple method of ranking is the so-called **Term Frequency-Inverse Document Frequency** or **TF-IDF** for short. Given a **Term Frequency (TF)** weight of $tf_{i,j}$ and a **Inverse Document Frequency (IDF)** weight of idf_j it is defined as $tf_{i,j} \times idf_j$.

¹Googlebot (<https://support.google.com/webmasters/answer/182072>)

²Bingbot (<http://www.bing.com/webmaster/help/which-crawlers-does-bing-use-8c184ec0>)

$$w_{i,j} = \begin{cases} (1 + \log f_{i,j}) \times \log \frac{N}{df_i} & \text{if } f_{i,j} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (5.2)$$

Where $w_{i,j}$ is the weight associated with (k_i, d_j) . Using this formula ensures that rare terms have a higher weight and more so if they occur a lot in one document.

The **TF** $tf_{i,j}$ is calculated and normalised using a log function as: $1 + \log f_{i,j}$ if $f_{i,j} > 0$ or 0 otherwise.

The total **TF** F_i is calculated as $\sum_{j=1}^N f_{i,j}$, where F_i is the total frequency of term k_i in the collection and $f_{i,j}$ is the frequency of occurrence of term k_i in document d_j and N is the total number of documents.

The **IDF** idf_j weight is calculated as $\log \frac{N}{df_i}$, where the document frequency df_i is the number of documents in a collection that contain a term k_i and idf_i is the **IDF** of term k_i . The more often a term occurs in different documents the lower the **IDF**.

5.1.1 IR Models

IR models describe ranking algorithms formally. ???

	<i>Faustroll</i>	<i>Gospel</i>	<i>Voyage</i>
<i>Faustroll</i>	77	0	0
<i>father</i>	1	28	2
<i>time</i>	34	16	129
<i>purpose</i>	2	0	3
<i>little</i>	28	16	81
<i>background</i>	0	0	0
<i>water</i>	29	7	120
<i>doctor</i>	30	0	0
<i>without</i>	27	7	117
<i>skiff</i>	35	0	0
<i>bishop</i>	27	0	2
<i>God</i>	25	123	2
<i>substance</i>	8	3	1
<i>issue</i>	0	2	2
<i>watch</i>	5	3	6

Figure 5.2: Various wordcounts in Faustroll, Gospel and Voyage

There are different models for different needs, for example a multimedia system is going to be different than a text based system, or a Web based system is going to be different than an offline database system. Even within one such category there could more than one model. Take text based search systems for example. Text can be unstructured or semi-structured. Web pages are typically semi-structured. They contain a title, different sections or paragraphs and so on. An unstructured page would have no such differentiations but only contain simple text. Classic example models are set theoretic, algebraic and probabilistic. The PageRank algorithm by Google is a link-based retrieval model.

The notation for **IR** models is as follows (adapted from [Baeza-Yates and Ribeiro-Neto 2011](#), p.58):

An **IR** model is a quadruple $[D, Q, F, R(q_i, d_j)]$ where:

- D is the set of documents,
- Q is the set of queries,
- F is the framework e.g. sets, Boolean relations, vectors linear algebra...
- $R(q_i, d_j)$ is the ranking function, where $q_i \in Q$ and $d_j \in D$,
- t is the number of index terms in a document collection,
- $V = \{k_1, \dots, k_t\}$ is the set of all distinct index terms in a document collection (vocabulary).

change font again because of ligatures?

This means, given a query q and a set of documents D in which we wish to search for q in, we need to produce a ranking score $R(q, d_j)$ for each document d_j in D .

decide on which method for highlighting words — italic or apostrophe

The Boolean Model

One such ranking score is the Boolean model. The similarity of document d_j to query q is defined as follows (quoted from [\(Baeza-Yates and Ribeiro-Neto 2011, p.65\)](#))

$$sim(d_j, q) = \begin{cases} 1 & \text{if } \exists c(q) \mid c(q) = c(d_j) \\ 0 & \text{otherwise} \end{cases} \quad (5.3)$$

A ‘conjunctive component’ describes which terms occur in a document and which ones do not. E.g. for vocabulary $V = \{k_1, \dots, k_t\}$, if the terms $[k_1, k_2, k_3]$ oc-

cur in document d_j then the conjunctive component would be $(1, 1, 1)$, or $(1, 0, 0)$ if only term k_1 appears in d_j .

- | | |
|--------|--|
| $c(d)$ | is the term conjunctive component for document d |
| $c(q)$ | is the term conjunctive component for query q |

Sometimes things are not quite black and white though and we need to weigh the importance of words somehow. The easiest way to do that is by looking at the frequency in which a word occurs.

The Vector Model

The vector model allows a more flexible scoring since it basically computes the various degrees of similarity between documents (taken from ([Baeza-Yates and Ribeiro-Neto 2011, p.78](#))).

$$\begin{aligned}\vec{d}_j &= (w_{1,j}, w_{2,j}, \dots, w_{t,j}) \\ \vec{q} &= (w_{1,q}, w_{2,q}, \dots, w_{t,q})\end{aligned}\tag{5.4}$$

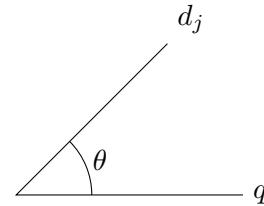


Figure 5.3: The Vector Model

Where t is the total number of terms in the index and $w_{i,j}$ is the TF-IDF weight for each component of the vector. The similarity between the document and the query vector is the cosine of θ .

$$\begin{aligned}sim(d_j, q) &= \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} \\ &= \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}}\end{aligned}\tag{5.5}$$

Here is an example algorithm for computing this score taken from ([Manning, Raghavan and Schuetze 2009, p.125](#)).

```

1   CosineScore (q)
2       float Scores[N] = 0
3       for each d
4           do Initialise Length[d] to the length of document d
5           for each query term t
6               do calculate wtq and fetch postings list for t
7                   for each pair (d, tftd) in postings list
8                       do add wftd to Scores[d]
9           Read the array Length[d]
10          for each d
11              do Divide Scores[d] by Length[d]
12          return Top K components of Scores[]

```

Source 5.1: Pseudo-code for computing vector scores

Where,

q	is the query
N	is the total number of documents
d	is a document
t	is a query term
wt_q	is the weight of the term in the query
tft_d	is the term frequency of t in d
wft_d	is the $tf - idf$ weight of t in d
K	is the number of results we want
$postingslist$	is the list of all (d, tft_d) for a given t .

There are several other common IR models that I won't discuss in detail here. These include the probabilistic, set-based, extended Boolean and fuzzy set (Miyamoto 2010; Miyamoto 1988; Srinivasan 2001; Widjiantoro and Yen 2001; Miyamoto and Nakayama 1986) models or latent semantic indexing (Deerwester et al. 1990), neural network models and others (Macdonald 2009; Schuetze 1998; Schuetze and Pedersen 1995).

Architecture

Search Algorithms

5.1.2 Ranking

Ranking signals contribute to the improvement of the ranking process. These can be content signals or structural signals. Content signals are referring to anything that is concerned with the text and content of a page. This could be simple word counts or the format of text such as headings and font weights. The structural signals are more concerned about the linked structure of pages. They look at incoming and outgoing links on pages. There are also Web usage signals

that can contribute to ranking algorithms such as the clickstream. This also includes things like the Facebook ‘like’ button or the Google+ ‘+1’ button which could be seen as direct user relevance feedback as well.

Ranking algorithms are the essence of any Web search engine and as such guarded with much secrecy. They decide which pages are listed highest in search results and if their ranking criteria were known publically, the potential for abuse (such as Google bombing³ for instance) would be much higher and search results would be less trustworthy. Despite the secrecy there are some algorithms like Google’s PageRank algorithm that have been described and published in academic papers. Here is a survey of the most notable algorithms.

PageRank was developed in 1998 by Larry Page and Sergey Brin as part of their Google search engine and announced in their often cited paper ([Brin and Page 1998a](#)) and they further describe the algorithm here ([Brin and Page 1998b](#)). PageRank is a link analysis algorithm, meaning it looks at the incoming and outgoing links on pages. It assigns a numerical weight to each document, where each link counts as a vote of support in a sense. PageRank is executed at indexing time, so the ranks are stored with each page directly in the index. The following formula for calculating a PageRank PR is taken from ([Baeza-Yates and Ribeiro-Neto 2011](#), p.472).

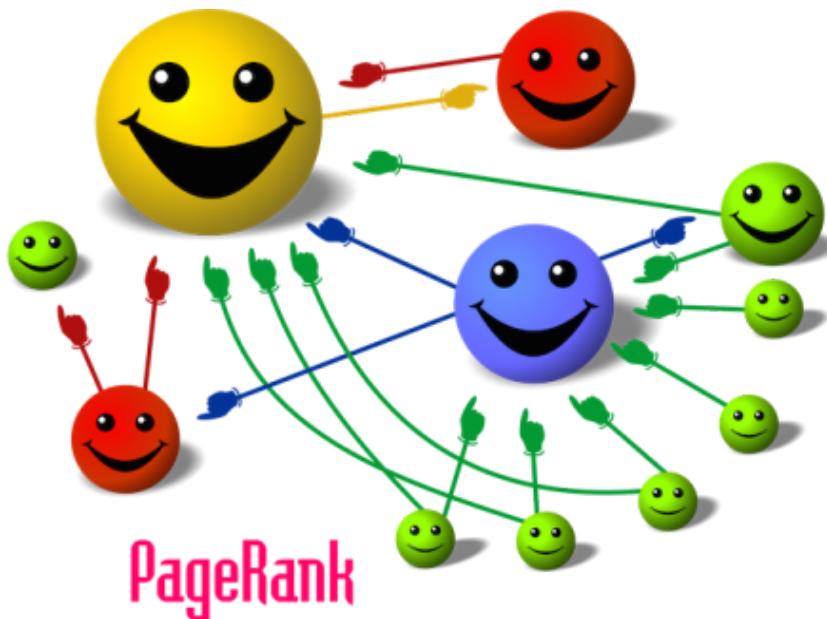


Figure 5.4: PageRank algorithm illustration from Wikipedia

³<http://www.searchenginepeople.com/blog/incredible-google-bombs.html>

$$PR(a) = \frac{q}{T} + (1 - q) \sum_{i=1}^n \frac{PR(p_i)}{L(p_i)} \quad (5.6)$$

Where,

- $L(p)$ is the number of outgoing links of page p ,
- a is the page we want to rank and is pointed to by pages p_1 to p_n ,
- T is the total number of pages on the Web graph, and
- q is the parameter to be set by the system (typically 0.15) needed to deal with dead ends in the graph.

The HITS algorithm also works on the links between pages. It was first described by Kleinberg (Kleinberg 1999; Kleinberg et al. 1999, p.472) in 1999. HITS stands for Hyperlink Induced Topic Search and its basic features are the use of so called hubs and authority pages. It is executed at query time. Pages that have many incoming links are called authorities and page with many outgoing links are called hubs. Again, the following formula is taken from (Baeza-Yates and Ribeiro-Neto 2011, p.471). S is the set of pages.

$$\begin{aligned} H(p) &= \sum_{u \in S | p \rightarrow u} A(u) \\ A(p) &= \sum_{v \in S | v \rightarrow p} H(v) \end{aligned} \quad (5.7)$$

Hilltop is a similar algorithm with the difference that it operates on a specific set of expert pages as a starting point. It was defined by Bharat and Mihaila in 2000 in (Bharat and Mihaila 2000). The expert pages they refer to should have many outgoing links to non-affiliated pages on a specific topic. This set of expert pages needs to be pre-processed at the indexing stage. The authority pages they define must be linked to by one of their expert pages. The main difference to the HITS algorithm then is that their “hub” pages are predefined.

Another algorithm is the so called Fish search algorithm. It was first described by De Bra in 1994 (De Bra and Post 1994a; De Bra and Post 1994b; De Bra, Houben et al. 1994). The basic concept here is that the search starts with the search query and a seed URL as a starting point. A list of pages is then built dynamically in order of relevance following from link to link. Each node in this directed graph is given a priority depending on whether it is judged to be relevant or not. URLs with higher priority are inserted at the front of the list while others are inserted at the back. Special here is that the “ranking” is done dynamically at query time.

There are various algorithms that follow this approach. For example the shark search algorithm ([Hersovici et al. 1998](#)). It improves the process of judging whether or not a given link is relevant or not. It uses a simple vector model with a fuzzy sort of relevance feedback. Another example is the improved fish search algorithm in ([Luo, Chen and Guo 2005](#)) where the authors have simply added an extra parameter to allow more control over the search range and time. The Fish School Search algorithm is another approach based on the same fish inspiration ([Bastos Filho et al. 2008](#)). It uses principles from genetic algorithms and particle swarm optimization. Another genetic approach is Webnaut ([Nick and Themis 2001](#)).

Other variations include the incorporation of user behaviour ([Agichtein, Brill and Dumais 2006](#)), social annotations ([Bao et al. 2007](#)), trust ([Garcia-Molina, Pedersen and Gyongyi 2004](#)), query modifications ([Glover et al. 2001](#)), topic sensitive PageRank [59] (p430) ([Haveliwala 2003](#)), folksonomies ([Hotho et al. 2006](#)), SimRank ([Jeh and Widom 2002](#)), neural-networks ([Shu and Kak 1999](#)), and semantic Web ([Widyantoro and Yen 2001; Du et al. 2007; Ding et al. 2004; Kamps, Kaptein and Koolen 2010; Taye 2009](#)).

5.1.3 Query Expansion and Relevance Feedback

Relevance feedback is an idea of improving the search results by explicit or implicit methods. Explicit feedback asks users to rate results according to their relevance or collects that kind of information through analysis of mouse clicks, eye tracking etc. Implicit feedback occurs when external sources are consulted such as thesauri or by analysis the top results provided by the search engine. There are two ways of using this feedback. It can be displayed as a list of suggested search terms to the user and the user decided whether or not to take the advice, or the query is modified internally without the user's knowledge. This is then called automatic query expansion.

Challenges of Web Search

Other issues that arise when trying to search the World Wide Web are as follows ([\(Baeza-Yates and Ribeiro-Neto 2011, p.449\)](#)).

- Data is distributed. Data is located on different computers all over the world and network traffic is not always reliable.
- Data is volatile. Data is deleted, changed or lost all the time so data is often out-of-date and links broken.
- The amount of data is massive and grows rapidly. Scaling of the search engine is an issue here.

- Data is often unstructured. There is no consistency of data structures.
- Data is of poor quality. There is no editor or censor on the Web. A lot of data is redundant too.
- Data is not heterogeneous. Different data types (text, images, sound, video) and different languages exist.

Since a single query for a popular word can result in millions of retrieved documents from the index, search engines usually adopt a lazy strategy, meaning that they only actually retrieve the first few pages of results and only compute the rest when needed (Baeza-Yates and Ribeiro-Neto 2011, p.459). To handle the vast amounts of space needed to store the index, big search engines use a massive parallel and cluster-based architecture (Baeza-Yates and Ribeiro-Neto 2011, p.459). Google for example uses over 15,000 commodity-class PCs that are distributed over several data centres around the world (Dean, Barroso and Holzle 2003).

Summary

IR refers to the retrieval of information from a collection. In terms of the Internet it is often called Web search. A Web search engine is divided into different components, being the crawler to build an index of the collection and a ranking algorithm which stands between the index and the user.

Different retrieval models exist including the Boolean and the Vector model. Other methods exist to make search results more accurate, including relevance feedback and query expansion.

Search quality is generally measured using the metrics of precision and recall but for Web search precision is more important and usually a metric called “precision at n” is used for measurements.

Challenges are the size of the World Wide Web and ambiguous, unstructured nature of Web pages among others.

Ranking can be done at different stages of the search process. Depending on how the index is formatted and what information can be pre-computed at that stage, the ranking algorithm evaluates every page for relevance and returns them in order. There exist lots of different approaches on ranking, including PageRank and HITS (both analyse the link structure of the WWW), or more dynamic models like Fish search or genetic approaches.

5.2 Natural Language Processing

describe NLTK and the core functionality

Natural Language Tool Kit (NLTK) Python library⁴.

PlaintextCorpusReader

Reader for corpora that consist of plaintext documents. Paragraphs are assumed to be split using blank lines. Sentences and words can be tokenized using the default tokenizers, or by custom tokenizers specified as parameters to the constructor.

Text

A wrapper around a sequence of simple (string) tokens, which is intended to support initial exploration of texts (via the interactive console). Its methods perform a variety of analyses on the text's contexts (e.g., counting, concordancing, collocation discovery), and display the results.

index (word)

Find the index of the first occurrence of the word in the text.

count (word)

Count the number of times this word appears in the text.

5.2.1 Damerau-Levenshtein

Damerau-Levenshtein for clinamen! https://en.wikipedia.org/wiki/Damerau%E2%80%93Levenshtein_distance

The Damerau–Levenshtein distance between two strings a and b is given by $d_{a,b}(|a|, |b|)$ where:

$$d_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0 \\ \min \begin{cases} d_{a,b}(i - 1, j) + 1 \\ d_{a,b}(i, j - 1) + 1 \\ d_{a,b}(i - 1, j - 1) + 1_{a_i \neq b_j} \\ d_{a,b}(i - 2, j - 2) + 1 \end{cases} & \text{if } i, j > 1 \text{ and } a_i = b_{j-1} \text{ and } a_{i-1} = b_j \\ \min \begin{cases} d_{a,b}(i - 1, j) + 1 \\ d_{a,b}(i, j - 1) + 1 \\ d_{a,b}(i - 1, j - 1) + 1_{a_i \neq b_j} \end{cases} & \text{otherwise.} \end{cases} \quad (5.8)$$

⁴<http://www.nltk.org/>

where $1_{(a_i \neq b_j)}$ is the indicator function equal to 0 when $a_i = b_j$ and equal to 1 otherwise.

Each recursive call matches one of the cases covered by the Damerau-Levenshtein distance:

$d_{a,b}(i - 1, j) + 1$ corresponds to a deletion (from a to b).

$d_{a,b}(i, j - 1) + 1$ corresponds to an insertion (from a to b).

$d_{a,b}(i - 1, j - 1) + 1_{(a_i \neq b_j)}$ corresponds to a match or mismatch, depending on whether the respective symbols are the same.

$d_{a,b}(i - 2, j - 2) + 1$ corresponds to a transposition between two successive symbols.

[Natural Language Processing \(NLP\)](#) blah blah blah...

Bird, S., Klein, E. and Loper, E., 2009. [NLP](#) with Python 1st ed., Sebastopol, CA: O'Reilly Media. ([Bird, Klein and Loper 2009](#))

Manning, C., Raghavan, P. and Schuetze, H., 2008. Introduction to Information Retrieval 1st ed., Cambridge: Cambridge University Press. ([Manning, Raghavan and Schuetze 2009](#))

Taken from ([Jurafsky and Martin 2009](#)), also known as:

- Speech and language processing
- Human language technology
- [NLP](#)
- Computational linguistics
- Speech recognition and synthesis

Goals of [NLP](#) are to get computers to perform useful tasks involving human language like:

- Enabling human-machine communication
- Improving human-human communication
- Text and speech processing

e.g. machine translation, automatic speech recognition, natural language understanding, word sense disambiguation, spelling correction, grammar checking...

Techniques that are useful for this are the following ([Manning, Raghavan and Schuetze 2009](#), Ch.2).

Tokenisation

discarding white spaces and punctuation and making every term a token

Normalisation

making sets of words with same meanings, e.g. car and automobile

Case-folding

converting everything to lower case

Stemming

removing word endings, e.g. connection, connecting, connected → connect

Lemmatization

returning dictionary form of a word, e.g. went → go

Regular Expressions

Used to specify text strings in text.

RE search requires a pattern that we want to search for and a corpus of texts to search through.

Errors can be false positives (FP) and false negatives (FN).

- Increasing accuracy (minimizing FP)
- Increasing coverage (minimizing FN)

RE's can be expressed as Finite-State Automata (FSA).

Language Models (LM)

Probabilities are based on counting things. Counting things in natural language is based on a corpus (pl corpora), a computer readable collection of text or speech.

Cats versus cat?

Same lemma but different wordforms.

- A lemma is a set of lexical forms that have the same stem. (e.g. go)
- A wordform is the full inflected or derived form of the word. (e.g. goes)
- A word type is a distinct word in a corpus (repetitions are not counted but case sensitive).
- A word token is any word (repetitions are counted repeatedly)

The process of converting all words in a text to their lemma (e.g. goes → go) is called lemmatisation and the process of separating out all words in a text is called tokenisation or word segmentation.

N-Grams

We can do word prediction with probabilistic models called *N-Grams*. They predict the probability of the next word from the previous $N - 1$ words.

We want to compute the probability for $P(w|h)$ where w is a word and h is a history (the previous words). How many times occurred h followed by w divided by how many times occurred h ?

$$P(w | h) = \frac{\text{count}(hw)}{\text{count}(h)} \quad (5.9)$$

Using the **chain rule of probability**:

$$\begin{aligned} P(w_1^n) &= P(w_1)P(w_2 | w_1)P(w_3 | w_1^2)\dots P(w_n | w_1^{n-1}) \\ &= \prod_{k=1}^n P(w_k | w_1^{k-1}) \end{aligned} \quad (5.10)$$

Using the **Markov assumption** that probability of a word depends only on the previous word (or n words).

$$P(w_1^n) = \prod_{k=1}^n P(w_k | w_{k-1}) \quad (5.11)$$

Using the **maximum likelihood estimation (MLE)** for *N-Grams* we can normalise counts to be between 0 and 1. C stands for count.

Maximum likelihood estimation (MLE)

$$P(w_n | w_{n-N+1}^{n-1}) = \frac{C(w_{n-N+1}^{n-1} w_n)}{C(w_{n-N+1}^{n-1})} \quad (5.12)$$

Usually instead of calculating the counts based on products we calculate them based on sums of logs.

So instead of $p_1 \times p_2 \times p_3 \times p_4 = \log p_1 + \log p_2 + \log p_3 + \log p_4$

Google offers its *N-Gram* data for free on:

- <http://bit.ly/1baDXAW>
- <http://books.google.com/ngrams/>
- <http://www.speech.sri.com/projects/srilm/>
- <http://bit.ly/1G3ZJmX>

Evaluating N-Grams

Extrinsic and intrinsic evaluation.

Extrinsic

: evaluate performance of a language model by embedding it into an independent application.

Intrinsic

: evaluate independent on any application, e.g. perplexity.

Perplexity

$$PP(W) = \sqrt[N]{\prod_{i=1}^N \frac{1}{P(w_i | w_{i-1})}} \quad (5.13)$$

Smoothing

Add-One: Laplace smoothing for bigrams

$$P_{Add-1}(w_i | w_{i-1}) = \frac{c(w_{i-1}, w_i) + 1}{c(w_{i-1}) + V} \quad (5.14)$$

Adjusted count

$$c_i^* = (c_i + 1) \frac{N}{N + V} \quad (5.15)$$

Add-1 smoothing is ok for text categorisation but not so much for language modelling.

Most commonly used is Kneser-Ney extended interpolated.

For very large N-grams like the Web “Stupid Backoff” is used.

Good Turing Discounting

N_c is the frequency of frequency c .

$$c^* = (c + 1) \frac{N_{c+1}}{N_c} \quad (5.16)$$

Naive Bayes

[3] page 234...

(Wikipedia): “A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes’ theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be ‘independent feature model’.”

Maximum Entropy Models (MaxEnt)

Page 227 ... in [1]

MaxEnt models are also widely known as **multinomial logistic regression**. They are used for sequence classification, e.g. part-of-speech tagging. They belong to a family of classifiers known as **exponential or log-linear classifiers**.

The task of classification is to take a single observation, extract some useful features describing the observation, and then, based on these features, to classify the observation into one of a set of discrete classes. A probabilistic classifier also gives the probability of the observation being in that class; it gives a probability distribution over all classes.

MaxEnt works by extracting some set of features from the input, combining them linearly (meaning that each feature is multiplied by a weight and then added up), and then using this sum as an exponent. Formula below shows how to calculate the probability of class c given an observed datum (a given data point) d and λ is a weight that is assigned to feature f . Taking the exponent makes the result always positive. Dividing by the Sum of that for all classes makes it a probability.

$$P(c | d, \lambda) = \frac{\exp \sum_i \lambda_i f_i(c, d)}{\sum_{cl} \exp \sum_i \lambda_i f_i(cl, d)} \quad (5.17)$$

To get the single best class with the highest probability we need to compute the following.

$$\hat{c} = \operatorname{argmax}_{c \in C} P(c | d, \lambda) \quad (5.18)$$

Features:

$$\begin{aligned} f1(c, d) &\equiv [c = \text{LOCATION} \wedge w - 1 = \text{"in"} \wedge \text{isCapitalized}(w)] \\ f2(c, d) &\equiv [c = \text{LOCATION} \wedge \text{hasAccentedLatinChar}(w)] \\ f3(c, d) &\equiv [c = \text{DRUG} \wedge \text{ends}(w, "c")] \end{aligned}$$

PERSON	LOCATION	DRUG
In Québec	In Québec	In Québec
0	1.8 + -0.6	0.3

Table 5.1: MaxEnt Example table

$$P(\text{LOCATION} \mid \text{in Québec}) = \frac{e^{1.8} e^{-0.6}}{e^{1.8} e^{-0.6} + e^{0.3} + e^0} = 0.586$$

$$P(\text{DRUG} \mid \text{in Québec}) = \frac{e^{0.3}}{e^{1.8} e^{-0.6} + e^{0.3} + e^0} = 0.238$$

$$P(\text{PERSON} \mid \text{in Québec}) = \frac{e^0}{e^{1.8} e^{-0.6} + e^{0.3} + e^0} = 0.176$$

The empirical expectation is the sum of all occurrences where a feature is true for one of our observed datums.

$$\text{empirical } E(f_i) = \sum_{(c,d) \in \text{observed}(C,D)} f_i(c, d) \quad (5.19)$$

Evaluation

$$\text{Precision} = \frac{\text{number of correctly labeled}}{\text{total number of extracted}} \quad (5.20)$$

$$\text{Recall} = \frac{\text{number of correctly labeled}}{\text{total number of gold}} \quad (5.21)$$

$$F_1 = \frac{2PR}{P+R} \quad (5.22)$$

Information Extraction

[1] Chapter 22, p 759...

"The process of information extraction (IE), also called text analytics, turns the unstructured information embedded in texts into structured data."

IE involves named entity recognition (NER), relation detection and classification, event detection and classification and temporal analysis.

Named Entity Recognition

A named entity can be anything that can be referred to by a proper name, such as person-, place- or organisation names and times and amounts.

Example (first sentence in Faustroll):

"In this year Eighteen Hundred and Ninety-eight, the Eighth day of February, Pursuant to article 819 of the Code of Civil Procedure and at the request of M. and Mme. Bonhomme (Jacques), proprietors of a house situate at Paris, 100 bis, rue Richer, the aforementioned having address for service at my residence and further at the Town Hall of Q borough."

In this [year Eighteen Hundred and Ninety-eight, the Eighth day of February]^{TIME}, Pursuant to article [819]^{NUMBER} of the [Code of Civil Procedure]^{DOCUMENT} and at the request of [M. and Mme. Bonhomme (Jacques)]^{PERSON}, proprietors of a house situate at [Paris, 100 bis, rue Richer]^{LOCATION}, the aforementioned having address for service at my residence and further at the [Town Hall]^{FACILITY} of [Q borough]^{LOCATION}.

Gazetteers (lists of place or person names for example) can help with the detection of these named entities.

Part of Speech Tagging

Parts of speech (POS) are lexical tags for describing the different elements of a sentence. The eight main parts-of-speech (originating from ca. 100 B.C.) are noun, verb, pronoun, preposition, adverb, conjunction, participle and article. Wikipedia:

Noun

: any abstract or concrete entity; a person (police officer, Michael), place (coastline, London), thing (necktie, television), idea (happiness), or quality (bravery)

Pronoun

: any substitute for a noun or noun phrase

Adjective

: any qualifier of a noun

Verb

: any action (walk), occurrence (happen), or state of being (be)

Adverb

: any qualifier of an adjective, verb, or other adverb

Preposition

: any establisher of relation and syntactic context

Conjunction

: any syntactic connector

Interjection

: any emotional greeting (or “exclamation”)

Building a Large Annotated Corpus of English (Marcus, Santorini and Marciniewicz 1993)

There exist other sets of tags, like the Penn Treebank with divides those 8 tags into a total of 45, for example *CC* for coordinating conjunction, *CD* for cardinal number, *NN* for noun singular, *NNS* for noun plural, *NNP* for proper noun singular, *VB* for verb base form, *VBG* for verb gerund, etc.

The process of adding tags to the words of a text is called parts-of-speech tagging or just tagging. This usually is done together with the tokenisation of the text.

Example (first sentence in Faustroll):

In/IN this/DT [year/NN Eighteen/CD Hundred/CD and/CC Ninety-eight/CD,/ the/DT Eighth/CD day/NN of/IN February/NNP]^{TIME},/
, Pursuant/JJ to/IN article/NN [819/CD]^{NUMBER} of/IN the/DT [Code/NN of/IN Civil/NNP Procedure/NNP]^{DOCUMENT} and/CC at/
IN the/DT request/NN of/IN [M./NN and/CC Mme./NN Bon-homme/NNP (/Jacques/NNP)/)]^{PERSON},/, proprietors/NNS of/
IN a/DT house/NN situate/JJ at/IN [Paris/NNP,/, 100/CD bis/
NN,/, rue/NN Richer/NNP]^{LOCATION},/, the/DT aforementioned/
JJ having/VBG address/NN for/IN service/NN at/IN my/PRP residence/NN and/CC further/JJ at/IN the/DT [Town/NNP Hall/
NNP]^{FACILITY} of/IN [Q/NNP borough/NN]^{LOCATION}./.

$$t_1^n = \underset{t_1^n}{\operatorname{argmax}} P(w_1^n | t_1^n) P(t_1^n) \quad (5.23)$$

$$P(t_i | t_{i-1}) = \frac{C(t_{i-1}, t_i)}{C(t_{i-1})} \quad (5.24)$$

For example: the probability of getting a common noun after a determiner is:

$$P(\text{NN} | \text{DT}) = \frac{C(\text{DT}, \text{NN})}{C(\text{DT})} = \frac{56,509}{116,454} = 0.49 \quad (5.25)$$

Given that there are 116,454 occurrences of DT in the corpus and of these 56,509 occurrences where a NN follows after the DT.

$$P(\text{is} \mid \text{VBZ}) = \frac{C(\text{VBZ, is})}{C(\text{VBZ})} = \frac{10,073}{21,627} = 0.47 \quad (5.26)$$

Or the probability of a third person singular verb being ‘is’ is 0.47.

Parsing

Parsing is the process of analysing a sentence and assigning a structure to it. Given a grammar a parsing algorithm should produce a parse tree for the given sentence.

Grammar

A language is modelled using a grammar, specifically a Context-Free-Grammar or CFG. Such a grammar normally consists of rules and a lexicon. For example a rule could be $\text{NP} \rightarrow \text{Det Noun}$, where NP stands for noun phrase, Det for determiner and Noun for a noun. The corresponding lexicon would then include facts like $\text{Det} \rightarrow \text{a}$, $\text{Det} \rightarrow \text{the}$, $\text{Noun} \rightarrow \text{book}$. This grammar would let us form the noun phrases “the book” and “a book” only. The two parse trees would then look like this:



Figure 5.5: Grammars

The parse tree for the previous example sentence from Faustroll is shown below, in horizontal for convenience.

```

(ROOT
  (S
    (PP (IN In)
      (NP (DT this) (NN year) (NNPS Eighteen) (NNP Hundred)
        (CC and)
        (NNP Ninety-eight)))
      (, ,)% chktex 26
      (NP
        (NP (DT the) (JJ Eighth) (NN day))
        (PP (IN of)
  
```

(NP (NNP February) (, ,) (NNP Pursuant)))% chktex 26
(PP
 (PP (TO to)
 (NP
 (NP (NN article) (CD 819))
 (PP (IN of)
 (NP
 (NP (DT the) (NNP Code))
 (PP (IN of)
 (NP (NNP Civil) (NNP Procedure)))))))
 (CC and)
 (PP (IN at)
 (NP
 (NP (DT the) (NN request))
 (PP (IN of)
 (NP (NNP M.))
 (CC and)
 (NNP Mme) (NNP Bonhomme))))))
 (PRIN (-LRB- -LRB-)
 (NP (NNP Jacques))
 (-RRB- -RRB-))
(, ,)% chktex 26
(NP
 (NP (NNS proprietors))
 (PP (IN of)
 (NP
 (NP (DT a) (NN house) (NN situate))
 (PP (IN at)
 (NP (NNP Paris))))))
(, ,)% chktex 26
 (NP (CD 100) (NN bis))
(, ,)% chktex 26
(VP (VBP rue)
 (NP
 (NP (NNP Richer))
 (, ,)% chktex 26
 (NP (DT the) (JJ aforementioned)
 (UCP
 (S
 (VP (VBG having)
 (NP

```

(NP (NN address))
(PP (IN for)
    (NP (NN service))))
(PP (IN at)
    (NP (PRP$ my) (NN residence))))
(CC and)
(PP
    (ADVP (RBR further))
    (IN at)
    (NP
        (NP (DT the) (NNP Town) (NNP Hall))
        (PP (IN of)
            (NP (NNP Q))))))
    (NN borough)))
(. .))% chktex 26

```

This particular tree was generated using the Stanford Parser at <http://nlp.stanford.edu:8080/parser/index.jsp>. Given the rather complicated nature of the words and sentence structure, some of the labels might be wrong.

5.3 Linguistics / WordNet

Here's my hypernym term. [holonym hypernym](#)

I looked into linguistics for the purpose of patadata. This section definitely needs some expanding. Some concepts that might be relevant include (taken from Wikipedia):

Hyponym

- subcategory of something

Hypernym

- top category of some things

Meronym

- member of something (e.g. finger is meronym to hand, wheel to car)

Holonym

- e.g. tree is holonym of bark, trunk, limb... opposite of meronym

Troponym

- presence of "manner" between things (e.g. to traipse and to mince = walk a certain way)

Homonym

- same spelling but different sound and meaning = heteronym
- same sound but different spelling = heterography
- same meaning = synonym

Antonym

- opposite

Metonym

- figure of speech (e.g. Hollywood for American movies) not quite metaphor but similar.

I need to find REFERENCES for this section.

5.4 Algorithm Formalisation

Algorithm Classification

By implementation:

- Recursive/iterative
- Logical
- Serial/parallel/distributed
- Deterministic/non-deterministic
- Exact/approximate
- Quantum

By field of study:

- Search
- Sorting
- Merge
- Numerical
- Graph
- String
- Computational geometrics
- Combinatorial
- Medical
- Machine learning
- Cryptography
- Data compression
- Parsing

By design paradigm:

- Brute-force/exhaustive search
- Divide and conquer
- Dynamic
- Greedy
- Linear
- Reduction
- Search and enumeration

By complexity:

- Big-O-Notation

High-Level Description

in prose, ignoring implementation details.

Implementation Description

in prose, describing implementation in detail.

Formal description

lowest level, most detailed.

$D = \{d_1, \dots, d_n\}$	is the set of documents
$Q = \{q_1, \dots, q_n\}$	is the set of queries
$q = \{t_1, \dots, t_n\}$	is the set of query terms
$V = \{v_1, \dots, v_t\}$	is the set of all distinct index terms in a document collection (the Vocabulary)
$R(q_i, d_j)$	is the ranking function, where $q_i \in Q$ and $d_j \in D$
N	is the total number of documents
$w_{t,q}$	is the weight of the term in the query
$tf_{t,d}$	is the term frequency of t in d
$wf_{t,d}$	is the tf-idf weight of t in d
P_t	is the postings list of all $(d, tf_{t,d})$ for a given t

Chapter 6

Evaluation

6.1	IR Evaluation	98
6.2	Approaches	99
6.3	Output Minus Input	101
6.4	Measurable Attributes	101
6.5	Linda Candy	103

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

Summary

- output – input
 - novelty + value
 - product + process
 - Mimicking
 - novelty + value + quality
 - randomness + serendipity

reformat

Useless Search Results

The word useless is defined as “not fulfilling or not expected to achieve the intended purpose or desired outcome” in the Oxford dictionary (2010). Given this definition most of the search results we have in mind would be classed as useless. That is at least if we considered every result individually, outside of context and with an information-lookup query in mind. If we have an exploratory search in mind however things get more interesting and we will explain why in the remainder of the paper.

Relevant versus Creative

When we say relevant results we mean the kind of search results that any mainstream search engine would produce, the kind of results you would immediately understand their connection to the query for, the kind of results that just makes sense. Consider the example results in table 1.

Results which might seem useless at first can be much more creative or even poetic. And creative results support exploratory search. Surprise and user expectations play a big role in creativity according to Boden (2003).

Fewer expectations (an open mind) allow creativity to happen more easily. Empirical experiences form expectations, which hinder our ability to accept creative ideas when they happen. In order to be able to recognise creative ideas we need to be able to see what they all have in common and in what way they differ and not reject unusual, unexpected ones.

Relevant Creative

UK action sports lifestyle brand (www.animal.co.uk) List of animals in the Emperor’s possession Wikipedia article (en.wikipedia.org/wiki/animal) Instructions about embalming animals Animal Planet — Discovery Channel (animal.discovery.com) Information on a society for animal training Table 1 — Results for a query for animal. The relevant results were retrieved from Google. The creative results were inspired by Borges’ “Chinese Encyclopaedia” from The Analytical Language of John Wilkins (1942) We can link this very nicely to the idea of exploratory search. Lowering expectations or opening the mind implies extending the task domain or problem space. Creativity and exploratory search seem predestined to work with each other.

Biases

Wikipedia defines bias as “an inclination to present or hold a partial perspective at the expense of (possibly equally valid) alternatives. Anything biased generally is one-sided, and therefore lacks a neutral point of view.” (2012)

However, biases can be good and bad. It is important to consider the implications of their existence though, especially when trying to measure the success of something objectively. An example of when biases can be advantageous is location signals that the search tool takes into account when producing results. An Englishmen would probably not have much use of a Chinese website and vice-versa, even if the actual content matches the original query (unless of course the user happens to understand both languages perfectly). Another example of this is location queries such as ‘Chinese restaurants in Cambridge’, which should return web pages about restaurants based in Cambridge, UK or Massachusetts, USA, depending on the user’s I.P. address. This might seem logical, but in the truest sense it is a bias employed by the search engine to help provide more relevant results to individuals. Truly unbiased search results are probably impossible to come by nowadays.

There is a general move from objectivity to subjectivity in the sense that users become the subject of search results as much as the query they pose. Instead of neutrally providing results for a query alone, the results are tailored around the information known about the user (e.g. language, location, clickstream, social media likes, bookmarks, etc.) to make up the missing context. The user becomes the subject and context of a query, while the results become an objective list of matches for all those values rather than just the query term (s).

Standard Web search: Subject/User Object/Results

Constraints

There are certain factors and constraints that influence the perception and success of the results. Some can be taken into account when building a search system but others cannot be avoided. User education is one way to deal with those issues. Earlier we briefly mentioned some external constraints such as the setting in which the search takes place. Is the user operating from a hand-held device or a desktop computer? Is he or she in a hurry to find answers or just leisurely browsing for them? Is the search system web-based or is the user querying a database?

User Expectations It is important to note that “search systems are not used in isolation from their surrounding context, i.e., they are used by real people who are influenced by environmental and situational constraints such as their current task” (White and Marchionini 2004). User expectations should be taken into consideration during the evaluation of search results. Users who are hoping to find precise answers to a specific question might not be satisfied by exploratory search results. Someone browsing for inspiration on a broad topic on the other

hand could benefit from them. Users should therefore be informed about the nature of the search tool in some way.

User Skill The searching skills of the user matter. Specifically his or her ability to articulate an information need and any knowledge of special search techniques (use of Boolean modifiers, quotation marks, wildcards, etc.) are two important factors that influence the results obtained greatly. This is very much based on the old idea of garbage-in, garbage-out (Lidwell et al. 2010). **Visual Representation** The way that results are presented affects how the user perceives them. A diversity of different document types, for example text, images, sound, or video results could improve how well the results are rated (Sawle et al. 2011). Johanna Drucker had already pointed out that “many information structures have graphical analogies and can be understood as diagrams that organise the relations of elements within the whole” (2009). An alphabetical list is a typical model for representing text data sets for example. But is a ranked list really the best way to represent search results? Other models could be a differently ranked or ordered list, a tree structure, a matrix, a one-to-many relationship, etc.

Structure of Results As suggested by Sawle et al (2011) we need to consider different ways to structure and measure search results. A single, perfectly good result might be deemed irrelevant and useless if it is surrounded by several unsuitable results. Therefore there might be certain advantages to measuring and evaluating the value or relevance of individual results over a whole set of results.

Direct User Relevance Feedback Relevance feedback lets users rate individual results or sets of results either directly (through manual ratings) or indirectly (through click-stream data). This data is then congregated and used for webpage rankings or other purposes such as suggesting other query terms. It can improve results for similar queries in the future but also lets the user stir the direction his search is taking in real-time. Users can adjust their query to manipulate the results; this basically means they adjust some of their own constraints.

“Relevance feedback—asking information seekers to make relevance judgments about returned objects and then executing a revised query based on those judgments—is a powerful way to improve retrieval.” (Marchionini 2006)

Automatic Query Expansion As opposed to integrating and involving the user actively in the refinement of a query, in automatic query expansion the improvements are done passively, often completely without the user’s knowledge. Information gathering methods include, for example, the analysis of mouse clicks, so called like buttons (e.g. Facebook, Google+) or eye tracking, etc. How the col-

lected data is then used varies. Simple examples of automatic query expansion are the correction of spelling errors or the hidden inclusion of synonyms when evaluating a query.

Depending on these factors and constraints, search results can be viewed as useful or useless. In a way the usefulness or correctness of an idea or result cannot always be judged fairly – there are always conditions that will affect how the outcome is interpreted. In the scenario of a creative search tool, results could be very useful, while they might be completely useless in another.

6.1 IR Evaluation

In this paper (Sawle, Raczinski and Yang 2011) we have discussed an initial approach to measure the creativity of search results. Based on a definition of creativity by Boden, we attempt to define creativity in a way which could be applied to search results and provide a simple metric to measure it.

Evaluating search results is not easy. The most widely used metric to measure the quality of results is that of precision and recall.

Precision is defined as the fraction of retrieved documents that are relevant.

$$Precision = \frac{\text{relevantdocumentsretrieved}}{\text{retrieveddocuments}} \quad (6.1)$$

Recall is defined as the fraction of relevant documents that are retrieved.

$$Recall = \frac{\text{relevantdocumentsretrieved}}{\text{relevantdocuments}} \quad (6.2)$$

Note the slight difference between the two. Precision tells us how many of all retrieved results were actually relevant (of course this should preferable be very high) and recall simply indicates how many of all possible relevant documents we managed to retrieve. This can be easily visualised as follows.

Precision is typically more important than recall in web search while it is the other way around in a database search system maybe. The mean average precision value (MAP) can be calculated following this formula (taken from Baeza-Yates and Ribeiro-Neto 2011, p.141):

$$MAP_i = \frac{1}{|R_i|} \sum_{k=1}^{|R_i|} P(R_i[k]) \quad (6.3)$$

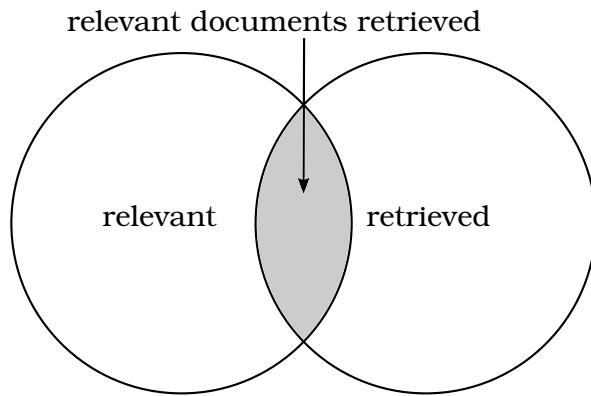


Figure 6.1: Precision and Recall

Where R_i is the set of relevant documents for query q_i .

But for many web searches is it not necessary to calculate the average of all results, since users don't inspect results after the first page very often and it is therefore desirable to have the highest level of precision in the first 5 to 30 results maybe. For this purpose it is common to measure the average precision of web search engines after only a few documents have been seen. This is called "Precision at n" or "P@n" (Baeza-Yates and Ribeiro-Neto 2011, p.140). So for example this could be P@5 or P@10 or P@20. For example, to compare two ranking algorithms, we would calculate P@10 for each of them over an average of 100 queries maybe and compare the results and therefore the performance of the algorithm.

The Text REtrieval Conference (TREC) is a conference that provides large test sets of data to participants and lets them compare results. They have specific test sets for web search comprised of crawls of .gov web pages for example, but unfortunately they have to paid for to get a copy.¹

There are certain other factors that can be or need be evaluated when looking at the complete search system. For example the time it takes to crawl for data, the time it takes to index data, the amount of storage needed for data and then also how fast the query response is and how many queries the system can handle in a given time period.

6.2 Approaches

Taking the debates about human creativity and directly applying them to machines seems logical but may be the wrong and lazy approach. Adapting Mayer's

¹http://ir.dcs.gla.ac.uk/test_collections/

five big questions to machines does not seem to capture the real issues at play.

1. Is creativity a property of programmers, users, machines, products, or processes?
 2. Is creativity a local, a network or an Internet phenomenon?
 3. Is creativity common or rare? (P or H creativity)
 4. Is creativity domain-general or domain-specific?
 5. Is creativity quantitative or qualitative?
-
- Can a machine judge whether a human is creative?
 - Is creativity a property of machines (hardware or software?)
 - Is mimicking human creativity really enough and appropriate?
 - Compare against “human creativity”? Or define machine creativity from scratch?
 - Who is creative? The programmer or the program?
 - Can creativity be objectively measured?
 - Quantitative or qualitative?
 - In respect to P or H creativity?
 - Output minus input? (we don't have the same strict judgement on humans)
 - Is it the product or the process or both?
 - Does context matter? (Blind deaf dumb person = computer?)
 - Does time matter?
 - Does purpose or intention matter?
 - AGI vs AI? Artificial general creativity vs artificial creativity?

On a more practical level, there are various problems that arise when trying to evaluate computer creativity. Anna Jordanous found that “evaluation of computational creativity is not being performed in a systematic or standard way” ([Jordanous 2011](#), p.2, her emphasis).

Please and Colton [27] divide it into two notions:

ref

- whether an idea or artefact is valuable or not, and
- whether a system is acting creatively or not.

6.3 Output Minus Input

It has been argued that “creativity may be seen as **output minus input.**” (Pease, Winterstein and Colton 2001, p.2, my emphasis). The output in this case is the creative product but the input is not the process. Rather, it is the “**inspiring set**” (comprised of explicit knowledge such as a database of information and implicit knowledge input by a programmer) of a piece of software.

The degree of creativity in a program is partly determined by the number of novel items of value it produces. Therefore we are interested in the set of valuable items produced by the program which exclude those in the inspiring set. (Colton, Pease and Ritchie 2001, p.3)

We are interested in how a single agent can come up with something that is novel **relative to its initial state of knowledge** (Ritchie 2007, p.72, his emphasis)

“Output minus input” might easily be misinterpreted as “product minus process”, however, that is not the case. In fact, Pease, Winterstein and Colton argue that “the process by which an item has been generated and evaluated is intuitively relevant to attributions of creativity” (2001, p.6), and that “two kinds of evaluation are relevant; the evaluation of the item, and evaluation of the processes used to generate it.” (2001, p.7). If a machine simply copies an idea from its inspiring set then it just cannot be considered creative, and they need to be disqualified so to speak.

6.4 Measurable Attributes

Simon Colton came up with an evaluation framework called the “creative tripod”. The tripod consists of three behaviours a system or artefact should exhibit in order to be called creative. The three legs represent **skill, appreciation, and imagination** and three different entities can sit on it, namely the programmer, the computer and the consumer. Colton argues that the perception “that the software has been skillful, appreciative and imaginative, then, regardless of the behaviour of the consumer or programmer, the software should be considered creative.” (2008b, p.5) + (2008a, p.5). As such a product can be considered creative, if it appears to be creative. If not all three behaviours are exhibited, however, it should not be considered creative. (Colton 2008b, p.5) + (Colton 2008a, p.5)

Imagine an artist missing one of skill, appreciation or imagination. Without skill, they would never produce anything. Without appreciation, they would produce things which looked awful. Without imagination, everything they produced would look the same (Colton 2008b)

Pease, Winterstein and Colton suggest that all creative products must be **novel and valuable** (2001, p.1) and provide several measures that take into consideration the context, complexity, archetype, surprise, perceived novelty, emotional response and aim of a product. In terms of the creative process itself they only discuss **randomness** as a measurable approach. Elsewhere, Pease et al discuss using **serendipity** as an approach (2013).

Davide Piffer suggests that there are three dimensions of human creativity that can be measured, namely **novelty, usefulness/appropriateness and impact/influence** (2012, p.258-259). As an example of how this applies to measuring a person's creativity he proposes citation counts (Piffer 2012, p.261). While this idea works well for measuring scientific creativity maybe, he does not explain how this would apply to a visual artist for example².

Graeme Ritchie identifies three main criteria for creativity as **novelty, quality and typicality** (2007, p.72-73), although he argues that “novelty and typicality may well be related, since high novelty may raise questions about, or suggest a low value for, typicality” (2007, p.73) (see also 2001). He proposes several evaluation criteria which fall under the following categories: (Ritchie 2007, p.91-92) basic success, unrestrained quality, conventional skill, unconventional skill, avoiding replication and various combinations of those.

This also somewhat suggests that variety is a criteria for creativity.

Dan Ventura later suggested the addition of **variety and efficiency** to Ritchie's model (2008, p.7).

Geraint Wiggins introduced a formal notation and set of rules for the description, analysis and comparison of creative systems (2006) which is largely based on Boden's theory of creativity (2003). The framework uses three criteria for measuring creativity: **relevance, acceptability and quality**.

Anna Jordanous proposed 14 key components of creativity, an ontology of creativity (2012, p.104-120), from a linguistic analysis of creativity literature which identified words that appeared significantly more often in discussions of creativity compared to unrelated topics. (2012, p.120).

²<http://www.artfacts.net> seems to provide just that though.

The themes identified in this linguistic analysis have collectively provided a clearer ‘working’ understanding of creativity, in the form of components that collectively contribute to our understanding of what creativity is. Together these components act as building blocks for creativity, each contributing to the overall presence of creativity; individually they make creativity more tractable and easier to understand by breaking down this seemingly impenetrable concept into constituent parts. ([Jordanous and Keller 2012](#), p.120)

The 14 components Jordanous collated are: ([2012](#), p.118-120)

1. Active Involvement and Persistence
2. Generation of Results
3. Dealing with Uncertainty
4. Domain Competence
5. General Intellect
6. Independence and Freedom
7. Intention and Emotional Involvement
8. Originality
9. Progression and Development
10. Social Interaction and Communication
11. Spontaneity / Subconscious Processing
12. Thinking and Evaluation
13. Value
14. Variety, Divergence and Experimentation

6.5 Linda Candy

Evaluation is well established in HCI. HCI is not unsimilar to creativity. Design too.

I guess her work is meant mostly for “interactive art” while mine is meant for creative computing, but clearly there are many overlaps.

In HCI, historically, the focus has been on people as users deploying task oriented systems. The criteria for evaluation has largely been in terms of ease of use, task efficiency and effectiveness- usability. However, attributes such as speed and productivity are, for the most part, meaningless in the context of creative interactive experience. ([Candy 2012](#), p.23)

Evaluation “is used to describe assessing and judging the value or worth of a particular idea or artifact both during the creative pro-

cess and afterwards. Whether the process is systematic or ad hoc, evaluation depends upon criteria and measures that are situated and domain specific." (Candy 2012, p.7)

"Whatever the context, evaluation is always tailored to the approach, needs, purpose and methodology of that context." (Candy 2012, p.7)

"For evaluation to contribute to a successful outcome, the practitioner needs to have the necessary information including constraints on the options under consideration." (Candy 2012, p.7)

"The matrix for evaluating creativity represents three standpoints: the capabilities of the creator, the audience, or more accurately, participant, experiences, and the features of the interactive systems as artworks. This initial matrix has been extended to include creative processes for both creator and audience participant (i.e. working practices and interaction experiences) and contextual factors in the form of the physical and technical environment in which the creative acts and events take place, including the influence of physical and technical resources and real world constraints." (Candy 2012, p.7-8)

"Evaluation studies are well established in the field of Human-Computer Interaction (HCI) as well as interaction design contexts in general." (Candy 2012, p.8)

"The evaluation of user, or rather, participant, experience of interactive artworks often involves measurement of aesthetic appreciation and the various engagement qualities which are dependent on personal traits, motivations, expectations, emotions and cognitive states of the audience. Those experiences that involve open-ended activity tend towards the creative end of human activity and, as such, are hardly ever measurable in quantitative ways." (Candy 2012, p.8)

"Evaluation is a key activity in creative design that can be revealed through documentation from design rationale. The introduction of rationale has been an important contribution to the quest for clarity and traceability in design decision-making. Design rationale

may be thought of as structured records of design that support the understanding of decisions taken and allow designers to give better informed reconsideration to them at a later stage." (Candy 2012, p.9)

"A software system can be viewed as an artifact that embodies implicit theoretical constructs that are realized as functional and operational requirements (Carroll and Campbell 1989). Structures are chosen because of their ability to achieve the intended functionality, and such choices may be evaluated against various criteria. During the design process, the ideas are modified and there is a clarification and refinement of intended functions and features. There may be additional factors arising from the context of the project that affect the way the design is carried out: for example the need to keep sight of general applicability whilst meeting the domain specific requirements, or the influence of the given hardware platform and software tools. Whatever the situation, the relationship between designers' decision making and the design outcome is not necessarily transparent and this is can be a problem when it comes to system maintenance. The explicit listing of decisions made during a design process, and the reasons why those decisions were made provides a means to record and communicate the reasoning and justification behind a design decision, including alternatives considered and constraints that affected the decision-making including why alternatives were rejected. The successful application of design rationale to software system design can provide a form of communication of intent from the designer to those who are to maintain the system." (Candy 2012, p.9)

"A promising approach to the externalization of decision-making during the design process is being explored within practice-based research in the creative arts in the form of documented reflective practice. The approach builds upon a normal part of creative practice whereby practitioners draw and note ideas, designs and options in their sketch and notebooks. In this way, the documentation of tentative ideas and how they are worked into firmer proposals through testing and evaluation is a familiar and integral part of creativity. In practice-based research, documented reflective practice and empirical studies are frequently brought together." (Candy 2012, p.10)

The Multi-dimensional Model of Creativity and Evaluation (MMCE) shown in Figure 1 has four elements: people, process, product and context. (Candy 2012, p.11)

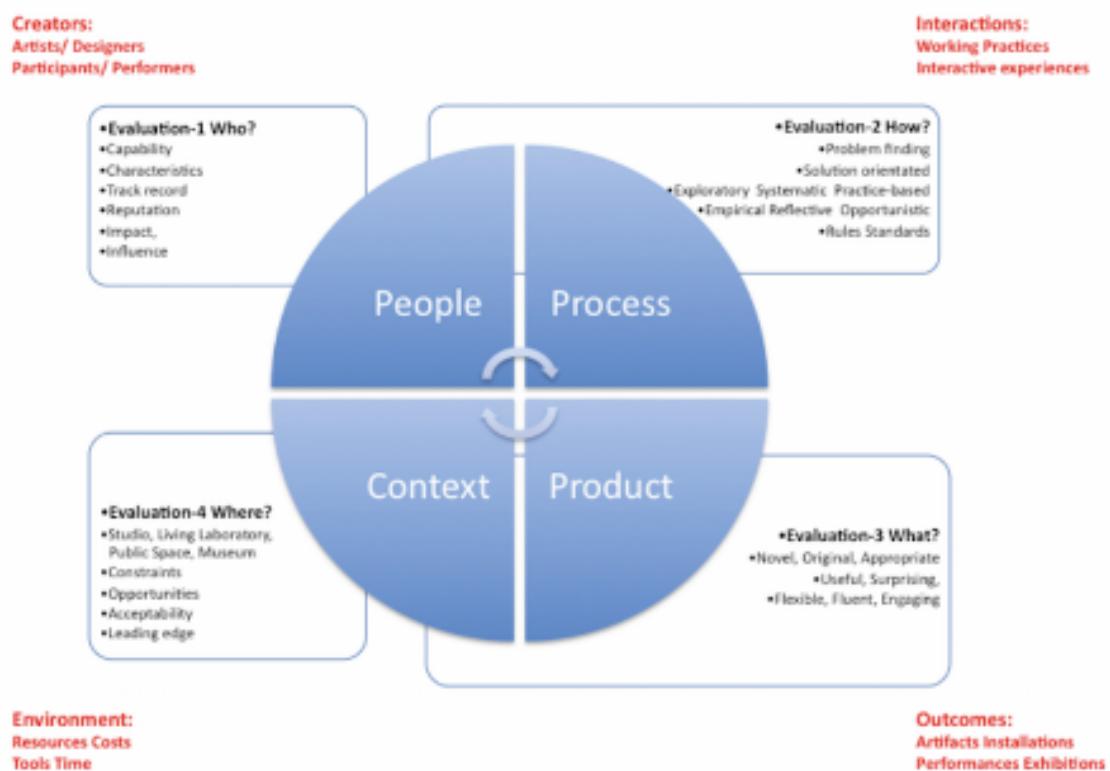


Figure 6.2: Candy's Multi-dimensional Model of Creativity and Evaluation

PEOPLE AS CREATORS: (Candy 2012, p.14-15)

Criteria for evaluating creator capability:

1. The creator must be able to demonstrate an ability to create an artistic outcome where subject matter, ideas and technique are combined well to produce a coherent outcome.
2. The creator must be able to demonstrate an ability to make work that is exploratory, creative and imaginative. Interesting ideas are presented in intelligent and surprising ways.
3. In respect of Composition and Interpretation, the creator must be able to demonstrate the ability to:
 - Select subject matter that is appropriate to a given theme
 - Manipulate ideas and techniques in a coherent manner
 - Express ideas visually (visual communication)
 - Respond in an individual and personal way

PROCESS AS INTERACTION: (Candy 2012, p.17)

Criteria for evaluation can be expressed as follows 1. For a work to be deemed

Categories in Creativity	Actors & Elements	Instances	Feature Descriptors	Outcomes	Criteria, Qualities, Values to Measure
People - Who is involved?	Creators	i) designer, artist	imagination, artistry, expertise, skill, experience, intention, reputation	ii) participant, performer	profile, demographic, motivation, skills, experience, curiosity, commitment
Process - How is it done?	Interactions	ii) working practices	inspirational, solution driven, goal oriented, problem solver, systematic, exploratory, reflective, risk taking	ii) interactive behaviors	opportunistic, adventurous, curious, cautious, expert, knowledgeable, experienced
Product - what comes out?	Outcomes	ii) artifacts, installations	novel, original, surprising, compelling, adaptable, aesthetically pleasing, effective	iii) performances, experiences	immediate, engaging, purposeful, enhancing, exciting, disturbing,
Context - where does it happen and with what?	Environments	ii) studio, living laboratory ii) public spaces, museums	physical space, light, technical facilities, rules, costs, time resources, effort, organizational constraints/support	ii) prototypes ii) exhibitions	design quality, usable, convincing, adaptable, effective, innovative, transcendent

Table 1: Categories of Creativity: Actors and Elements with Evaluation Measures: Criteria, Qualities, Value

Figure 6.3: Caption text under figure

engaging, the participant should exhibit observable responses. There are likely to be different levels of engagement depending on whether or not the audience has had prior experience of this kind of artwork or installation or similar.

2. The participant responses demonstrate active engagement in three ways: Immediate, Sustained or Creative. The categories are defined as follows:

- Immediate engagement: the work grabs immediate attention and yet is not so mundane as to create boredom.
- Sustained engagement: the work must excite curiosity in the and also be accessible to a general audience.
- Creative engagement: the work must excite immediate attention and encourage an audience to interact with it in a playful/purposeful way. As attention declines with familiarity and time, changes take place in the work that renew audience engagement.

PRODUCT AS OUTCOME: (Candy 2012, p.18)

Typical features for judging artworks include composition, aesthetic, affect, content, and technique. Criteria for evaluation can be expressed as follows:

- the composition of work should be coherent, exhibit shape and balance between order and complexity.
- the work should exhibit outstanding visual and sound qualities in color, line and form.
- the work should be pleasing, challenging, exciting, original etc.
- the content should be appropriate and effective for the chosen subject matter
- the execution should demonstrate high quality technique that fits the form.

It is interesting, therefore, to consider how criteria for judging the digital arts are specified by the Prix Ars Electronica, an international competition for Cyber Arts and the foremost event of its kind today.

Entries are judged by a Jury of experts in the order of their arrival and according to the following categories:

- Aesthetics • Originality
- Excellence of execution
- Compelling conception
- Innovation in technique of the presentation

CONTEXT AS ENVIRONMENT: (Candy 2012, p.21)

Establishing a workable living laboratory for interactive art and evaluation involved setting down acceptance criteria for assessing whether or not a new interactive art system was ready to be deployed.

These included:

- degree of robustness of the art system in expectation of heavy public use
- appropriate accessibility in respect of type of audience (e.g. children)
- adherence to safety and house rules required by the museum
- impact of other coinciding exhibits (sound, noise, light impacts)
- attention to participant orientation and training
- attention to art system maintenance by creator and technical support

“we need to apply strategies for generating clear and unambiguous data that can be turned into meaningful information. From meaningful information, we can then derive understandings related to the context of use, the outcome of which might take the form

of a coherent model." ([Candy 2012](#), p.21)

"Observation as a method for data collection raises issues as to its reliability in creativity evaluation. Data from observing creativity depends upon the interpretation of what the individual observer sees." ([Candy 2012](#), p.22)

"However, in order to 'measure' creativity, we have to conduct research outside of controlled laboratory conditions, and cannot rely on fixed criteria that can be applied to all cases. The shifting ground and the ever-changing contexts often renders consistency out of reach." ([Candy 2012](#), p.22)

"If the term 'measurement' does not match what we are doing within the creativity domain, then why do we still use this word?" ([Candy 2012](#), p.22)

"Whether an action is successful or unsuccessful depends on whether the intended result is achieved." ([Candy 2012](#), p.23)

"Measuring success is more likely to be dependent on factors such as whether or not the system has engaged the audience in a playful or immersive way or whether it has elicited curiosity or excitement or concentrated attention and so on." ([Candy 2012](#), p.23)

Part III

THE C \ominus RE: T Σ CHN \ominus -L \ominus GIC

Chapter 7

Theoretical Foundations

7.1	Searching vs. Browsing	112
7.2	Creativity	113
7.3	Pataphysics	114
7.3.1	and Creativity	114
7.3.2	and Computers	118
7.4	Patalgorithms	121
7.4.1	Pataphysicalisation	121
7.4.2	Patadata	123
7.4.3	Ranking — Pranking?	124
7.5	User experience	125

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

Why pataphysics? How does pataphysics relate to creativity and how does it support creativity in computers?

- link the various creativity models
- how do they apply to computers?
- link pataphysics to creativity — and computers
- ranking and pataphysics
-

7.1 Searching vs. Browsing

What do we actually mean by searching? Usually it implies that there is something to be found, an **Information Need (IN)**; although that doesn't necessarily mean that the searcher knows what he or she is looking for or how to conduct the search and satisfy that need.

From the users' point of view the search process can be broken down into four activities (**Sutcliffe and Ennis 1998**) reminiscent of classic problem solving techniques (**Polya 1957**):

Problem identification

IN,

Need articulation

IN in natural language terms,

Query formulation

translate IN into query terms, and

Results evaluation

compare against IN.

This model poses problems when we consider a situation where an IN cannot easily be articulated or in fact is not existent and the user is not looking for anything. This is not the only constraining factor though and Marchionini and Shneiderman have pointed out that "the setting within which information-seeking takes place constrains the search process" (**Marchionini and Shneiderman 1988**) and they laid out a framework with the following main elements.

- Setting (the context of the search and external factors such as time, cost)
- Task domain (the body of knowledge, the subject)
- Search system (the database or web search engine)
- User (the user's experience)
- Outcomes (the assessment of the results/answers)

Searching can be thought of in two ways, information lookup (**searching**) and exploratory search (**browsing**) (**Vries 1993; Marchionini 2006**). A situation where an IN cannot easily be articulated or in fact is not existent (the user is not looking for anything specific) can be considered a typical case of exploratory search and describes the kind of search that is most suited to our proposed tool. The former can be understood as a type of simple question answering while the latter is a more general and broad knowledge acquisition process without a clear goal.

Current web search engines are tailored for information lookup. They do really well in answering simple factoid questions relating to numbers, dates or names (e.g. fact retrieval, navigation, transactions, verification) but not so well in providing answers to questions that are semantically vague or require certain extend of interpretation or prediction (e.g. analysis, evaluation, forecasting, transformation).

When it comes to exploratory search though, the user's success in finding the right information depends a lot more on constraining factors such as those mentioned earlier and can sometimes benefit from a combination of information lookup and exploring ([Marchionini 2006](#)).

Much of the search time in learning search tasks is devoted to examining and comparing results and reformulating queries to discover the boundaries of meaning for key concepts. Learning search tasks are best suited to combinations of browsing and analytical strategies, with lookup searches embedded to get one into the correct neighbourhood for exploratory browsing. ([Marchionini 2006](#))

De Vries called this form of browsing an “enlargement of the problem space”, where the problem space refers to the resources that possibly contain the answers/solutions to the information need ([Vries 1993](#)). This is a somewhat similar idea to that of Boden’s conceptual spaces which she called the “territory of structural possibilities” and exploration of that space “exploratory creativity” ([Boden 2003](#)).

All of these ideas, however, seem to be concerned with how users interact with a search system, rather than how the system acts itself. So we need to shift our perspective and think about how a search tool can be more supportive for exploratory search directly and by what means.

7.2 Creativity

comparison table Poincare, Wallas, Polya, CC?

We had previously differentiated between creative computing and computational creativity:

Intuitively the former is about doing computations in a creative way, while the latter is about achieving creativity through computation.

Boden vs Kaufman & Beghetto

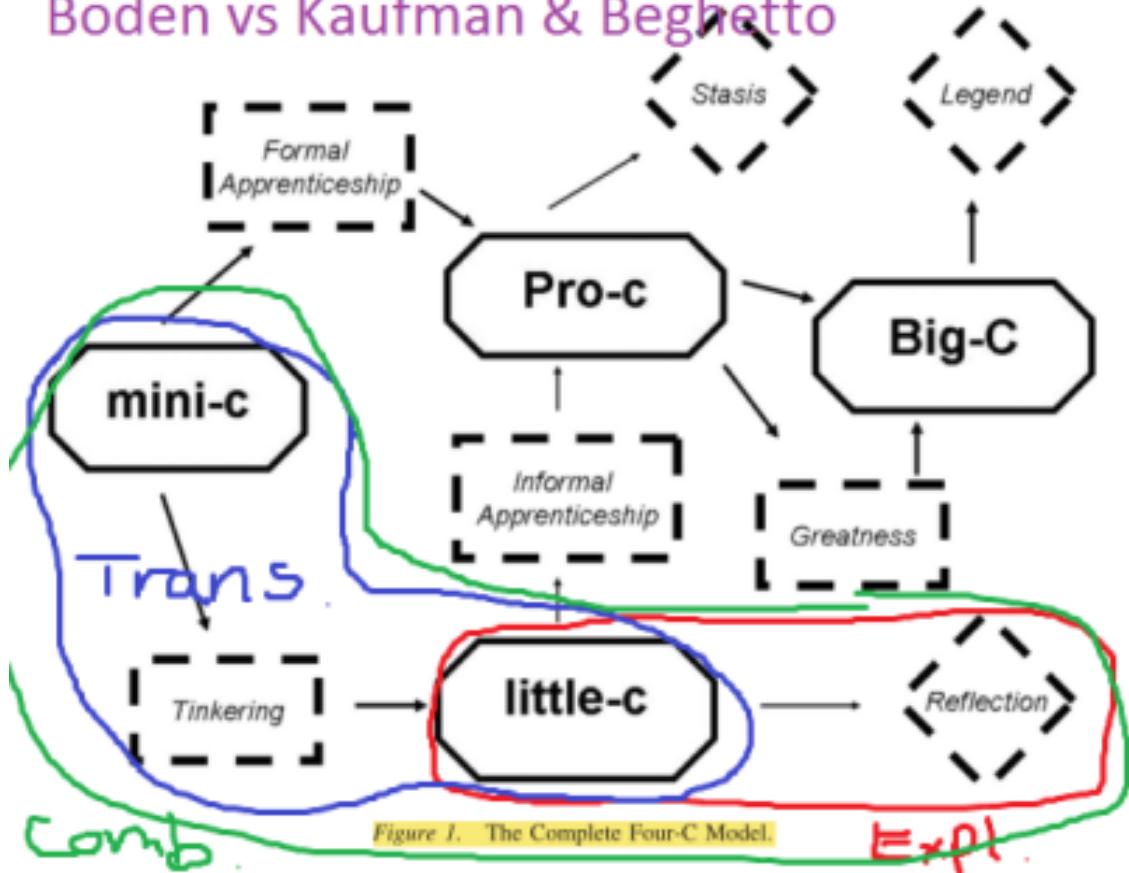


Figure 7.1: Kaufman's 4 C Model vs. Boden

You can think of the latter falling into the artificial intelligence category (using formal computational methods to mimic creativity as a human trait, see also [18]) and the former being a more poetic endeavour of how the computing itself is done, no matter what the actual purpose of the program is. (Hugill 2013)

The differences are subtle but clear.

How does ethics fit into this?

7.3 Pataphysics

7.3.1 and Creativity

The creative process normally involves a move from the known to the unknown and sometimes from the named to the unnamed. In bringing something new into existence, the human qualities of openness and tolerance of ambiguity are

Creative Computing	Digital Humanities	Computational Creativity
Motivation	Design	Intentionality
Ideation	Curation, Analysis, Editing, Modeling and Networks, Infrastructure	Framing
Implementation	Computation, Processing	Process
Operation	Versioning, Prototyping, Failures	Product

Table 7.1: Comparison of Creative Computing vs Digital Humanities vs Computational Creativity

4 step model	4 P Model	Problem Solving
Preparation	Person	Understand
Incubation	Press/Context	Plan
Illumination	Process	Carry Out
Verification	Product	Look back

Table 7.2: Comparison of 4 Step Model vs 4 P Model vs Problem Solving

generally regarded as highly desirable. We may define creativity as **the ability to use original ideas to create something new and surprising of value**. We generally speak of creative ‘ideas’ rather than ‘products’, which merely provide evidence of a creative process that has already taken place. Both the originality and the value of an idea are evaluated using subjective criteria. **Pataphysics**, which represents an extreme form of subjectivity, is therefore a highly appropriate framework within which to encourage and enable creative thinking and operations.

Both the originality and the value of a creative idea are always evaluated using purely subjective criteria. Pataphysics, which represents an extreme form of subjectivity, is therefore a highly appropriate framework within which to encourage and enable creative thinking and operations.

The ambiguity of experience is the hallmark of creativity, that is captured in the essence of pataphysics. Traversing the representations of this ambiguity using algorithms inspired by the syzygy, clinamen and anomaly of pataphysics, using a panalogical mechanism applied to metadata, should be able to humanize and even poeticize the experience of searching the Web. ([Hendler and Hugill 2013](#))

Pataphysics is highly subjective and particular and is as such very suitable for this kind of transformation from relevant to creative.

[Pataphysics] can only be defined in a new undiscovered language because too obvious: tautology. ([Baudrillard 2007](#))

It is instructive to overlay these ideas on existing theories of creativity. Margaret Boden ([Boden 2003](#)), for example, has defined **P-creativity** (short for psychological creativity) as the personal kind of creativity that is novel in respect to the individual mind and **H-creativity** (short for historical creativity) as fundamentally novel in respect to the whole of human history. This allows for subjective evaluation of any idea.

Using Boden's definition we can call an idea 'new' if it is new to the individual who came up with it, making the idea P-creative. We can say that a creative idea can be seen from two perspectives: the subjective (P-creative) and the objective (H-creative) view. She argues that constraints support creativity, and are even essential for it to happen. "Constraints map out a territory of structural possibilities which can then be explored, and perhaps transformed to give another one" ([Boden 2003](#), p.82).

This echoes the ideas of groups such as the **Oulipo** (which began as a Sub-Commission of the Collège de 'Pataphysique), who investigate 'potential literature' by creating constraints that frequently have a ludic element. Various other groups, the Ou-x-Po, perform similar operations in fields as diverse as cinema, politics, music and cooking ([Motte 2007](#)).

Boden's conceptual space is the "territory of structural possibilities". So, the conceptual space of a teacup might be that it is meant to carry a certain amount of tea without breaking or burning fingers. It wouldn't be wise to create a teacup made out of paper. But whether we make a cup out of glass or porcelain, or how we shape the cup or the handle is pretty much up the individual's creativity. Being able to move around in this conceptual space, experiment (in thought or in reality) and play with different ideas while still following a given set of constraints is a good starting point for creativity to happen. Boden defines three sub-types of creativity.

The Oulipo similarly classifies its conceptual space under two broad headings: the synthetic and the analytic:

[...] In the research which the Oulipo proposes to undertake, one

may distinguish two principal tendencies, oriented respectively towards Analysis and Synthesis. The analytic tendency investigates works from the past in order to find possibilities that often exceed those their authors had anticipated. [...] The synthetic tendency is more ambitious: it constitutes the essential vocation of the Oulipo. It's a question of developing new possibilities unknown to our predecessors. This is the case, for example, of [Raymond Queneau's] 100,000,000,000 Poems or the Boolean haikus. ([Motte 2007](#), p.27)

Later writings develop these ideas in more detail. *La Littérature Potentielle Oulipo 1973*, is divided into several sections, dealing with clusters of methods, that include: anoulipisms (analytical oulipisms, such as combinatorial literature); use of preexisting structures such as lipograms (omitting a letter or letters), palindromes and snowballs (in which each successive word adds or subtracts a letter), homophonic translation, tautogram, and definitional literature; lexical, syntactic, or prosodic manipulations (such as the celebrated S+7, in which each substantive is replaced by the seventh word after it in a standard dictionary); lexicographical or prosodic synthoulipisms (early algorithmic methods); and perimathematical synthoulipisms (such as the Boolean poetry and combinatorial works already mentioned).

Boden links her three aspects of creativity to three sorts of surprise. She says that creative ideas are surprising because they go against our expectations. “The more expectations are disappointed, the more difficult it is to see the link between old and new.” ([Boden 2003](#), p.84) This suggests that fewer **expectations** (an open mind) allow creativity to happen more easily. Empirical experiences form expectations, which hinder our ability to accept creative ideas when they happen. In order to be able to recognise creative ideas we need to be able to see what they all have in common and in what way they differ and not reject unusual, unexpected ones.

“Unless someone realizes the structure which old and new spaces have in common, the new idea cannot be seen as the solution to the old problem. Without some appreciation of shared constraints, it cannot even be seen as the solution to a new problem intelligibly connected with the previous one.” ([Boden 2003](#), p.84)

It is clear that the Oulipo has a similar approach in its theorising of potential literature. Releasing creativity through constraint is its essential *raison d'être*.

This is not to say that experience and knowledge are necessarily bad for creativity. To appreciate creativity we need to be knowledgeable in the relevant domain to be able to recognise old and new connections and transformations. But we also need a certain level of openness and tolerance for ambiguity to overcome our expectations. Perhaps it is for this reason that ‘creative people’ are often assumed to have particular personality traits. Sternberg ([Sternberg 1999](#); [Sternberg 1999](#)), for example, proposes that these comprise: independence of judgement, self-confidence, and attraction to complexity, aesthetic orientation, and tolerance for ambiguity, openness to experience, psychoticism, risk taking, androgyny, perfectionism, persistence, resilience, and self-efficacy. More empirically, Heilman, Nadeau and Beversdorf ([Heilman, Nadeau and Beversdorf 2003](#)) have investigated the possible brain mechanisms involved in creative innovation. While a certain level of domain specific knowledge and special skills are necessary components of creativity, they point out that “co-activation and communication between regions of the brain that ordinarily are not strongly connected” might be equally important.

Newell, Shaw and Simon add to the above with their report on the creative thinking process ([Newell, Shaw and Simon 1963](#)). They identify three main conditions for creativity: the use of imagery in problem solving; the relation of unconventionality to creativity; and the role of hindsight in the discovery of new heuristics. Other issues they point out are abstraction and generalisation. So, for example, poets transform the grammar of their conceptual space (in this case, language) to create new sentence structures in a poetic form. By doing so, they go against the expectations, the possibilities of the language and cause surprise. Some people might not understand the transformations and therefore the jokes or beauty of a poem simply because they are either not able to recognise connections between the old and newly transformed elements (maybe due to a lack of knowledge in the poems topic or in that particular language) or because they do not want to accept unconventional methods.

Consider the following table, which compares some of the key ideas of creativity ([Boden 2003](#); [Indurkhya 1997](#); [Koestler 1964](#)) with the main pataphysical operations. It will be seen that pataphysics succeeds in bringing into sharp relief the more generalised scientific ideas. The pataphysical terms are taken from the natural sciences or philosophy, but always with an ironic twist, betraying their underlying humour. They connect quite strongly with the primary descriptors of creativity, while adding a certain layer of jouissance. **Pataphysics is self-avowedly useless, but its principles may prove surprisingly useful within this context.**

7.3.2 and Computers

CREATIVITY	PATA PHYSICS
Combinational: Juxtaposition of dissimilar, bisociation, deconceptualisation	Antinomy: Symmetry, duality, mutually incompatible, contradicting, simultaneous existence of mutually exclusive opposites
	Syzygy: Alignment of three celestial bodies in a straight line, pun, conjunction of things, something unexpected and surprising
Exploratory: Noticing new things in old places	Anomaly: Exceptions, equality
Transformative: Making new thoughts possible by transforming old conceptual space, altering its own rules	Clinamen: Unpredictable swerve, the smallest possible aberration that can make the greatest possible difference

Table 7.3: Creativity vs Pataphysics

ANDREW:

Since our solutions will be imaginary, our aim is not so much to have the computer generate creative artefacts as to engage in a creative dialogue with the user. Therefore, we do not intend to move as close to artificial intelligence as Colton's framework seems to suggest. In the pataphysical universe, ideas such as 'human skill', 'human imagination' and 'human appreciation' are too generalised to be useful. One may very well ask: which human? And when, where and even why? Rather, our project will aim to produce an exceptional computational entity that consistently generates surprising and novel provocations to the users, who in turn may navigate and modify these by deploying their own skills, appreciation and imagination. The relationship between the two will develop quite rapidly into one of mutual subversion since, however apparent the 'rules of the game' may become, the outcomes will always be particular or exceptional.

We are not the first people to attempt to apply pataphysical ideas in computer science. Johanna Drucker focused specifically on the cleft between formal logic and subjective judgement. She introduced the discipline of 'Speculative Computing' as a solution to that problem (Drucker and Nowviskie 2007). The concept can be understood as a criticism of mechanistic, logical approaches that distinguish between subject and object.

Speculative computing takes seriously the destabilization of all categories of entity, identity, object, subject, interactivity, process, or instrument. In short, it rejects mechanistic, instrumental, and formally logical approaches, replacing them with concepts of autopoiesis (contingent interdependency), quantum poetics and emergent systems, heteroglossia, indeterminacy and potentiality, intersubjectivity, and deformance. Digital Humanities is focused on texts, images, meanings, and means. Speculative Computing engages with interpretation and aesthetic provocation. ([Drucker 2009](#), p.29)

For Drucker, aesthesis (ambiguous and subjective knowledge) is fundamentally opposed to mathesis (formal objective logic) and subjectivity is always in opposition to objectivity. Knowledge is a matter of interpretation of information, which can be represented digitally as data and metadata. She introduces what she calls a '**patacritical**' method of including exceptions as rules, even if repeatability and reliability are compromised. Bugs and glitches are privileged over functionality, and are "valuable to speculation in a substantive, not trivial, sense." As she says: "Pataphysics inverts the scientific method, proceeding from and sustaining exceptions and unique cases" ([Drucker and Nowviskie 2007](#)).

In order to break out of the formal logic and defined parameters of computer science, she asserts, we need speculative capabilities and pataphysics. "The goal of pataphysical and speculative computing is to keep digital humanities from falling into mere technical application of standard practices." She links interface design with other speculative computing principles, and refers to Kant's idea of art as '**purposiveness without purpose**'. She says that the appreciation of design as a thing in itself (regardless of utility) is a goal of speculative aesthetics.

The projects Johanna Drucker describes in her book SpecLab ([Drucker 2009](#)) could certainly be considered related work. Not only in their theoretical foundations but also in some aspects of their implementation. One project in particular is worth mentioning here: the 'Patacritical Demon, an "interactive tool for exposing the structures that underlie our interpretations of text", although it remained a purely conceptual piece of work and was never implemented. Her idea if the "patacritical" method is quite interesting. Pataphysical exceptions and anomalies can thus be justified in a computational system. But it is not just this concept that deserves mention here. Her ideas on structured data, metadata and knowledge representation link very nicely into my project. How can we represent and structure data so that it does not lose its subjectivity, context and meaning? Her reference to graphical analogies is inspiring in that regard as well. I am certain I will refer back to her concepts throughout my thesis.

7.4 Patalgorithms

The proposed concept for a pataphysical algorithm requires precise data structures to represent the transformations that have taken place during the pataphysicalisation, such as the patadata. The system's index has to be adapted to accommodate this new type of data structure. It also needs to be flexible enough to allow algorithms to fit in at different stages or locations of the system, for example the inverted-index, ranking functions or query itself. Whilst this new style of algorithm has been proposed, current architectures are not capable of supporting them. As such a new flexible component-based software architecture has been proposed which will allow for a range of different style systems to be developed with little overhead. As such improving the chance of creative outcomes occurring in a different way.

Linking back to some of the creative, pataphysical concepts we have discussed earlier, let us put some of the ideas for our tool into perspective. The constraints for our conceptual space are the pataphysical rules that we want to apply to our data. We use those rules to explore, combine and transform our space; giving us the flexibility and freedom we need to find interesting results.

We developed the idea of pataphysicalising data as the process of applying such pataphysical rules in order to produce creative search results. This pataphysicalisation process forms a central component of our system (see Figure 7.2) and influences all areas of the search tool.

redraw figure

7.4.1 Pataphysicalisation

The conceptual space for our project is 'pataphysical Web searching'. There are some very simple rules or constraints that form an initial definition of the project. For example it is clear that we want to search the World Wide Web (rather than a library database), that we want to return a list of search results (and not a pile of books) and that we want the search process and its results to be creative/pataphysical (rather than relevant). In a more technical sense, we have the query term (s), the index (of all web pages that we have crawled) and some pataphysical rules in our conceptual space. How we structure our search system, how we format the index or how we go about finding our results, is not in our conceptual space however. We can explore the space to its limits and we can transform it if we want to or feel like we need to. Our pataphysical rule set will include methods for transforming the space. By applying pataphysical rules to find results to our query we are pataphysicalising the query.

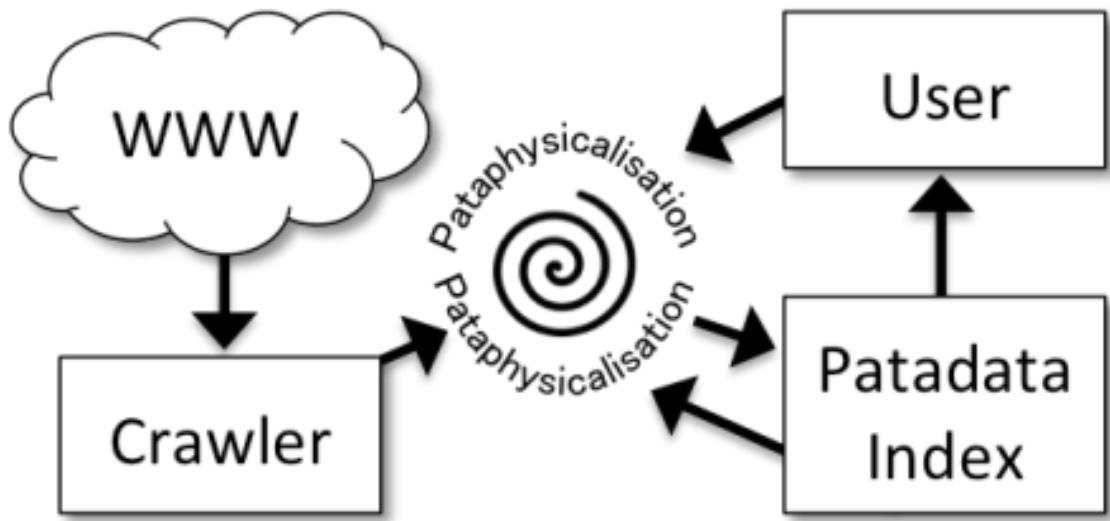


Figure 7.2: Pata central

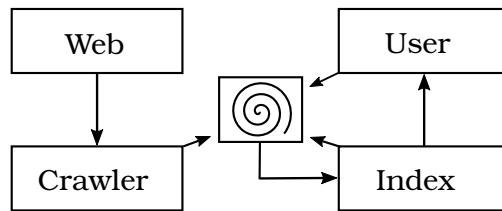


Figure 7.3: Pata central

Definitions:

To pataphysicalise

(verb) – applying pataphysical transformations

Pataphysicalisation

(noun) – the process of pataphysicalising

Patadata

(noun) – any data which has been pataphysicalised

But what exactly does the process of pataphysicalisation include? The kinds of transformations we are thinking of could be for example replacing or adding to the query term (s) with synonyms, antonyms, opposites, syzygies, clinamens etc. This can be done with the help of thesauri or dictionaries and ontologies. Whether we pataphysicalise our query term (s), the index or the results does not matter at this point. They are all possible and will maybe be done all at the same time (see figure 7.3 below). We can consider the possibility of a ‘pa-

tametric index', rather than a parametric index or a 'patasaurus' (pataphysical thesaurus/ontology).

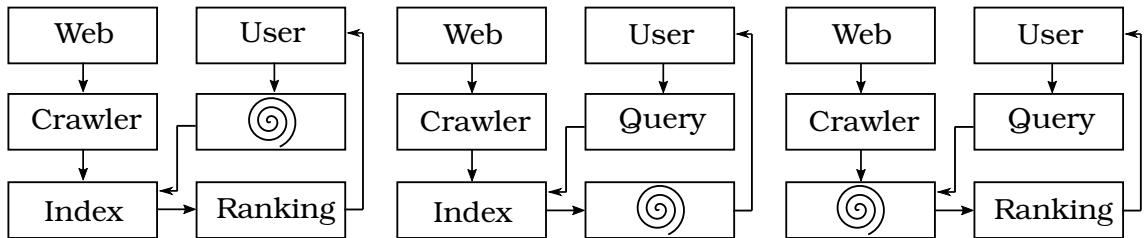


Figure 7.4: Pataphysicalisation

Arguably, few other textual forms will have greater impact on the way we read, receive, search, access, use and engage with the primary materials of humanities studies than the metadata structures that organize and present that knowledge in digital form. (Drucker 2009, p.9)

7.4.2 Patadata

The idea of patadata is derived from the idea below:

Physics → Metaphysics → Pataphysics

Data → Metadata → Patadata

Patadata will allow us to engage with digital knowledge in a more creative way even. If metadata helps us organise information semantically then patadata is for organising information pataphysically. If metadata is objective then patadata is subjective. Drucker also points out that "many information structures have graphical analogies and can be understood as diagrams that organise the relations of elements within the whole." (Drucker 2009, p.16) So maybe patadata could allow us to represent these graphical analogies in some way? An alphabetical list is a typical model for representing text data sets for example. Or an otherwise ranked list, a tree structure, a matrix, a one-to-many relationship, etc. But is a ranked list really the best way to represent search results? Ranking itself seems unpataphysical. It contradicts the philosophy of pataphysics, although we can argue that this contradiction makes it pataphysical again. Maybe this dilemma can be solved simply by adopting another type of graphical analogy to structure the results such as a tree structure instead of a ranked list.

Example: Let's say our patadata is represented by a list of keywords that each stands for a pataphysicalisation of the original query term. This list is added to each item in the index.

Query = 'Tree'

Patadata = [Tree (equivalent), Car (opposite), Paper (antinomy),
Narwhal (anomaly), Book (syzygy), Venus Fly Trap (clinamen)]

Query = 'Sun God Ra'

Patadata = [Sun God Ra (equivalent), Slave (opposite), Holiday (antinomy),
Blue Balloon (anomaly), Pyramid (syzygy), Sphinx (clinamen)]

7.4.3 Ranking — Pranking?

In traditional Web search, ranking signals contribute to the improvement of the ranking process. These can be content signals or structural signals. Content signals are referring to anything that is concerned with the text and content of a page. This could be simple word counts or the format of text such as headings and font weights. The structural signals are more concerned about the linked structure of pages. They look at incoming and outgoing links on pages. There are also Web usage signals that can contribute to ranking algorithms such as the clickstream. This also includes ideas such as the Facebook 'like' button or the Google '+1' button which could be seen as direct user relevance feedback.

Ranking can be done at different stages of the search process. Depending on how the index is formatted and what information can be pre-computed at that stage, the ranking algorithm evaluates every Web page for relevance and returns them in order. There exist lots of different approaches on ranking, including PageRank ([Brin and Page 1998b](#)) and HITS ([Kleinberg 1999](#)), which both analyse the link structure of the World Wide Web. They analyse the incoming and outgoing links on pages. PageRank for example assigns a numerical weight to each document, where each link counts as a vote of support in a sense. It is executed at indexing time, so the ranks are stored with each page directly in the index. HITS stands for 'Hyperlink Induced Topic Search' and its basic features are the use of so called hubs and authority pages. It is executed at query time. Pages that have many incoming links are called authorities and pages with many outgoing links are called hubs.

Given a query term X, what is considered a relevant match though? Do we simply return a list of Web pages where X appears in the heading of each page? It is obviously not that easy. Several ranking signals are combined together; Google states that they use over 200 signals including PageRank and they personalise results using signals such as the web history and location (Google n.d.). What kinds of ranking signals do we need for our pataphysical Web search tool? We could say that a page Y is relevant if it matches the patadata for query X. So, for example, Y would be a relevant result if it is a clinamen or syzygy to X.

The more patadata matches there are the higher the ranking maybe. We don't necessarily have to assign a numerical ranking value to each page. Depending on how we structure our results page that might not be necessary. Shuffling the results list or the results tree could be an option.

7.5 User experience

Whilst developing a system that returns creative results to the end user has numerous advantages, the assumptions that are made about and the decisions we take for the user must still be considered. For example, presume that the user inputs a search request 'The Cat in the Hat' after reading a Dr. Seuss book to their child, and the system employs an anomalous method on the query and searched 'sunglasses'. Whilst there is logic to the new search request, it is anomalous to the initial request, if the user receives these results without being told what method was used, the results will appear random, and therefore are likely to be detrimental to the user. Therefore the level of interaction the user has with the system and the feedback the system gives to the user on decisions it is making will have a large influence on the overall effectiveness and appreciation of the search tool. A quick and simple solution to this problem would be to add an icon to the side of each search result which displays how the original query was pataphysicalised.

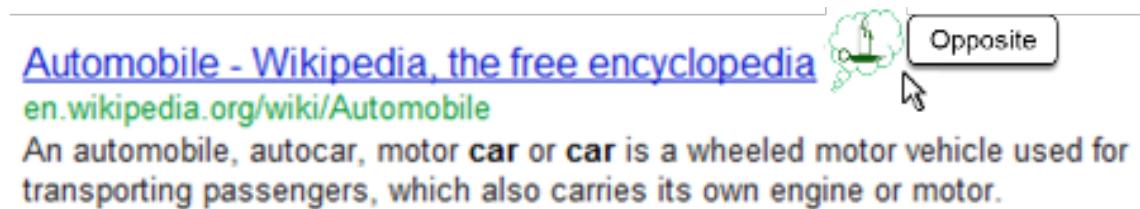


Figure 7.5: Feedback button

The above image (figure 7.5) shows an example of how this could be implemented. The little green candle (a reference to pataphysics in itself by the way) shows a pop-up note when hovered over with the mouse pointer. In this case the original query could have been 'tree' and 'car' was returned as an opposite to that.

Chapter 8

Practical Implementation

8.1	Corpus	128
8.2	Setup	130
8.3	Text	133
8.3.1	Clinamen	133
8.3.2	Syzygy	136
8.3.3	Antinomy	137
8.4	Image & Video	138
8.4.1	REST & API	139
8.5	Design	141
8.6	Prototypes	144
8.6.1	Prototype 2	144
8.6.2	Prototype 1	145

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

style inline code

run code on laptop and get snippets of all variable contents, e.g. faustroll, froll_dict, ...

give examples of different results if using different base documents!

add section about which pieces of code are not written by me

The website <http://pata.physics.wtf> showcases the current proof-of-concept algorithms. This chapter gives an overview of the structure of the website and the development process.

Typically, software development is divided into so-called front and back ends. The frontend includes web design and web development and is meant to provide an interface for the end-user to communicate with the backend which involves a server, an application and a database (although this is not completely true in this project).

The frontend design is created using the **w3.css** stylesheet as a basis. The website is mostly responsive, meaning it can be viewed well on phones, tablets and screens (the poems and image spirals for example unfortunately have a fixed width which does not scale down well). The site contains various scripts written in **Javascript** (e.g. scramble letters, randomise poem, send email and tabbed content).¹

The backend relies heavily on a **Python** framework called **Flask**. Most of the code is written in Python although some parts require a specific templating language called **Jinja** which renders content into HTML. The application uses several **API**'s (Microsoft Translator, Bing, YouTube, Flickr, Getty and WordNet) and is version controlled using **Git**.²

The folder structure is as follows:

```
- app
  -- static
    -- css
    -- images
    -- corpus
    -- templates
  - .git
- dev.py
- guni.py
- live.py
- .gitignore
- README.md
```

¹frontend links: <http://www.w3schools.com/w3css/>, <https://www.javascript.com/>

²backend links: <https://www.python.org/>, <http://flask.pocoo.org/>, <http://jinja.pocoo.org/>, <https://git-scm.com/>

- TODO.txt

folder structure

To provide a short overview, the tool's workflow can be described as follows:

1. Tokenise texts and remove stopwords to build index,
2. a query triggers the three pataphysical algorithms,
3. each algorithm finds results for the query,
4. retrieve some words before/after match for context, and
5. render the resulting sentences.

add audio? update this section depending on what i do

From the homepage users can choose between text, image and video search. Then they can enter a query — in the case of text search this should be single words only, image and video search support multi word queries.

8.1 Corpus

Instead of crawling the Internet the present tool uses a local collection of texts in its text-search. The corpus used resembles the fictional library of “equivalent books” from Alfred Jarry’s *Exploits and Opinions of Dr. Faustroll, 'Pataphysician* (1996, p.10-12)³. In principle the corpus is just a folder within the tool’s directory structure which contains the following files:

0. Alfred Jarry: *Exploits and Opinions of Dr. Faustroll, 'Pataphysician*
1. Edgar Allan Poe: *Collected Works*
2. Cyrano de Bergerac: *A Voyage to the Moon*
3. Saint Luke: *The Gospel*
4. Leon Bloy: *Le Desespere* (French)
5. Samuel Taylor Coleridge: *The Rime of the Ancient Mariner*
6. Georges Darien: *Le Voleur* (French)
7. Marceline Desbordes-Valmore: *Le Livre des Mères et des Enfants* (French)
8. Max Elskamp: *Enluminures* (French)
9. Jean-Pierre Claris de Florian: *Les Deux Billets* (French)
10. *One Thousand and One Nights*

³“In addition, three prints hanging on the walls, a poster by TOULOUSE-LAUTREC, *Jane Avril*; one by BONNARD, advertising the *Revue Blanche*; a portrait of Doctor Faustroll, by AUBREY BEARDSLEY; and an old picture, which appeared to us to be valueless, *Saint Cado*, issued by the Oberthuer printing house of Rennes.”(Jarry 1996, p.12)

11. Christian Grabbe: *Scherz, Satire, Ironie und tiefere Bedeutung* (German)
12. Gustave Kahn: *Le Conte de l'Or et Du Silence* (French)
13. Le Comte de Lautreamont: *Les Chants de Maldoror* (French)
14. Maurice Maeterlinck: *Aglavaine and Selysette*
15. Stephane Mallarme: *Verse and Prose* (French)
16. Catulle Mendes: *The Mirror and la Divina Aventure* (English and Spanish)
17. Homer: *The Odyssey*
18. Josephin Peladan: *Babylon* (EMPTY FILE)⁴
19. Francois Rabelais: *Gargantua and Pantagruel*
20. Jean de Chilra: *L'Heure Sexuelle* (EMPTY FILE)⁴
21. Henri de Regnier: *La Canne de Jaspe* (EMPTY FILE)⁴
22. Arthur Rimbaud: *Poesies Completes* (French)
23. Marcel Schwob: *Der Kinderkreuzzug* (German)
24. Alfred Jarry: *Ubu Roi* (French)
25. Paul Verlaine: *Poems*
26. Emile Verhaeren: *Poems*
27. Jules Verne: *A Journey to the Centre of the Earth*

The original list as it appears in “Faustroll” is shown below. Only three of the items have not been found as a resource. Some others have been approximated by using another text by the same author for example. Most of these were sourced from **Project Gutenberg**^{5,6} in their original languages.

1. BAUDELAIRE, a volume of E.A. POE translations.
2. BERGERAC, Works, volume II, containing the *History of the States and Empires of the Sun*, and the *History of Birds*.
3. *The Gospel according to SAINT LUKE*, in Greek.
4. BLOY, *The Ungrateful Beggar*.
5. COLERIDGE, *The Rime of the ancient Mariner*.
6. DARIEN, *The Thief*.
7. DESBORDES-VALMORE, *The Oath of the Little Men*.
8. ELSKAMP, *Illuminated Designs*.

⁴I have not been able to find any source texts online.

⁵See <https://www.gutenberg.org/>

⁶**A note on copyright:** Duration of copyright: §5. “For literary, dramatic, musical or artistic works 70 years from the end of the calendar year in which the last remaining author of the work dies.” (<https://www.copyrightservice.co.uk/ukcs/docs/edupack.pdf>) Maurice Maeterlinck and Marguerite Vallette-Eymery (a.k.a. Rachilde or Jean de Chilra) died less than 70 years ago and their work should still be under copyright. Alfred Jarry in the Simon Watson Taylor translation is a derivative work and is probably also still protected. (http://www.copyrightservice.co.uk/copyright/p22_derivative_works) Fair dealing: §7. “Private and research study purposes”, so for the purposes of this project copyright should not apply.

9. An odd volume of the *Plays* of FLORIAN.
10. An odd volume of *The Thousand and One Nights*, in the GALLAND translation.
11. GRABBE, *Scherz, Satire, Ironie und tiefere Bedeutung*, comedy in three acts.
12. KAHN, *The Tale of Gold and of Silence*.
13. LAUTREAMONT, *The Lays of Maldoror*.
14. MAETERLINCK, *Aglavaine and Selysette*.
15. MALLARME, *Verse and Prose*.
16. MENDES, *Gog*.
17. *The Odyssey*, Teubner's edition.
18. PELADAN, *Babylon*.
19. RABELAIS.
20. JEAN DE CHILRA, *The Sexual Hour*.
21. HENRI DE REGNIER, *The Jasper Cane*.
22. RIMBAUD, *The Illuminations*.
23. SCHWOB, *The Childrens' Crusade*.
24. Ubu Roi.
25. VERLAINE, *Wisdom*.
26. VERHAEREN, *The Hallucinated Landscapes*.
27. VERNE, *Voyage to the Center of the Earth*.

8.2 Setup

When the server is first started various setup functions are executed before any HTML is rendered. The search algorithms are triggered once a user enters a search term into the query field on any of the text, image or video pages.

Each plain text file in the corpus is added to the internal library one by one. Source 8.1 shows how this is done. The `PlaintextCorpusReader` is a feature of the `NLTK` Python library⁷ for Natural Language Processing.

```

1     library = PlaintextCorpusReader(corpus_root, '.*\.txt')
2     l_00 = library.words('00.faustroll.txt')
3     l_01 = library.words('01.poe1.txt')
4     ...
5     l_27 = library.words('27.verne.txt')

```

Source 8.1: Adding text files to the corpus library.

⁷<http://www.nltk.org/>



Figure 8.1: Toulouse-Lautrec's "Jane Avril"

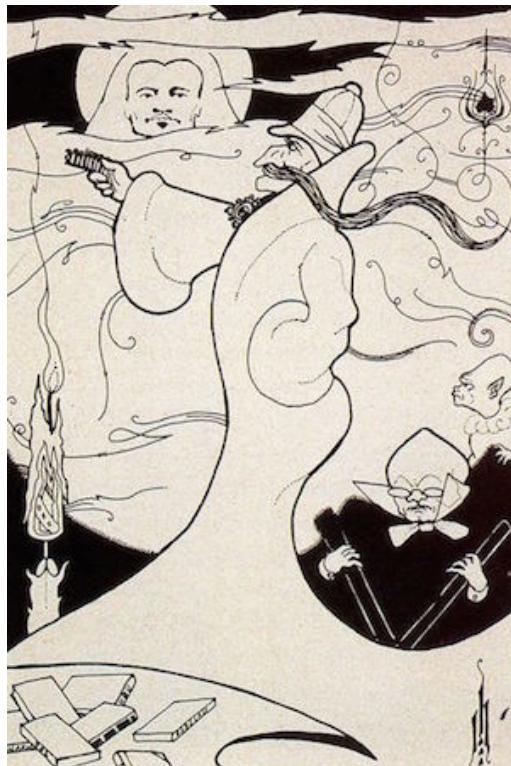


Figure 8.3: Aubrey Beardsley's "Docteur Faustroll"

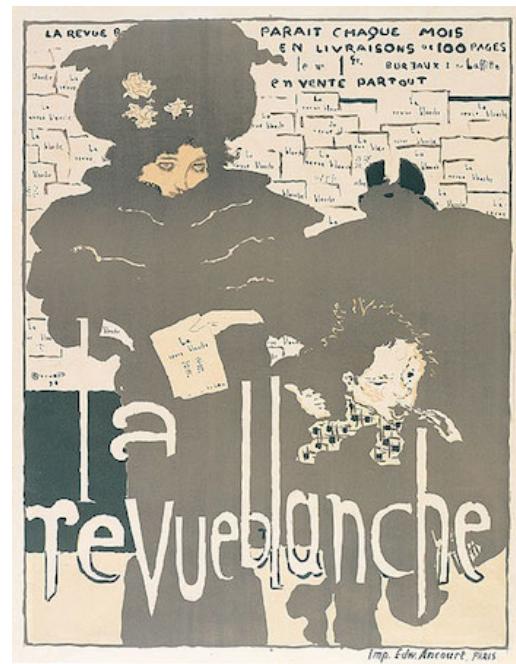


Figure 8.2: Bonnard's "Revue Blanche"



Figure 8.4: Oberthuer's "Saint Cado"

The `setupcorpus` function (see source 8.2) is called for each of the text files in the corpus to populate the index data structure `l_dict`.

```
l_dict = dictionary { dictionary { list [ ] } }
```

A dictionary in Python is what is known as an ‘associative array’ in other languages. Essentially they are unordered sets of **key: value** pairs. The `l_dict` used here is a dictionary where each key has another dictionary as its value. Each nested dictionary has a list as the value for each key.

```
1     # f = input text file variable
2     # l = stopwords file variable
3     def setupcorpus(f, l):
4         # x = counter/position
5         # w = word in file
6         for x, w in enumerate(f):
7             if w.isalpha() and (w.lower() not in l):
8                 y = '1_'
9                 y += (re.search(r"(\d\d).(\w)+.txt", f.fileid)).group(2)
10                l_dict[w.lower()][y].append(x)
```

Source 8.2: ‘setupcorpus’ function to process the corpus and create the index.

Line 6 in source 8.2 starts looping through file `f`. Line 7 checks if the current word `w` contains anything other than alphabetical characters and whether or not `w` is contained in the relevant stopword file `l` (for a list of english stopwords see appendix A). If both of those conditions are true variable `y` is created on line 8 (such as ‘1_00’ based on ‘00.faustroll.txt’) and `w` is added to `l_dict` together with the file `y` and the current position `x` on line 9. After all files are processed, the index looks like this:

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

Using one of the terms from figure 5.2 on page 72, here are their entries in the index file (the files are represented by their number in the `corpus` (see page 128), i.e. **1_00** is the ‘Faustroll’ file, **1_01** is the ‘Poe’ file, etc.). An excerpt from the actual `l_dict` can be found in the appendix A.

```
{
    doctor: {
```

```

    1_00: [253, 583, 604, 606, 644, 1318, 1471, 1858, 2334, 2431, 2446,
→ 3039, 4743, 5034, 5107, 5437, 5824, 6195, 6228, 6955, 7305, 7822, 7892,
→ 10049, 10629, 11055, 11457, 12059, 13978, 14570, 14850, 15063, 15099,
→ 15259, 15959, 16193, 16561, 16610, 17866, 19184, 19501, 19631, 21806,
→ 22570, 24867],
    1_01: [96659, 294479, 294556, 294648, 296748, 316773, 317841, 317854,
→ 317928, 317990, 318461, 332118, 338470, 340548, 341252, 383921, 384136,
→ 452830, 453015, 454044, 454160, 454421, 454596, 454712, 454796, 454846,
→ 455030, 455278, 455760, 455874, 456023, 456123, 456188, 456481, 456796,
→ 457106, 457653, 457714, 457823, 457894, 458571, 458918, 458998, 459654,
→ 459771, 490749],
    1_02: [11476, 12098, 28151, 36270],
    1_10: [53085, 53118, 53220, 53266, 53364, 53469, 53573, 53592, 53621,
→ 53718, 54873, 55262, 55525, 55577, 55614, 55683, 55741, 56058, 62709,
→ 113969, 114131, 114405, 114794],
    1_19: [14928, 15702, 49560, 82710, 167218, 180210, 189817, 189908,
→ 190020, 190235, 190905, 199430, 226663, 275454, 275928, 278097, 287375,
→ 291383, 304731, 306055, 324757, 330488],
    1_27: [16270, 79245]
},
...
}

```

8.3 Text

After the setup stage is completed and the webpage is fully loaded, user input in the form of a text query is required to trigger the three pataphysical algorithms.

Image and Video search do not use all three algorithms — where relevant this is highlighted in each section. Generally the following descriptions refer to the text search functionality.

Explain difference in Text, Image and Video

8.3.1 Clinamen

The clinamen is the unpredictable swerve that Bök calls “the smallest possible aberration that can make the greatest possible difference” ([Boek 2002](#)).

In simple terms, the clinamen algorithm works in two steps:

1. get clinamen words based on dameraulevenshtein and faustroll,
2. get sentences from corpus that match clinamen words.

find ref for dameraulevenshtein in baeza-yates book?

It uses the ‘faustroll’ text by Alfred Jarry (1996) as a base document and the Damerau-Levenshtein algorithm (Damerau 1964; Levenshtein 1966), which measures the distance between two strings (with 0 indicating equality), to find words that are similar but not quite the same. The distance is calculated using insertion, deletion, substitution of a single character, or transposition of two adjacent characters. This means that we are basically forcing the program to return matches that are of distance two or one, meaning they have two or one spelling errors in them.

```

1  # String w = query word
2  # Int i = assigned distance
3  def clinamen(w, i):
4      words = set([item for item in l_00 if dameraulevenshtein(w, item) <= i])
5      out, sources, total = get_results(words, 'Clinamen')
6      return out, words, sources, total

```

Source 8.3: Clinamen function

Source 8.3 line 4 creates the set of clinamen words using a list comprehension. It retrieves matches from the ‘faustroll’ file `l_00` with the condition that they are of Damerau-Levenshtein distance `i` or less to the query term `w` (see appendix A). Duplicates are removed. Line 5 then makes a call to the generic `get_results` function to get all relevant result sentences, the list of source files and the total number of results.

```

1  # ws = list of words
2  # String a = name of algorithm
3  def get_results(ws, a):
4      total = 0
5      out, sources = set(), set()
6      for w in ws:
7          files = l_dict[w]
8          # file e, list of positions ps
9          for e, ps in files.items():
10             f = get_title(e)
11             sources.add(f)
12             sent = pp_sent(w.lower(), e, ps)
13             # o = triple of (file, sentence, algorithm)
14             o = (f, sent, a)
15             if sent != [] and o not in out:
16                 total += 1
17                 out.add(o)
18     return out, sources, total

```

Source 8.4: ‘get_results’ function to get all sentences for a list of words.

The `get_results` function (see source 8.4) is used by all three algorithms (climamen, syzygy and antinomy). Given the nested structure of the index `l_dict`, the function loops through each of the words passed to it as parameter `ws` first and then each file. Line 7 retrieves the dictionary of files from `l_dict`. Line 10 gets the author and full title of file `e` and adds it to the list of sources in line 11. Line 12 makes use of yet another function called `pp_sent` (see source 8.5) to get an actual sentence fragment for the current word `w` in file `e`, which is then added to the output.

```

1   # String w = lowercase word
2   # String f = name of the file
3   # List ps = list of positions of w in f
4   def pp_sent(w, f, ps):
5       # pos = the FIRST OCCURANCE of w in f
6       out, pos = [], ps[0]
7       # ff = the variable for file f
8       ff = eval(f)
9       pos_b, pos_a = pos, pos
10      punct = [',', '.', '!', '?', '(', ')', ':', ';', '\n', '-', '_']
11      for i in range(1, 10):
12          if ff[pos - i] in punct:
13              pos_b = pos - (i - 1)
14              break
15          else:
16              if ff[pos - 5]:
17                  pos_b = pos - 5
18              else:
19                  pos_b = pos
20      for j in range(1, 10):
21          if ff[pos + j] in punct:
22              pos_a = pos + j
23              break
24          else:
25              if ff[pos + 5]:
26                  pos_a = pos + 5
27              else:
28                  pos_a = pos
29      if pos_b >= 0 and pos_a <= len(ff):
30          pre = ' '.join(ff[pos_b:pos])
31          post = ' '.join(ff[pos+1:pos_a])
32          out = (pre, w, post)
33      return out

```

Source 8.5: ‘`pp_sent`’ function to retrieve a sentence from a file.

In function `pp_sent` (source 8.5) line 6 is important to note because it is a key functionality point. Even though the index `l_dict` stores a full list of all possible

positions of a given word in each file, the `pp_sent` function **only retrieves the sentence of the very first occurrence of the word** rather than each one. This decision was taken to avoid overcrowding of results for the same keyword.

Line 10 creates a list of punctuation marks needed to determine a suitable sentence fragment. Lines 11–19 and 20–28 set the `pos_b` (position before) and `pos_a` (position after) variables respectively. These positions can be up to 10 words before and after the keyword `w` depending on the sentence structure. In line 30 the actual sentence fragment up to the keyword is retrieved, while in line 31 the fragment just after the keyword is retrieved. `ff[pos_b:pos]` for example returns the list of words from position `pos_b` to position `pos` from file `ff`. The built-in Python `.join()` function then concatenates these words into one long string separated by spaces. On line 32 a triple containing the pre-sentence, keyword and post-sentence is set as the output and then returned.

The image/video searches don't use the clinamen function at all.

8.3.2 Syzygy

The syzygy surprises and confuses. It originally comes from astronomy and denotes the alignment of three celestial bodies in a straight line. In a pataphysical context it is the pun. It usually describes a conjunction of things, something unexpected and surprising. Unlike serendipity, a simple chance encounter, the syzygy has a more scientific purpose.

In simple terms, the syzygy algorithm works in two steps:

1. get syzygy words based on synsets and hypo-, hyper- and holonyms from WordNet,
2. get sentences from corpus that match syzygy words.

```
1  # w = input query term
2  def syzygy(w):
3      words = set()
4      wordsets = wn.synsets(w)
5      for ws in wordsets:
6          words.update(get_nym('hypo', ws))
7          words.update(get_nym('hyper', ws))
8          words.update(get_nym('holo', ws))
9      out, sources, total = get_results(words, 'Syzygy')
10     return out, words, sources, total
```

Source 8.6: Syzygy function.

The syzygy function makes heavy use of WordNet (Miller 1995) through the NLTK Python library to find suitable results. Specifically, as shown in source 8.6, the algorithm fetches the set of synonyms (synsets) on line 4. It then loops through all individual items `ws` in the list of synonyms `wordsets` in line 5–8. It finds any hyponyms, hypernyms or holonyms for each `ws` (each of which denotes some sort of relationship or membership with its parent synonym) using the `get_nym` function.

explain reasoning behind algorithms like this for all:

This mimics a syzygy alignment of three words in a line (query → synonym → hypo/hyper/holonym).

Line 9 makes use of the `get_results` function (see source 8.4) in the same was as the clinamen function does.

rewrite getnzym function to automatically get all three without the ifs

The image and video searches both use the syzygy function as part of their `pataphysicalise` function (see source 8.8).

8.3.3 Antinomy

The antimony, in a pataphysical sense, is the mutually incompatible.

In simple terms, the antinomy algorithm works in two steps:

1. get antinomy words based on synsets and antonyms from WordNet,
2. get sentences from corpus that match antinomy words.

```
1  # w = input query term
2  def antinomy(w):
3      words = set()
4      wordsets = wn.synsets(w)
5      for ws in wordsets:
6          anti = ws.lemmas()[0].antonyms()
7          if len(anti) > 0:
8              for a in anti:
9                  if str(a.name()) != w:
10                     words.add(str(a.name()))
11
12 out, sources, total = get_results(words, 'Antinomy')
13 return out, words, total
```

Source 8.7: Antinomy function.

For the antinomy we simply used WordNet's antonyms (opposites) (see source 8.7). This algorithm is very similar to the algorithm for the syzygy. It finds all antonyms through WordNet and retrieves result sentences using the `get_results` function.

8.4 Image & Video

In simple terms, the image and video search works in three steps:

1. pataphysicalise query terms using syzygy algorithm
2. translate each pataphysicalised term
3. retrieve images/videos using API calls

The `pataphysicalise` function (see source 8.8) transforms the original query terms ready for the next step. In line 5 the `syzygy` algorithm (source 8.6) is used to make this transformation. Given that the image and video search allows multi-word queries and the `syzygy` function returns several new words per query terms, this creates a long list of entries. On top of that the output is the inner product (line 8) of all these results. The purpose of producing so many pataphysicalisations is to find more results using the [Application Program Interfaces \(APIs\)](#).

```
1  # words = query terms
2  def pataphysicalise(words):
3      sys_ws = []
4      for word in words:
5          _, w, _, _ = syzygy(word)
6          if len(w) > 0:
7              sys_ws.append(list(w))
8      out = itertools.product(*sys_ws)
9      return list(out)
```

Source 8.8: Function to pataphysicalise image and video query terms.

For example, running the pataphysicalise function with the terms 'clear' and 'sky' will produce two intermediary lists (shortened here for the demonstration) which are then combined into one list using the Cartesian product:

```
[ "disembarrass", "bear", "judge", "remove", "elucidate", "modify", "free",
  ↪ "approve", "certify", "determine", "strip", "empty", "purge", "vanish",
  ↪ "disappear", "sell", "pay", "make", "take", "disforest", "formalize",
  ↪ "okay", "allow", ... ],
```

```

["blue", "atmosphere", "fling", "throw_back", "lag", "blue_sky",
→ "submarine", "toss_back", "blue_air", "mackerel_sky", "wild_blue_yonder"]

[("disembarrass", "blue"), ("disembarrass", "atmosphere"), ..., ("strip",
→ "fling"), ..., ("empty", "submarine"), ..., ("allow", "mackerel_sky"),
→ ("allow", "wild_blue_yonder")]

```

The next step is to translate the pataphysicalised search terms as shown in source [8.9](#) before any [API](#) calls are made.

```

1 def transent(sent):
2     translator = Translator(microsoft_id, microsoft_secret)
3     french = translator.translate(sent, "fr")
4     japanese = translator.translate(french, "ja")
5     patawords = translator.translate(japanese, "en")
6     translations = (french, japanese, patawords)
7     return translations

```

Source 8.9: Translation function.

8.4.1 REST & API

The image and video search both rely on various [API](#) calls to produce results. Currently used are Microsoft Translate, Bing Image Search and YouTube.

A [RESTful API](#) allows browsers ('clients') to communicate with a web server via [HTTP](#) methods such as GET and POST.

For this project I used the Python library **Requests**⁸.

<https://api.datamarket.azure.com/Bing/Search/Image?format=json&Query='kittens'>

```

"d": { "results": [
  { "__metadata":
    { "uri":
      → "https://api.datamarket.azure.com/Data.ashx/Bing/Search/Image?Query=\u0027kittens\u0027&$skip=0&$top=10",
        "type": "ImageResult"
      }, // __metadata
      "ID": "e09072a2-faf3-47ac-b77d-46a8df8941aa",
      "Title": "Cute Kittens - Pictures - The Wondrous Pics",
      "MediaUrl":
      → "http://wondrouspics.com/wp-content/uploads/2011/12/Cute-Kitten2.jpg",
        "SourceUrl": "http://wondrouspics.com/cute-kittens-pictures/",
        "DisplayUrl": "wondrouspics.com/cute-kittens-pictures",
        "Width": "1440",

```

⁸<http://docs.python-requests.org/en/latest/>

```

1 def get_Bing(words):
2     out = []
3     trans = ''
4     bing_key = 'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx'
5     base = "https://api.datamarket.azure.com/Bing/Search/"
6     params = "Image?format=json&Query="
7     after = ''
8     for x in words:
9         y = ' '.join(x)
10        z = transient(y) # (french, japanese, patawords)
11        url = ''.join([base, params, z[2], after])
12        bing_img = requests.get(url, auth=HTTPBasicAuth(None, bing_key))
13        if bing_img.json()['d']['results']:
14            trans = z
15            for result in bing_img.json()['d']['results']:
16                phototitle = result['Title']
17                photoimg = result['MediaUrl']
18                photolink = result['SourceUrl']
19                out.append((phototitle, photoimg, photolink))
20            break
21        else:
22            out = []
23    return out, trans

```

Source 8.10: Using the Microsoft Bing API to retrieve images.

```

    "Height": "900",
    "FileSize": "238015",
    "ContentType": "image/jpeg",
    "Thumbnail":
    { "__metadata":
        { "type": "Bing.Thumbnail"
        },
        "MediaUrl":
        ← "http://ts2.mm.bing.net/th?id=OIP.M5692e5d79242507e30600fd54639316cH0&pid=15.1",
            "ContentType": "image/jpg",
            "Width": "480",
            "Height": "300",
            "FileSize": "13856"
        } // Thumbnail
    }, ...
], // results
"__next":
← "https://api.datamarket.azure.com/Data.ashx/Bing/Search/Image?Query=%u0027kittens%u0027&$skip=10&$top=10"
} // d

```

8.5 Design

Once the three algorithms have produced their respective results, the page displaying these results can be rendered. This is done using the templating language Jinja and HTML (with CSS stylesheets and some JavaScript).

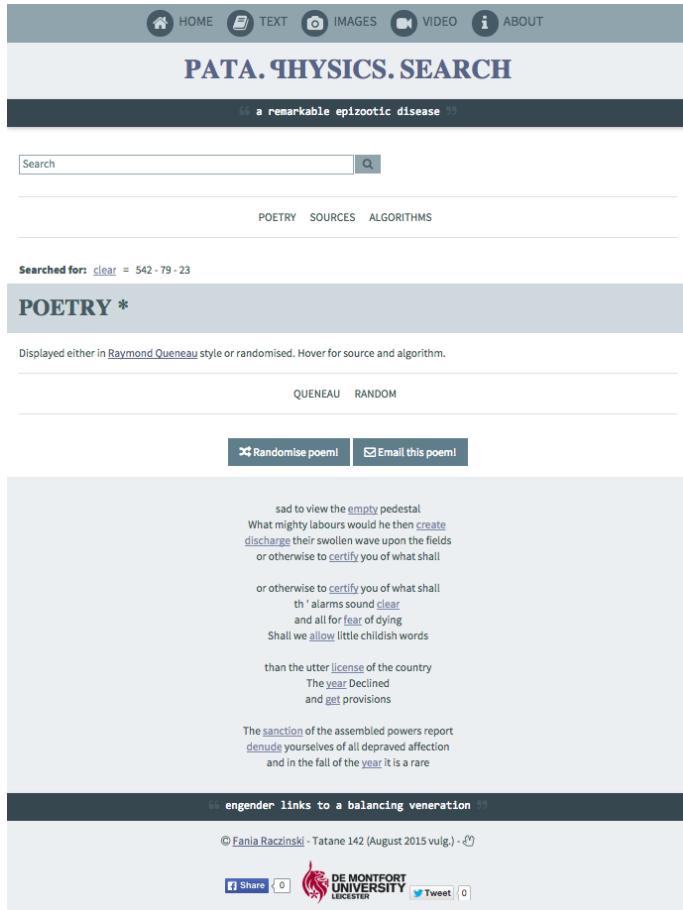


Figure 8.5: Poetry results screenshot

The text results page has three options for how the results are presented, with 'Poetry — Queneau' being the default.

Poetry

Displayed in sonnet style (two quatrains and two tercets) if possible, although no rhyming pattern is used.⁹

- Queneau — Each line can be changed manually.
- Random — The whole poem can be randomised.

⁹<https://en.wikipedia.org/wiki/Sonnet>

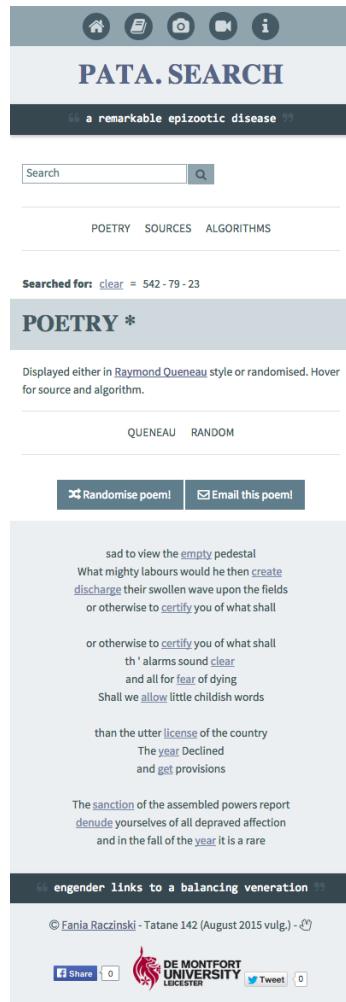


Figure 8.6: Poetry results screenshot for mobile screens

Sources

Ordered by source text.

Algorithms

Ordered by algorithm.

get proper ref for sonnet style

The image and video results pages work the same way. They both have two display options, with the 'Spiral' option being the default.

Spiral

Displayed square images/videos as a golden spiral.

List Displayed as a simple list.

The image and video spirals are modelled on the Fibonacci spiral ¹⁰.

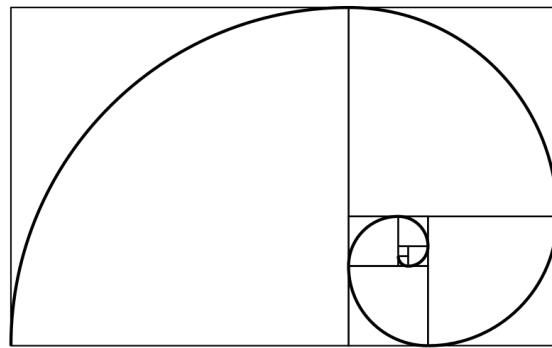


Figure 8.7: Fibonacci Spiral

```
1  <div class="subtab_content" id="q_tab">
2      <p class="w3-center">
3          <a class="emailbutton w3-btn w3-blue-grey" href="#" onclick="return
4              getContent(this)">
5                  Email this poem!
6              </a>
7          </p>
8          <div class="poetry w3-container w3-theme-15">
9              {% for n in range(1, lol|length + 1) %}
10             {% set wid = ['wn', n|string]|join %}
11             {% set lid = ['lyr', n|string]|join %}
12             {% set sid = ['scrollLinks', n|string]|join %}
13             {% set aid = lol[n-1] %}
14             <div id="poems">
15                 <div id="{{wid}}" class="wn">
16                     <div id="{{lid}}" class="lyr">
17                         {% for sens in aid %}<span title="{{ sens[0] }}, {{ sens[2]
18                             }}>{{ sens[1][0] }} <form class="inform" action="../textresults"
19                             method="post"><input class="inlink" type="submit" name="query" value="{{
20                             sens[1][1] }}" onclick="loading();"></input></form> {{ sens[1][2]
21                             }}</span>{% endfor %}
22                         </div>
23                     </div>
24                     <div id="{{sid}}" class="scrollLinks"></div>
25                 </div>
26             {% endfor %}
27         </div>
28     </div>
```

Source 8.11: Code for rendering Queneau style poems.

¹⁰Image taken from https://upload.wikimedia.org/wikipedia/commons/9/93/Fibonacci_spiral_34.svg

Source 8.11 shows the segment of HTML/Jinja code that renders the Queneau Poetry. Lines 2-6 creates a button for sending the currently showing poem per email. Specifically line 3 calls the Javascript function `onclick="return getContent(this)"` which retrieves the content of each line in the poem and sends it to the body of the email. Lines 7-22 render the 4 stanzas of the poem. This is done using two nested Jinja 'for' loops (line 8 and line 16). Line 8 loops through the (ideally) 14 lines of the poem. `1o1` can be considered a masterlist of all sublists for each poem line.

```
# all_sens list:
[
    (title, (pre, word, post), algorithm),
    ...
]
```

8.6 Prototypes

8.6.1 Prototype 2

This version introduced the move from Django to Flask. It also included the first major re-design of the website. It is still available online at pata.fania.eu.



Figure 8.8: Prototype 2 screenshot

The main differences between the current version and prototype 2 are:

- the corpus consisted of the faustroll text only

```

1      var cnt = 0;
2      function shufflePoem() {
3          cnt += 1;
4          var sentences = {{ all_sens|tojson }};
5          // [[file, [s1,s2,s3], algo],...]
6          var n = {{ all_sens|length }};
7          var rlist = [];
8          for (var i = 0; i < 14; i++) {
9              var r = Math.floor(Math.random() * n);
10             var t = sentences[r][0];
11             var al = sentences[r][2];
12             var b = sentences[r][1][0];
13             var m = sentences[r][1][1];
14             var a = sentences[r][1][2];
15             var str1 = "<span title='>" + t + ', ' + al;
16             var str2 = "'>" + b + " <form class='inform' action='../textresults' "
17             ← method='post'><input class='inlink' type='submit' name='query' value='";
18             var str3 = m + "' onclick='loading();'></input></form> " + a;
19             var str4 = "</span>";
20             var fullsent = str1 + str2 + str3 + str4;
21             rlist[i] = fullsent;
22         }
23         rlist[3] = rlist[3].concat('<br>');
24         rlist[7] = rlist[7].concat('<br>');
25         rlist[10] = rlist[10].concat('<br>');
26         var output = rlist.join('<br>');
27         document.getElementById('clickcount').innerHTML = cnt;
28         document.getElementById('random_poem').innerHTML = output;
29         return false;
}

```

Source 8.12: Code for randomising poems.

- results were keyword ± 5 words per line
- text results were displayed sorted by algorithm only
- image and video results were displayed as spiral only

8.6.2 Prototype 1

get new screenshots for prototype 1

don't mention James?

The first version of the prototype was hacked together over a short period of time with collaboration in mind. It was originally build to demonstrate the three algorithms in action before James' architecture was finished. The design of the website was simple and plain.

Pataphysical Search!

Clinamen - 19 pataphysicalised reverberations found for: "clear"

...s webbed feet . Pitiful **pleas** swim up against the stream...
 ...was impossible to enter the **cellar** due to the flooding thereof...
 ...the length of a cabbage **leaf** , paying no attention to...
 ...with Mendacious in the **lead** . Since the episcopal nature...
 ...gigantic , black , mass **altar** . At the blunt point...
 ...blue diadem , He was **clad** , too , in sky...
 ...TO ARTICLE B19 In this **year** Eighteen Hundred and Ninety -...
 ...route ; but such a **leap** is not within everyone '...
 ...any need for Faustroll to **fear** that his scalp - hair...
 ...which followed us and those **near** us which crossed our path...
 ...maple , oak , **cedar** , sorb wood and poplar...
 ...yellow skin , his face **clean** - shaven , apart from...
 ...exclaims : Never , I **swear** , shall I forget the...
 ...the pale forehead , the **dear** face , this terrible little...
 ...LAURENT TAILHADE We could already **hear** bells - as loud as...
 ...content to be black . **Fear** , turning away its head...
 ...Nage 's right **ear** and four of his teeth...
 ...doctor informed me , discern **clearly** through these mirrors those ultraviolet...
 ...WITCH Her hump to the **rear** , belly to the fore....

Syzygy - 22 pataphysicalised reverberations found for: "clear"

...French language , he could **pronounce** fairly correctly a few words...
 ...as his tonsure , laying **bare** the optic nerve and the...
 ...few quarter - centuries will **determine** their periods . Soon ,...
 ...allows air and steam to **pass** through but is impermeable to...
 ...zero , if these dimensions **vanish** on both sides of our...
 ...the Mayor , who did **certificate** the original thereto ; within...
 ...four hours , to **pay** to the claimant into my...
 ...choice of the two asphyxiating **make** - ups called white hanged...
 ...hereunder . The sale will take place on whatever day shall...
 ...web , leaves the holes **empty** - the number of which...
 ...guide had given him absolutely **free** ; one represented realistically ...
 ...a day ... serene **countenance** ... supreme image , so...
 ...usual example of water , let us reflect , in this...
 ...our dead drunk credits and **gain** , without wasting our talent...
 ...as far as I could **judge** , understood these prodigies very...
 ...examine any disturbances which this **change** in size might involve in...
 ... , except perhaps in the **country** , he will rarely see...
 ...the Snout , to the **clear** anxiety of those present
 ...found by experiment that the **benefit** extends only to those whose...
 ...meshes are wide enough to **allow** the passage of a large...
 ...the globe by attraction , **permit** me , I pray
 ... And I 'll **declare** He 's mooning up...

Antinomy - 1 pataphysicalised reverberations found for: "clear"

...colors were locked in an **opaque** box ; until he was...

This prototype was written in python 2.7 and the website using django 1.4.
 It uses the NLTK natural language processing library.
 Fania Raczkowski - De Montfort University - Palatin, 139 (May 2012 vulg.)

Figure 8.9: Prototype 1 screenshot

The main differences between prototype 1 and prototype 2 are:

- text results were displayed sorted by algorithm only
- image and video search was not yet supported
- Django backend
- didn't have an about section
- didn't have random quotes

Chapter 9

Impact and Applications — Case Study

9.1	Patakosmos	149
9.2	Soeren and the other guy	149
9.3	Digital Opera	149
9.3.1	Use	149
9.3.2	Result	149
9.3.3	Interview	150

discuss problems with algorithms, pros and cons...

This function exhibits the same problem as mentioned above for the syzygy, just much worse. Arguably, some words just do not appear to have an opposite, but the pataphysical antinomy should still be able to find a match. A better thesaurus or a larger index (e.g. based on more than one book — or, of course, the Web) could improve this method.

In this section we consider the possible uses and applications for the proposed creative search tool.

Our target audience is not quite as broad as that of a general search engine like Google. Instead, we aim to specifically cater for users who can appreciate creativity or users in need of creative inspiration. Users should generally be educated about the purpose of the search tool so that are not discouraged by

what might appear to be nonsensical results. Users could include artists, writers or poets but equally anybody who is looking for out-of-the-box inspirations or simply a refreshingly different search engine to the standard. The way we display and label results produced by the tool can influence how the user perceives them. The current prototype for example separates the results into its three components but we could have equally just mixed them all together. The less transparent the processes in the background (e.g. which algorithm was used, how does the result relate to the query precisely, etc.) are for the user, the more difficult it might be to appreciate the search.

There are many ways a pataphysical search tool could be used across disciplines. In literature, for example, it could be used to write or generate poetry, either practically or as a simple aid for inspiration. We are not limited to poetry either; novels, librettos or plays could benefit from such pataphysicalised inspirations. One can imagine tools using this technology that let you explore books in a different ordering of sentences (a sort of pataphysical journey of paragraph hopping), tools that re-write poems or mix and match them together. Even our simple prototype shows potential in this area and could be even more powerful if we extended it to include more base texts, for example the whole set of books contained in Faustroll's library ([20] and also [12]). A richer body of texts (by different authors) would produce a larger index which would possibly find many more matches through WordNet and end in a more varied list of results.

From a computer science perspective it could be used as one of the many algorithms used by traditional search engines for purposes like query feedback or expansion (e.g. "did you mean ... " or "you might also be interested in ... "). Depending on how creative we want the search engine to be, the higher we would rank the importance of this particular algorithm. One of the concepts related to the search tool, namely patadata, could have an impact on the development of the Semantic Web. Just as the Semantic Web is about organizing information semantically through objective metadata, patadata could be used to organize information pataphysically in a subjective way.

The prototype tool is already being used in the creation of an online opera, provisionally entitled from [place] to [place], created in collaboration with The Opera Group, an award-winning, nationally and internationally renowned opera company, specialising in commissioning and producing new operas. In particular, it is being used to create the libretto for one of the virtual islands whose navigation provides the central storyline for the opera. The opera will premiere in 2013, and will continue to develop thereafter, deploying new versions of the tool as they appear.

9.1 Patakosmos

www.patakosmos.com

9.2 Soeren and the other guy

9.3 Digital Opera

9.3.1 Use

“There is an official and an unofficial way that I used the prototype. Officially, I threw keywords based on mood ‘sad’, ‘lively’ etc into it and used the results as the libretto for small sections of music that reflect said mood. Unofficially I used lots and lots of different words to retrieve the lines that worked.” Lee Scott (22 May 2014)

9.3.2 Result

Confusing

...my tuning fork. imagine the perplexity of a man outside time...

...mandrills or clowns, spread their caudal fins out wide like acrobats...

...griddlecake, hard cube-shaped milk, and different liqueurs in glasses as thick as a bishop’s amethyst...

Playful

...peacocks’ tails, gave us a display of dancing on the glassy...

Busy

...wasps and bumblebees and the vibration of a fly’s wing...

Driving

...bodies striking the hours of union and division of the black...

Disjointed

...tangential point of the universe, distorting it according to the sphere’s...

Sadness

...others: may your dire sorrow flyaway...

...no longer deep enough to satisfy our honour...

...other side of the green sleep of hulls; ships passed away...

Sweeping

...loved her like the infinite series of numbers...

...the veritable portrait of three persons of god in three escutcheons...

Fear

...it will set. fear creates silence nothing is terrifying...
...forth revealing the distinction and evil engraved in the wood...
... underground arose from ali baba screaming in the pitiless oil...

Joy

...sibyls record the formula of happiness, which is double: be amorous...
...the lord of the island gloried that his creation was good...

Awe

...like earth; the enemy of fire and renascent from it...
...awesome figure, warlike and sacerdotal, glared at the assembly...
... is not an island but a man...

Clocked

...quincuncial trees...

Tension

...the vigilant gaze of the spirit of the dead...
...do not make as much noise as a single drum...
... the oars made a clangourous sound as they scraped along the bow....

Calm

...a strange upon a clam sea quilted with sand; faustroll...
...each person present threw a pebble into the sea...
...depth and with edges that tend to ebb and flow...
(textures)

Morphing

...in a striking metamorphosis the mourning color of the hangings turned...

9.3.3 Interview

Part IV

M^ΣT[∀]-LOGIC[∀]LYSIS

Chapter 10

Interpretation

10.1 Measurable Attributes	154
10.2 Problems with Evaluation	156
10.3 5 P's: product, process, people, place and purpose	158

Part of this research has been described in a journal article in Digital Creativity in 2013, and I presented a paper at the Creativity and Cognition conference 2013 in Sydney.

Creativity should be seen as a continuum, there is no clear cut-off point or Boolean answer to say precisely when a person or piece of software has become creative or not.

ref

Evaluating creative software is not an easy task. Pease and Colton [27] divide it into two notions:

- whether an idea or artefact is valuable or not, and
- whether a system is acting creatively or not.

We would need to investigate each individual search result in terms of its value and creativity. This could be done by user ratings or satisfaction questionnaires. Rather than measuring the success of individual results we could look at evaluating them as one set instead, similar to the blind side-by-side comparisons by Bing or MillionShort.

The search results produced by our tool can be quite surprising sometimes and

it not always clear how they connect to the initial query (especially if the inner workings of the algorithms are unknown), even if we identify through which function a result has been obtained. Obviously these keywords might not be helpful to users unfamiliar with the concept of pataphysics and might therefore appear rather nonsensical. Whilst there is a clear logic to each search result, they might appear anomalous to the user's expectations if he received these results without knowing the philosophy of the search tool. The results could possibly appear random then, and would therefore likely to be detrimental to the user.

To prevent that, the level of interaction the user has with the system and the feedback the system gives to the user on decisions it is making will have a large influence on the overall effectiveness and appreciation of the tool. A quick and simple solution to this problem would be to add an icon to the side of each search result, which displays exactly how the original query was pataphysicalised.

Creativity does not have a universally accepted definition. Creativity is a human quality and definitions don't necessarily lend themselves to be applied to computers as well. There are aspects that come up in many, like novelty and value, but some that rarely pop up, like relevance and variety. Creativity can be studied at various **levels** (neurological, cognitive, and holistic/systemic), from different **perspectives** (subjective and objective) and **characteristics** (combinational, exploratory and transformative). Creativity should be seen as a continuum, there is no clear cut-off point or Boolean answer to say precisely when a person or piece of software has become creative or not.

Current evaluation methodologies have concentrated on only a handful of the points raised above, for example studying only the creative end-product itself (out of context), only judging it by its novelty objectively, assigning an arbitrary thresholds, etc. This also includes the assumption that machines "only" mimic humans and are therefore not judged by full potential.

What does it mean, how can it be measured?

Subjectivity vs objectivity is a theme throughout

How is it defined and measured in humans, what can we just take directly from those concepts and apply them directly to machines and what needs to be completely redefined?

This paper discusses issues related to the study of creativity in a computer science context. Two transdisciplinary fields of study have emerged from the variety of disciplines concerned. These are computational creativity and creative

computing. The former lies at the cross section of artificial intelligence and cognitive science and the latter is mostly distinguished by its involvement in art. Creative computing focuses on the process of creativity rather than just the outcome as in computational creativity. In the remainder of this paper, CC will always denote creative computing unless otherwise specified.

Many of these (if not all) spawn from the computational creativity discipline.

Introduce the difference between human and computer evaluation/creativity?

In research communities, approaches to the study of creativity differ in three main respects: 1) the type of research design, whether experimental, psychometric, observational etc. 2) the focus of the research, whether on human attributes cognitive processes or features of creative outcomes, and 3) the type of information that is used for the basis of evidence, by which is meant whether the time frame is present (real-time observation) or past (historical data) and whether the situation is artificial (laboratory) or natural (real world settings). (Candy 2012, p.3)

distinguishing between person's and product's creativity (Piffer 2012, p.258) it is concluded that a person's creativity can only be assessed indirectly (for example with self report questionnaires or official external recognition) but it cannot be measured (Piffer 2012, p.258)

10.1 Measurable Attributes

Novelty	Value	Quality	Ephemeral Uncontrolled	Temporal Controlled	Purpose
Originality	Usefulness	Skill	Serendipity	Persistence	Intention
Newness	Appropriateness	Efficiency	Randomness	Results	Communication
Variety	Appreciation	Competence	Uncertainty	Development	Evaluation
Typicality	Relevance	Intellect		Progression	Aim
Imagination	Impact	Acceptability		Spontaneity	
	Influence		Experimentation		Independence

Table 10.1: Summary of all creativity attributes

Summary

- Mimicking
- novelty + value + quality
- randomness + serendipity

10.2 Problems with Evaluation

Evaluating **human creativity** is problematic.

There are many debates across the involved disciplines. Specifically, Mayer identified five big questions of human creativity research: (Mayer 1999, p.450-451)

1. Is creativity a property of people, products, or processes?
2. Is creativity a personal or social phenomenon?
3. Is creativity common or rare?
4. Is creativity domain-general or domain-specific?
5. Is creativity quantitative or qualitative?

An important challenge for the next 50 years of creativity research is to develop a clearer definition of creativity and to use a combination of research methodologies that will move the field from speculation to specification. (Mayer 1999, p.459)

Taking these debates about human creativity and directly applying them to machines seems logical but may be the wrong and lazy approach. Adapting Mayer's five big questions to machines does not seem to capture the real issues at play.

1. Is creativity a property of programmers, users, machines, products, or processes?
 2. Is creativity a local, a network or an Internet phenomenon?
 3. Is creativity common or rare? (P or H creativity)
 4. Is creativity domain-general or domain-specific?
 5. Is creativity quantitative or qualitative?
- Can a machine judge whether a human is creative?
 - Is creativity a property of machines (hardware or software?)

- Is mimicking human creativity really enough and appropriate?
- Compare against “human creativity”? Or define machine creativity from scratch?
- Who is creative? The programmer or the program?
- Can creativity be objectively measured?
- Quantitative or qualitative?
- In respect to P or H creativity?
- Output minus input? (we don't have the same strict judgement on humans)
- Is it the product or the process or both?
- Does context matter? (Blind deaf dumb person = computer?)
- Does time matter?
- Does purpose or intention matter?
- AGI vs AI? Artificial general creativity vs artificial creativity?

On a more practical level, there are various problems that arise when trying to evaluate computer creativity. Anna Jordanous found that “evaluation of computational creativity is not being performed in a systematic or standard way” (p.2 **Jordanous 2011**, her emphasis).

(neurological, cognitive and systemic) in the computer sense!

Since most problems with evaluating creativity in computers (and humans alike) stems from the lack of a universal definition it seems logical to try and remedy this first and foremost.

Creativity is studied in many disciplines.

- understanding the chemical mechanisms within the brain (neurological)
- understanding the thought processes associated with creativity (cognitive)
- understanding creativity in children and adults, novices and professionals
- understanding creativity in individuals and society (holistic)

Creativity is a continuum, which means that being creative and not being creative form the two distinct extreme ends of a scale.

"a continuous sequence in which adjacent elements are not perceptibly different from each other, but the extremes are quite distinct"
[OED]

(subjective and objective)

How can we model Koestler's bisociative creativity in computers? Boden/Kaufman: Subjective and objective types (pandh or little-candbig-C) (Boden 2003; Kaufman and Beghetto 2009) (product+process)

DIGITAL CREATIVITY ?!?!?! Mix between digital humanities and creative computing/computational creativity --- see Digital Creativity Journal!!!! Unified theory of creativity! (Koestler?) Unified definition!

1. What is the definition of creativity?
2. Who is being creative? WHO
3. What was the aim/intention, if any?
4. What was the process, approach? HOW
5. What factors influenced the person/process? WHERE
6. What is the result/product, if any? Is it original, relevant? WHAT
7. What is the impact, if any?
8. What is the maintenance plan, if any?

10.3 5 P's: product, process, people, place and purpose

One way of characterizing these processes is to use an alliteration that allows us to keep track of some of the core features of RRI in ICT, namely the four P's, which are: product, process, purpose and people. The purpose of using the four P's is to draw attention to the fact that, in addition to the widely recognized importance of both product and process of technical development, the purpose of the development needs to be considered and people involved in the innovation need to be incorporated in RRI. (Stahl, Jirotka and G. Eden 2013, p.203, my emphasis)

combine the 4 P's with purpose// 5 P's: product, process, people, place and purpose// Why is the purpose important?// Interpreting or Measuring?// Maybe we should not be looking for metrics but rather guidelines for interpretations of creativity.

In the end I believe it is impossible to measure creativity objectively. I don't just think it is impossible, I think it is unwise to try and do so. It would be silly to put a percentage on how creative something is just like it would be silly to say a certain product is 50percent ethical. In fact there are lots of parallels between (computer) ethics and (computer) creativity. Both are subjective, both are highly dependent on context.

What is important is to study and consider the factors that influence our perception of whether something is creative (or ethical) and what the implications are.

Creativity in a process or product will mean different things to different people, in different environments and contexts. Common sense.

Just as there are people who just cannot see any creativity in modern art, there will always be people who won't accept anything produced by a computer as creative.

Chapter 11

Patacritical Analysis

11.1 Problems Encountered	160
11.2 Design Aspects	160
11.3 Search Results	160

11.1 Problems Encountered

11.2 Design Aspects

11.3 Search Results

Part V

HAPPILY ΣVΣR ∨FTΣR?

Chapter 12

Aspirations

12.1 Code	162
12.2 Interface	163
12.3 Algorithms	163
12.4 Architecture	163
12.5 Research	163

PROBLEMS ENCOUNTERED AND SUGGESTED SOLUTIONS

SHORTCOMINGS AND MISSING FUNCTIONALITY

- Research in science and art
- Review paper? Pataphysics and creativity?
- Quantitative research questions
- Working definition for Pataphysics
- Examples for Pataphysics concepts
- Examples for types of creativity
- Examples for creative process
- Explain Leary's tables
- How can we use creative concepts discussed?

12.1 Code

FURTHER DEBUGGING OF CODE (IF NECESSARY)

12.2 Interface

DESIGN ASPECTS

IMPROVEMENTS / ALTERNATIVES TO USER INTERFACE DESIGN

12.3 Algorithms

IMPROVEMENTS / ALTERNATIVES TO ALGORITHMS

12.4 Architecture

IMPROVEMENTS / ALTERNATIVES TO ARCHITECTURE

12.5 Research

USER FEEDBACK (IF NECESSARY)

Chapter 13

Observations

13.1 Achievements	165
13.2 Implications	165
13.3 Recommendations	165

From here, we can try to implement different algorithms or different pataphysical concepts within our existing tool or built a different system. The next logical step would be to implement a fully functioning Web search engine using the algorithms described in this paper. But before we go into further development, it might be worth evaluating and interpreting the results produced by the prototype.

In this paper we have introduced a new approach for a creative search tool that uses pataphysics as an underlying philosophy. We have explained how pataphysics can be used in search algorithms to produce interesting results with a humorous twist. Our initial experiments within a limited domain have shown that the generated results can indeed be novel, surprising and useful. We have also briefly discussed ideas for applications of the tool and issues that may trigger possible further research in the field of Computing. We have presented some thoughts on evaluation of our tool and future work.

While we only return matches that actually appear in the book (i.e. they exist in the index), and by doing so eliminate the introduction of new words like Jarry's merdre, the swerve or aberration is still evident.

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

Table 13.1: Comparison of algorithms

13.1 Achievements

13.2 Implications

13.3 Recommendations

POST ☹

Appendix A

Code

setup

text setup

video setup

image setup

Clinamen

algorithm

get results

render results

etc

put cross refs for each piece of code here

A.1 Setup

stopwords:

```
[i, me, my, myself, we, our, ours, ourselves, you, your, yours, yourself,  
↳ yourselves, he, him, his, himself, she, her, hers, herself, it, its, itself,  
↳ they, them, their, theirs, themselves, what, which, who, whom, this, that,  
↳ these, those, am, is, are, was, were, be, been, being, have, has, had, having,  
↳ do, does, did, doing, a, an, the, and, but, if, or, because, as, until, while,  
↳ of, at, by, for, with, about, against, between, into, through, during, before,  
↳ after, above, below, to, from, up, down, in, out, on, off, over, under, again,  
↳ further, then, once, here, there, when, where, why, how, all, any, both, each,  
↳ few, more, most, other, some, such, no, nor, not, only, own, same, so, than,  
↳ too, very, s, t, can, will, just, don, should, now]
```

index:

```

1  defaultdict(<function <lambda> at 0x101d17938>, {u'fawn': defaultdict(<type
2   'list'>, {u'l_17': [101330, 111976], u'l_19': [345609], u'l_01': [92598,
3   469332, 469716, 469757, 469830, 469950]}}, u'pl\xe4tze': defaultdict(<type
4   'list'>, {u'l_11': [18287, 35736]}), u'lenitives': defaultdict(<type 'list'>,
5   {u'l_19': [121161]}), u'malheureuses': defaultdict(<type 'list'>, {u'l_06':
6   [18950, 52631, 65053, 65792, 79960, 127861, 176200], u'l_04': [51545,
7   93611]}), u'\xe9vit\xe2mes': defaultdict(<type 'list'>, {u'l_13': [82999]}),
8   u'nunney': defaultdict(<type 'list'>, {u'l_19': [38182, 160331]}),
9   u'\xe9difie': defaultdict(<type 'list'>, {u'l_12': [20278]}), u'nuzzing':
10  defaultdict(<type 'list'>, {u'l_19': [147035]}), u'huileux': defaultdict(<type
11  'list'>, {u'l_04': [69998]}), u'lentuli': defaultdict(<type 'list'>, {u'l_19':
12  [217454]}), u'porphyrogene': defaultdict(<type 'list'>, {u'l_01': [137368,
13  480308]}), u'woods': defaultdict(<type 'list'>, {u'l_14': [16256], u'l_05':
14  [2890], u'l_17': [11445, 34923, 48413, 59186, 61062, 78084, 78681, 81374,
15  101319, 137624], u'l_10': [12500, 21691, 33136], u'l_02': [4596, 6221],
16  u'l_01': [63176, 63622, 74535, 74807, 326433, 326464, 326472, 419835, 441374,
17  467791, 481003, 481113, 500201, 500331, 501595], u'l_27': [60731], u'l_26':
18  [1120, 9622], u'l_19': [149538, 232296, 294503]}), u'clotted':
19  defaultdict(<type 'list'>, {u'l_17': [92418], u'l_01': [53612, 133153, 200952,
20  241165]}), u'spiders': defaultdict(<type 'list'>, {u'l_19': [285105, 301556],
21  u'l_00': [3085]}), u'hanging': defaultdict(<type 'list'>, {u'l_17': [84239],
22  u'l_10': [35501, 92813, 126657], u'l_02': [6033, 33307, 34297], u'l_01':
23  [10317, 102489, 114549, 134507, 193056, 210178, 251167, 266320, 287765,
24  311802, 342798, 379707, 417326, 417956, 448394, 469204], u'l_00': [1831, 5490,
25  6891, 11088, 13846], u'l_27': [19757, 45058, 83470], u'l_26': [6500], u'l_19':
26  [8309, 15850, 17056, 20629, 25899, 47147, 47186, 49544, 108943, 177323,
27  200639, 252551, 334120, 341621]}), u'woody':

```

A.2 Damerau Levenshtein

```

1  # http://mwh.geek.nz/2009/04/26/python-damerau-levenshtein-distance/
2  # MIT license.
3  def dameraulevenshtein(seq1, seq2):
4      oneago = None
5      thisrow = range(1, len(seq2) + 1) + [0]
6      for x in xrange(len(seq1)):
7          twoago, oneago, thisrow = oneago, thisrow, [0] * len(seq2) + [x + 1]
8          for y in xrange(len(seq2)):
9              delcost = oneago[y] + 1
10             addcost = thisrow[y - 1] + 1
11             subcost = oneago[y - 1] + (seq1[x] != seq2[y])
12             thisrow[y] = min(delcost, addcost, subcost)
13             if (x > 0 and y > 0 and seq1[x] == seq2[y - 1] and
14                 seq1[x - 1] == seq2[y] and seq1[x] != seq2[y]):
15                 thisrow[y] = min(thisrow[y], twoago[y - 2] + 1)
16     return thisrow[len(seq2) - 1]

```

A.3 WordNet

A.3.1 Antinomy

```
synsets:  
[Synset('clear.n.01'), Synset('open.n  
.01'), Synset('unclutter.v.01'),  
Synset('clear.v.02'), Synset(''  
clear_up.v.04'), Synset('authorize  
.v.01'), Synset('clear.v.05'),  
Synset('pass.v.09'), Synset('clear  
.v.07'), Synset('clear.v.08'),  
Synset('clear.v.09'), Synset(''  
clear.v.10'), Synset('clear.v.11')  
, Synset('clear.v.12'), Synset(''  
net.v.02'), Synset('net.v.01'),  
Synset('gain.v.08'), Synset('clear  
.v.16'), Synset('clear.v.17'),  
Synset('acquit.v.01'), Synset(''  
clear.v.19'), Synset('clear.v.20')  
, Synset('clear.v.21'), Synset(''  
clear.v.22'), Synset('clear.v.23')  
, Synset('clear.v.24'), Synset(''  
clear.a.01'), Synset('clear.s.02')  
, Synset('clear.s.03'), Synset(''  
clear.a.04'), Synset('clear.s.05')  
, Synset('clear.s.06'), Synset(''  
clean.s.03'), Synset('clear.s.08')  
, Synset('clear.s.09'), Synset(''  
well-defined.a.02'), Synset('clear  
.a.11'), Synset('clean.s.02'),  
Synset('clear.s.13'), Synset(''  
clear.s.14'), Synset('clear.s.15')  
, Synset('absolved.s.01'), Synset  
('clear.s.17'), Synset('clear.r  
.01'), Synset('clearly.r.04')]  
  
synset item:clear.n.01  
synset item:open.n.01  
synset item:unclutter.v.01  
antonym out:clutter  
antonym in:clutter  
synset item:clear.v.02  
synset item:clear_up.v.04  
antonym out:overcast  
antonym in:overcast  
synset item:authorize.v.01  
  
synset item:clear.v.05  
synset item:pass.v.09  
synset item:clear.v.07  
antonym out:bounce  
antonym in:bounce  
synset item:clear.v.08  
synset item:clear.v.09  
synset item:clear.v.10  
synset item:clear.v.11  
synset item:clear.v.12  
synset item:net.v.02  
synset item:net.v.01  
synset item:gain.v.08  
synset item:clear.v.16  
synset item:clear.v.17  
synset item:acquit.v.01  
antonym out:convict  
antonym in:convict  
synset item:clear.v.19  
synset item:clear.v.20  
synset item:clear.v.21  
synset item:clear.v.22  
synset item:clear.v.23  
synset item:clear.v.24  
synset item:clear.a.01  
antonym out:unclear  
antonym in:unclear  
synset item:clear.s.02  
synset item:clear.s.03  
synset item:clear.a.04  
antonym out:opaque  
antonym in:opaque  
synset item:clear.s.05  
synset item:clear.s.06  
synset item:clean.s.03  
synset item:clear.s.08  
synset item:clear.s.09  
synset item:well-defined.a.02  
antonym out:ill-defined  
antonym in:ill-defined  
synset item:clear.a.11  
antonym out:cloudy
```

antonym in:cloudy	synset item:absolved.s.01
synset item:clean.s.02	synset item:clear.s.17
synset item:clear.s.13	synset item:clear.r.01
synset item:clear.s.14	synset item:clearly.r.04
synset item:clear.s.15	

A.3.2 Syzygy

synsets:	hypernym out:country
[Synset('clear.n.01'), Synset('open.n. .01'), Synset('unclutter.v.01'), Synset('clear.v.02'), Synset('' clear_up.v.04'), Synset('authorize .v.01'), Synset('clear.v.05'), Synset('pass.v.09'), Synset('clear .v.07'), Synset('clear.v.08'), Synset('clear.v.09'), Synset('' clear.v.10'), Synset('clear.v.11') , Synset('clear.v.12'), Synset('' net.v.02'), Synset('net.v.01'), Synset('gain.v.08'), Synset('clear .v.16'), Synset('clear.v.17'), Synset('acquit.v.01'), Synset('' clear.v.19'), Synset('clear.v.20') , Synset('clear.v.21'), Synset('' clear.v.22'), Synset('clear.v.23') , Synset('clear.v.24'), Synset('' clear.a.01'), Synset('clear.s.02') , Synset('clear.s.03'), Synset('' clear.a.04'), Synset('clear.s.05') , Synset('clear.s.06'), Synset('' clean.s.03'), Synset('clear.s.08') , Synset('clear.s.09'), Synset('' well-defined.a.02'), Synset('clear .a.11'), Synset('clean.s.02'), Synset('clear.s.13'), Synset('' clear.s.14'), Synset('clear.s.15') , Synset('absolved.s.01'), Synset('' clear.s.17'), Synset('clear.r .01'), Synset('clearly.r.04')]	hypernym in:country [] synset item:unclutter.v.01 hypernym out:change hypernym in:change hypernym out:alter hypernym out:modify [] synset item:clear.v.02 hypernym out:make hypernym in:make hypernym out:create [] synset item:clear_up.v.04 [] synset item:authorize.v.01 hyponym out:approbate hyponym out:approve hyponym out:O.K. hyponym out:okay hyponym out:sanction hyponym out:certificate hyponym in:certificate hyponym out:commission hyponym out:declare hyponym in:declare hyponym out:license hyponym out:licence hyponym out:certify hyponym out:validate hyponym out:formalize hyponym out:formalise hyponym out:permit hyponym in:permit hyponym out:allow hyponym in:allow
synset item:clear.n.01	
hypernym out:innocence	
[]	
synset item:open.n.01	
hypernym out:area	

hypernym out:let
hypernym in:let
hypernym out:countenance
hypernym in:countenance
[]
synset item:clear.v.05
hyponym out:clear-cut
hyponym out:deforest
hyponym out:disforest
hyponym out:disafforest
hyponym out:denude
hyponym out:bare
hyponym in:bare
hyponym out:denudate
hyponym out:strip
hyponym out:stump
hypernym out:remove
hypernym out:take
hypernym in:take
hypernym out:take_away
hypernym out:withdraw
[]
synset item:pass.v.09
hyponym out:clear
hyponym in:clear
hypernym out:succeed
hypernym in:succeed
hypernym out:win
hypernym out:come_through
hypernym out:bring_home_the_bacon
hypernym out:deliver_the_goods
[]
synset item:clear.v.07
[]
synset item:clear.v.08
hypernym out:vanish
hypernym in:vanish
hypernym out:disappear
hypernym out:go_away
[]
synset item:clear.v.09
hyponym out:hop
hypernym out:pass
hypernym in:pass
hypernym out:overtake
hypernym out:overhaul
[]
synset item:clear.v.10
hypernym out:clarify
hypernym out:clear_up
hypernym out:elucidate
[]
synset item:clear.v.11
hypernym out:free
hypernym in:free
hypernym out:discharge
[]
synset item:clear.v.12
hypernym out:rid
hypernym out:free
hypernym in:free
hypernym out:disembarrass
[]
synset item:net.v.02
hypernym out:yield
hypernym out:pay
hypernym in:pay
hypernym out:bear
[]
synset item:net.v.01
hypernym out:profit
hypernym out:gain
hypernym in:gain
hypernym out:benefit
hypernym in:benefit
[]
synset item:gain.v.08
hyponym out:eke_out
hyponym out:squeeze_out
hyponym out:gross
hyponym out:profit
hyponym out:turn_a_profit
hyponym out:rake_in
hyponym out:shovel_in
hyponym out:rake_off
hyponym out:take_home
hyponym out:bring_home
hyponym out:yield
hyponym out:pay
hyponym in:pay
hyponym out:bear
hypernym out:get

hypernym out:acquire []
synset item:clear.v.24
hypernym out:remove
hypernym out:take
hypernym in:take
hypernym out:take_away
hypernym out:withdraw []
synset item:clear.a.01
[]
synset item:clear.s.02
[]
synset item:clear.s.03
[]
synset item:clear.a.04
[]
synset item:clear.s.05
[]
synset item:clear.s.06
[]
synset item:clean.s.03
[]
synset item:clear.s.08
[]
synset item:clear.s.09
[]
synset item:well-defined.a.02
[]
synset item:clear.a.11
[]
synset item:clean.s.02
[]
synset item:clear.s.13
[]
synset item:clear.s.14
[]
synset item:clear.s.15
[]
synset item:absolved.s.01
[]
synset item:clear.s.17
[]
synset item:clear.r.01
[]
synset item:clearly.r.04
[]

A.4 Random Quotes

```
1 def getrandquote():
2     # RANDOM QUOTES
3     root_path = os.path.dirname(os.path.abspath(__file__))
4     root_path = root_path[:-4]
5     corpus_root = root_path + '/app/static/corpus'
6     path_b = corpus_root + '/quotes.txt'
7     quotes_text = codecs.open(path_b, "r", encoding='utf-8')
8     quotestext = quotes_text.readlines()
9     quotes_text.close()
10    # print random.choice(quotestext)
11    return random.choice(quotestext)
```

Appendix B

Publications

This chapter includes copies of the following publications and talks:

finish list here

1. Presentation slides for a [Computer Arts Society \(CAS\) Institute of Creative Technologies \(IOCT\) talk at the Phoenix in Leicester, UK \(14 Oct 2015\)](#).¹
2. [IOCT Leicester Media School \(LMS\) Showcase, De Montfort University \(DMU\), Leicester, UK \(5 Nov 2014\)](#).
3. [IOCT PhD Research Showcase, Phoenix Cube, Leicester, UK \(15-18 Aug 2014\)](#).²
4. Journal article 'The pataphysics of creativity: developing a tool for creative search'. Routledge: Digital Creativity, Volume 24, Issue 3 (2013).
5. Talk given at Creativity and Cognition conference in Sydney, Australia (20 June 2013).³
6. Conference paper 'Creative Search Using Pataphysics'. Proceedings of the 9th ACM conference on Creativity and Cognition (2013).
7. [DMU Talk, Leicester, UK \(13 Feb 2013\)](#).⁴
8. Conference paper 'A Framework for Creativity in Search Results'. Proceedings of the 3rd International Conference on Creative Content Technologies (2011).⁵

¹<http://interactlabs.co.uk/news/2015/10/ioct-talks---videos-now-available>, <https://vimeo.com/142947457>

²Sean's Gallery <https://www.flickr.com/photos/seancuttlefish/sets/72157646116801940/>, Interact Blog <http://interactlabs.co.uk/diary/2014/08/dmu-phd-showcase>

³<http://cc13.creativityandcognition.com/>

⁴<https://www.youtube.com/watch?v=UxYUZMyPEOo>

Pata-computed Poetry

Fania Raczinski

pata.physics.wtf



Raymond Queneau



Oulipo



Pataphysics



Technologies

- Python
- Javascript, HTML & CSS
- Flask, Jinja,
- NLTK & WordNet

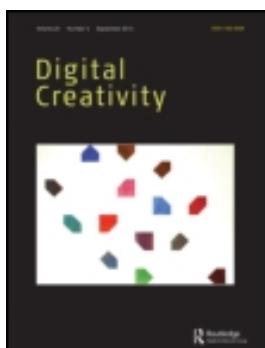
creative computing vs. computational creativity

<h2>Faustroll's Corpus</h2> <p>Alfred Jarry, Edgar Allan Poe, Cyrano de Bergerac, Saint Luke, Lean Bloy, Samual Taylor Coleridge, Georges Darien, Marceline Desbordes-Valmore, Max Elskamp, Jean-Pierre Claris de Florian, V.A., Christian Dietrich, Gustave Kahn, Le Compte de Lautreamont, Maurice Maeterlinck, Stephane Mallarme, Mendes, Homer, Josephin Peladan, Francois Rabelais, Jean de Chirra, Henri de Regnier, Arthur Rimbaud, Marcel Schwob, Paul Verlaine, Emile Verhaeren and Jules Verne</p>	<h2>Concepts</h2> <p>[Clinamen] “smallest possible aberration that can make the greatest possible difference” Boek</p> <p>[Syzygy] alignment of three celestial bodies, pun</p> <p>[Antinomy] mutually incompatible, paradox</p>
<h2>Patalgorithms</h2> <p>[Clinamen] Introduction of spelling errors</p> <p>[Syzygy] Moving within the semantic hierarchy</p> <p>[Antinomy] Opposite</p>	<h2>Love Poem</h2>  <pre> < kissing me like a little girl <u>home</u> from school > < like the <u>pope</u>'s mule > < after a <u>long</u> period > < we prepared to draw <u>lots</u> > < I have been a <u>long</u> time on this earth > < agape they heard me call > < il sera <u>pair</u> d' Angleterre avant peu ... Ah > < so many acres to be <u>lost</u> > < Chewing with teeth that grind and <u>crush</u> > < And each in jovial mood his <u>mate</u> address'd > < what though from <u>pole</u> to pole resounds her name > < elate and <u>loud</u> > < if it was a <u>note</u> for the Reckoning > < -- the bottom of a <u>pair</u> of breeches for a vessel full of fart > </pre>
<h2>Love Poem</h2> <p>Maeterlinck, Rabelais, Verne, Poe Mendes, Coleridge, Darien, Rabelais Verhaeren, Homer, Homer Homer, Bergerac, Rabelais</p> <p>CCCC CSSC SSC CCS</p>	<h2>Future</h2> <ul style="list-style-type: none"> • rhyming pattern • sentence structure • corpus or web • multi-keyword • Digital Opera • PhD

- IOCT LMS Showcase. 5 Nov 2014. <!-- ![Tweets](Citations/tweets.png) -->

Phoenix Exhibit. Phoenix Cube Showcase. 15-18 Aug 2014. [Sean's Gallery](<https://www.flickr.com/photos/seancuttlefish/sets/72157646116801940/>), [Interact Blog](<http://interactlabs.co.uk/diary/2014/08/dmu-phd-showcase>)

This article was downloaded by: [De Montfort University]
On: 13 November 2013, At: 06:34
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office:
Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Digital Creativity

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/ndcr20>

The pataphysics of creativity: developing a tool for creative search

Andrew Hugill^a, Hongji Yang^a, Fania Raczinski^a & James Sawle^a

^a Institute of Creative Technologies, De Montfort University

Published online: 12 Nov 2013.

To cite this article: Andrew Hugill, Hongji Yang, Fania Raczinski & James Sawle (2013) The pataphysics of creativity: developing a tool for creative search, *Digital Creativity*, 24:3, 237-251, DOI: [10.1080/14626268.2013.813377](https://doi.org/10.1080/14626268.2013.813377)

To link to this article: <http://dx.doi.org/10.1080/14626268.2013.813377>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Digital Creativity, 2013
Vol. 24, No. 3, 237–251, <http://dx.doi.org/10.1080/14626268.2013.813377>

The pataphysics of creativity: developing a tool for creative search

**Andrew Hugill, Hongji Yang,
Fania Raczinski and James Sawle**

Institute of Creative Technologies, De Montfort University

a.hugill@bathspa.ac.uk; hyang@dmu.ac.uk; fania@dmu.ac.uk; jsawle@dmu.ac.uk

Abstract

We introduce the idea of a new kind of web search tool that uses the literary and philosophical idea of pataphysics as a conceptual framework in order to return creative results. Pataphysics, the science of exceptions and imaginary solutions, can be directly linked to creativity and is therefore very suitable to guide the transformation from relevant into creative search results. To enable pataphysical algorithms within our system we propose the need for a new type of system architecture. We discuss a component-based software architecture that would allow the flexible integration of the new algorithms at any stage or location and the need for an index suitable to handle patadata, data which have been transformed pataphysically. This tool aims to generate surprising, novel and provocative search results rather than relevant ones, in order to inspire a more creative interaction that has applications in both creative work and learning contexts.

Keywords: pataphysics, creativity, information retrieval, creative computing, component-based software engineering

1 Introduction

In this article we propose a new type of web search engine, reminiscent of the experience of ‘surfing the Web’. This is in contrast to current search engines which value relevant results over creative ones. ‘Surfing’ used to be a creative interaction between a user and the web of information on the Internet, but the regular use of modern search engines has changed our expectations of this sort of knowledge acquisition. It has drifted away from a *learning process* by exploring the Web to a straightforward process of information retrieval similar to looking up a word in a dictionary.

Jorge Luis Borges has provided us with a very useful example to illustrate our idea. His ‘Chinese Encyclopaedia’ (Borges 2000, 231) lists the following results under the category of ‘animal’:

- (1) those that belong to the emperor;
- (2) embalmed ones;
- (3) those that are trained;
- (4) suckling pigs;
- (5) mermaids;
- (6) fabulous ones;
- (7) stray dogs;
- (8) those that are included in this classification;
- (9) those that tremble as if they were mad;
- (10) innumerable ones;
- (11) those drawn with a very fine camel’s hair brush;
- (12) etcetera;
- (13) those that have just broken the flower vase;
- (14) those that at a distance resemble flies.

Hugill et al.

Although these are all perfectly valid results, it is clear that they form a more creative, even poetic, view of what an animal might be than the Oxford English Dictionary's prosaic: 'a living organism which feeds on organic matter' (Oxford Dictionaries n.d.).

To achieve this sort of creativity in search results we propose the use of pataphysical methods. Pataphysics is highly subjective and particular, and as such is very suitable for this kind of transformation from relevant to creative. We hope that the tool will prove useful as a source for information and inspiration and at the same time challenge the way we think about information retrieval on the Web. The Web is not a place limited to one discipline, and in fact creating a transdisciplinary field of 'web science' was suggested by Hendlér et al. in 2008. Our project will therefore span several disciplines as well.

Given the breadth of the Web and its inherently multi-user (social) nature, its science is necessarily interdisciplinary, involving at least mathematics, [computer science], artificial intelligence, sociology, psychology, biology, and economics

(Hendlér et al. 2008).

Over the rest of the article, we will examine how pataphysics and creativity map onto one another, give an outline of the field of information retrieval, and discuss how this new type of search could be implemented in future systems. We conclude with a short discussion and summary of the article.

2 Creativity and pataphysics

[Pataphysics] can only be defined in a new undiscovered language because too obvious: tautology

(Baudrillard 2007).

The creative process normally involves a move from the known to the unknown, and sometimes from the named to the unnamed. In bringing something new into existence, the human qualities of openness and tolerance of ambiguity are generally regarded as highly desirable. We may define crea-

tivity as *the ability to use original ideas to create something new and surprising of value*. We generally speak of creative 'ideas' rather than 'products', which merely provide evidence of a creative process that has already taken place. Both the originality and the value of an idea are evaluated using subjective criteria. Pataphysics, which represents an extreme form of subjectivity, is therefore a highly appropriate framework within which to encourage and enable creative thinking and operations.

2.1 Pataphysics

Pataphysics¹ was invented by a group of French schoolboys at the Lycée de Rennes in the 1880s. One of their number was the author and playwright Alfred Jarry (1873–1907), who later developed the concept both in his celebrated *Ubu* plays and in his novels and speculative writings. He defined it as follows:

Pataphysics ... is the science of that which is superimposed upon metaphysics, whether within or beyond the latter's limitations, extending as far beyond metaphysics as the latter extends beyond physics. Ex: an epiphenomenon being often accidental, Pataphysics will be, above all, the science of the particular, despite the common opinion that the only science is that of the general. Pataphysics will examine the laws which govern exceptions, and will explain the universe supplementary to this one; or, less ambitiously, will describe a universe which can be—and perhaps should be—envisioned in the place of the traditional one, since the laws which are supposed to have been discovered in the traditional universe are also correlations of exceptions, albeit more frequent ones, but in any case accidental data which, reduced to the status of unexceptional exceptions, possess no longer even the virtue of originality.

DEFINITION. Pataphysics is the science of imaginary solutions, which symbolically attributes the properties of objects, described by their virtuality, to their lineaments

(Jarry 1996, 21).

This may be summarised in following way: pataphysics:

- is the science of imaginary solutions;
 - is the science of the particular;
 - is the science of the laws governing exceptions and contradictions;
 - is to metaphysics as metaphysics is to physics.
- The conceptual space of pataphysics is a ‘universe supplementary to this one’ (Jarry 1996, 21). We argue that pataphysics can facilitate creative computing. *Constraints* are the rules that we set in our space, the grammar that we want to use. A pataphysical grammar would consist of exceptions, syzygies, anomalies, clinamen, antinomies, contradictions, equivalents and imaginaries. Such constraints can transform the ways in which we may navigate the new space. Pataphysical concepts will cause surprise and therefore could be considered unconventional.

Since pataphysics is concerned with the laws governing *exceptions*, its application in creative computing will focus on the ludic aspects of unique occurrences, rather than predictable recurrence of expected outcomes (Bök 2002). It is axiomatic that no single viewpoint may predominate, an understanding that was codified by Jarry and subsequent theorists as the ‘doctrine’ of Equivalence. Abstraction and generalisation in creative computing may therefore be founded upon a parallel we would draw between meta-metaphysics (pataphysics) and meta-metadata (patadata), which will be discussed in more detail below. Since pataphysics is the science of imaginary solutions, *imagination* (specifically a poetic imagination) provides the guiding principle for our work. Domain-specific knowledge and skill is described by the final line of Jarry’s *Exploits and Opinions of Doctor Faustroll, Pataphysician*: ‘Pataphysics is the science’ (Jarry 1996, 114).

2.2 Creativity

It is instructive to overlay these ideas on existing theories of creativity. Margaret Boden (2003), for example, has defined *P-creativity* (short for psychological creativity) as the personal kind of creativity that is novel in respect to the individual mind, and *H-creativity* (short for historical

creativity) as fundamentally novel in respect to the whole of human history. This allows for subjective evaluation of any idea. A child that builds a corbelled arch out of woodblocks, without any knowledge of physics or architecture, could be called creative. The child created something new and valuable within its own constraints and could therefore be called P-creative, but since the technique was already known historically it cannot be considered H-creative.

Using Boden’s definition we can call an idea ‘new’ if it is new to the individual who came up with it, making the idea P-creative. We can say that a creative idea can be seen from two perspectives: the subjective (P-creative) and the objective (H-creative) view. She argues that constraints support creativity, and are even essential for it to happen. ‘Constraints map out a territory of structural possibilities which can then be explored, and perhaps transformed to give another one’ (Boden 2003, 82).

This echoes the ideas of groups such as the *Oulipo* (which began as a Sub-Commission of the Collège de ’Pataphysique), who investigate ‘potential literature’ by creating constraints that frequently have a ludic element. Various other groups, the Ou-x-Pos, perform similar operations in fields as diverse as cinema, politics, music and cooking (Motte 1998).

Boden’s conceptual space is the ‘territory of structural possibilities’. So, the conceptual space of a teacup might be that it is meant to carry a certain amount of tea without breaking or burning fingers. It wouldn’t be wise to create a teacup made out of paper. But whether we make a cup out of glass or porcelain or how we shape the cup or the handle is pretty much up to the individual’s creativity. Being able to move around in this conceptual space, experiment (in thought or in reality) and play with different ideas while still following a given set of constraints is a good starting point for creativity to happen. Boden defines three sub-types of creativity:

- combinational creativity: making unfamiliar combinations of familiar ideas;
- exploratory creativity: exploration of conceptual spaces;

Hugill et al.

- transformative creativity: transformation of space.

The Oulipo similarly classifies its conceptual space under two broad headings: the synthetic and the analytic:

In the research which the Oulipo proposes to undertake, one may distinguish two principal tendencies, oriented respectively towards Analysis and Synthesis. The analytic tendency investigates works from the past in order to find possibilities that often exceed those their authors had anticipated. . . . The synthetic tendency is more ambitious: it constitutes the essential vocation of the Oulipo. It's a question of developing new possibilities unknown to our predecessors. This is the case, for example, of [Raymond Queneau's] 100,000,000,000 Poems or the Boolean haikus

(Motte 1998, 27).

Later writings develop these ideas in more detail. *La Littérature Potentielle* (Oulipo 1973), is divided into several sections, dealing with clusters of methods, that include: *anoulipisms* (analytical oulipisms, such as combinatorial literature); *use of pre-existing structures* such as lipograms (omitting a letter or letters), palindromes and snowballs (in which each successive word adds or subtracts a letter), homophonic translation, tautogram, and definitional literature; *lexical, syntactic, or prosodic manipulations* (such as the celebrated S + 7, in which each substantive is replaced by the seventh word after it in a standard dictionary); *lexicographical or prosodic synthoulipisms* (early algorithmic methods); and *perimathematical synthoulipisms* (such as the Boolean poetry and combinatorial works already mentioned).

Boden links her three aspects of creativity to three sorts of surprise. She says that creative ideas are surprising because they go against our *expectations*. ‘The more expectations are disappointed, the more difficult it is to see the link between old and new’ (Boden 2003, 84). This suggests that fewer expectations (an open mind) allow creativity to happen more easily. Empirical experiences form expectations, which hinder our ability to accept creative ideas when they

happen. In order to be able to recognise creative ideas, we need to be able to see what they all have in common and in what way they differ, and not reject unusual, unexpected ones.

Unless someone realizes the structure which old and new spaces have in common, the new idea cannot be seen as the solution to the old problem. Without some appreciation of shared constraints, it cannot even be seen as the solution to a new problem intelligibly connected with the previous one

(Boden 2003, 84).

It is clear that the Oulipo has a similar approach in its theorising of potential literature. Releasing creativity through constraint is its essential *raison d'être*.

This is not to say that experience and knowledge are necessarily bad for creativity. To appreciate creativity we need to be knowledgeable in the relevant domain to be able to recognise old and new connections and transformations. But we also need a certain level of openness and tolerance for ambiguity to overcome our expectations. Perhaps it is for this reason that ‘creative people’ are often assumed to have particular personality traits. Sternberg (1988, 1999), for example, proposes that these comprise: independence of judgement, self-confidence, attraction to complexity, aesthetic orientation, tolerance for ambiguity, openness to experience, psychoticism, risk-taking, androgyny, perfectionism, persistence, resilience, and self-efficacy. More empirically, Heilman, Nadeau, and Beversdorf (2003) have investigated the possible brain mechanisms involved in creative innovation. While a certain level of domain-specific knowledge and special skills are necessary components of creativity, they point out that ‘co-activation and communication between regions of the brain that ordinarily are not strongly connected’ (Heilman, Nadeau, and Beversdorf 2003, 269) might be equally important.

Newell, Shaw, and Simon (1963) add to the above with their report on the creative thinking process. They identify three main conditions for creativity: the use of imagery in problem

solving; the relation of unconventionality to creativity; and the role of hindsight in the discovery of new heuristics. Other issues they point out are abstraction and generalisation. So, for example, poets transform the grammar of their conceptual space (in this case, language) to create new sentence structures in a poetic form. By doing so, they go against the expectations, the possibilities of the language and cause surprise. Some people might not understand the transformations and therefore the jokes or beauty of a poem simply because they are either not able to recognise connections between the old and newly transformed elements (maybe due to a lack of knowledge in the poems topic or in that particular language) or because they do not want to accept unconventional methods.

2.3 Creative computing

But how may we apply the insights into creativity described above in computing? One approach is described by Simon Colton (2008), who suggests we should adopt human skill, appreciation and imagination:

Without skill, [computers] would never produce anything. Without appreciation, they would produce things which looked awful. Without imagination, everything they produced would look the same.

(Colton 2008, 6)

He thinks that evaluating the worth of an idea or product is the biggest challenge facing *computational creativity*. Whereas in conventional problem-solving success is defined as finding a solution, in a creative context more aesthetic considerations have to be taken into account. He suggests three ways for computer programs to generate creative artefacts:

- (1) mimicking human skill
- (2) mimicking human appreciation
- (3) mimicking human imagination.

Since our solutions will be imaginary, our aim is not so much to have the computer generate creative artefacts as to engage in a creative dialogue with the user. Therefore, we do not intend to move as close to artificial intelligence as

Colton's framework seems to suggest. In the pataphysical universe, ideas such as 'human skill', 'human imagination' and 'human appreciation' are too generalised to be useful. One may very well ask: *which human?* And *when, where* and even *why*? Rather, our project will aim to produce an exceptional computational entity that consistently generates surprising and novel provocations to the users, who in turn may navigate and modify these by deploying their own skills, appreciation and imagination. The relationship between the two will develop quite rapidly into one of mutual subversion since, however apparent the 'rules of the game' may become, the outcomes will always be particular or exceptional.

2.4 Pataphysical computing

We are not the first people to attempt to apply pataphysical ideas in computer science. Johanna Drucker focused specifically on the cleft between formal logic and subjective judgement. She introduced the discipline of 'speculative computing' as a solution to that problem (Drucker and Nowviskie 2007). The concept can be understood as a criticism of mechanistic, logical approaches that distinguish between subject and object.

Speculative computing takes seriously the destabilization of all categories of entity, identity, object, subject, interactivity, process, or instrument. In short, it rejects mechanistic, instrumental, and formally logical approaches, replacing them with concepts of autopoiesis (contingent interdependency), quantum poetics and emergent systems, heteroglossia, indeterminacy and potentiality, intersubjectivity, and deformance. Digital humanities is focused on texts, images, meanings, and means. Speculative computing engages with interpretation and aesthetic provocation

(Drucker 2009, 29).

For Drucker, *aesthesia* (ambiguous and subjective knowledge) is fundamentally opposed to *mathesis* (formal objective logic) and subjectivity is always in opposition to objectivity. Knowledge is a matter of interpretation of information, which can be represented digitally as data and metadata. She

introduces what she calls a '*patacritical*' method of including exceptions as rules, even if repeatability and reliability are compromised. Bugs and glitches are privileged over functionality, and are 'valuable to speculation in a substantive, not trivial, sense' (Drucker 2009, 26). As she says: 'Pataphysics inverts the scientific method, proceeding from and sustaining exceptions and unique cases' (Drucker and Nowviskie 2007, 434).

In order to break out of the formal logic and defined parameters of computer science, she asserts, we need speculative capabilities and pataphysics. 'The goal of pataphysical and speculative computing is to keep digital humanities from falling into mere technical application of standard practices' (Drucker and Nowviskie 2007, 441). She links interface design with other speculative computing principles, referring to Kant's idea of art as 'purposiveness without purpose' and saying that the appreciation of design as a thing in itself (regardless of utility) is a goal of speculative aesthetics (Drucker and Nowviskie 2007, 437).

Table 1.

Creativity	Pataphysics
Combinational	Antinomy
Juxtaposition of dissimilar, Bisociation, Deconceptualisation	Symmetry, duality, mutually incompatible, contradicting, simultaneous existence of mutually exclusive opposites
Exploratory	Syzygy
Noticing new things in old places	Alignment of three celestial bodies in a straight line, pun, conjunction of things, something unexpected and surprising
Transformative	Anomaly
Making new thoughts possible by transforming old conceptual space, altering its own rules	Exceptions, equality
	Clinamen
	Unpredictable swerve, the smallest possible aberration that can make the greatest possible difference

2.5 Creativity and pataphysics compared

To conclude this discussion, consider Table 1, which compares some of the key ideas of creativity (Boden 2003; Bök 2002; Indurkha 1997; Koestler 1964) with the main pataphysical operations. It will be seen that pataphysics succeeds in bringing into sharp relief the more generalised scientific ideas. The pataphysical terms are taken from the natural sciences or philosophy, but always with an ironic twist, betraying their underlying humour. They connect quite strongly with the primary descriptors of creativity, while adding a certain layer of *jouissance*. *Pataphysics is self-avowedly useless, but its principles may prove surprisingly useful within this context.*

3 Information retrieval systems

Information retrieval is one of the common processes that a person carries out day-to-day, usually without even thinking about it. The amount of information that a human comes in contact with on a daily basis is overwhelming, and as such we have developed very sophisticated methods of finding the *relevant information* instantaneously. However, it is also possible to see how this relates to a large number of commonly used computer systems.

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually on local computer servers or on the Internet)

(Manning, Raghavan, and Schütze 2008, 1).

It is important to note that whilst a large proportion of information retrieval (IR) is focused on web search engines, this is not the only application. The reason that such a large focus is on this area is due to the unique challenges it holds: huge quantities of *unstructured data* which change over time and can be in a number of formats. The true aim of any research into search engines is that it can be applied back to the general field of IR and enhance a much larger ecosystem of systems.

However, research in all of IR focuses on arbitrary values of success, called precision and recall, the fraction of retrieved instances that are relevant and the fraction of relevant instances that are retrieved respectively. Whilst these measures are logical, they are arbitrary due to the subjectiveness of relevance. And due to the clinical nature of the measures, returning results that are partially related to the request would be detrimental to the perceived quality of the system, irrelevant of the insightful knowledge they may provide.

Whilst IR systems can take many different forms, Baeza-Yates and Ribeirio-Neto (2011) defined a standard model, which allows all systems to be broken down into similar components:

An IR model is a quadruple [D, Q, F, R(q_i, d_j)] where

- D is the set composed of logical views (or representations) for the documents in the collection;
 - Q is the set composed of logical views (or representations) for the user information needs. Such representations are called queries;
 - F is a framework for modelling document representations, queries, and their relationships...;
 - R(q_i, d_j) is a ranking function which associates a real number with a query representation q_i ∈ Q and a document representation d_j ∈ D. Such ranking defined an ordering among the documents with regard to the query q_i.
- (Baeza-Yates and Ribeirio-Neto 2011, 58)

It is possible, under this definition, that there is no ranking function; such is the case for the Boolean model. Whilst this may not appear logical when considering search engines, there are a number of cases where returning all possible results which match our ‘need’ without bias can be useful. It is not possible, however, for an IR system to exist without any of the other components.

3.1 Classical IR models

The classification of classical IR models typically includes the Boolean, vector space and probabilistic models (Dominich 2000). Each of the models is built on pure mathematical underpinnings, which has also lead to research into a unified model for them.

The Boolean retrieval model is based on set theory and Boolean algebra. The model views documents as a collection of words or, more precisely, a collection of indexed terms present in those documents. A user request (query) is usually a Boolean expression written as a series of terms connected by Boolean operators such as AND, OR and NOT.

In the vector space model, documents are represented as vectors (Wong and Raghavan 1984). The success or failure of this method is based on term weighting. Terms are words, phrases, or any other indexing units used to identify the contents of a text. As such, term weighting is assigning a value to each term in order to define its importance in relation to the rest of the terms within that context (Salton and Buckley 1987). Polettini (2004) points out that term weighting schemes play an important role for the similarity measure, which plays a key role in the retrieval performance of IR systems.

Due to the fact that vector space models only link documents through related terms, we have no in-built technique to handle relevance. The aim of probabilistic methods is to rank a collection of documents in decreasing probability of their relevance to a query. This is often referred to as the probabilistic ranking principle (Cooper 1968). The idea of using probability was suggested as no system can predict with certainty the documents that a requester might find useful (Maron and Kuhns 1960).

3.2 Latent semantic indexing

Latent semantic indexing (LSI) is an indexing and retrieval model that attempts to identify patterns in the relationships between the terms and concepts contained in an unstructured collection of text. A key feature of LSI is its ability to extract the conceptual content of a body of text by establishing

Hugill et al.

associations between those terms that occur in similar contexts (Deerwester et al. 1990). The model is based upon a mathematical method called single value decomposition as well as correspondence analysis (Benzécri 1973). LSI overcomes two of the most problematic constraints of Boolean keyword queries: synonymy and polysemy. Synonymy and polysemy are often the cause of mismatches in the vocabulary used by the authors of documents and the users of information retrieval systems (Furnas et al. 1987). There has been some promising research into using LSI instead of the vector space model (Chen and Tai 2009).

3.3 Artificial intelligence models

Numerous artificial intelligence-inspired models have been proposed, from neural networking, genetic algorithms, knowledge bases and natural language processing. Each of these different systems manages to solve a different problem within the field of IR; however, trying to generalise these models has not proved as fruitful as they were in their specialised fields.

4 Beyond the realm of traditional IR systems

Most modern web search engines, excluding semantic search engines, have a similar architecture, irrespective of the IR model on which they are based (see Figure 1). The main reason for this is due to the generic data store at the heart of them, the inverted index, which is a very efficient method of storing and searching over the contents of documents.

In an inverted index, the contents of a document are broken into various different combinations or terms by the indexer, and a link to the original document is stored with each of these terms. This means that when searching for a *keyword*, instead of having to look at every document and its contents, the system just looks for all terms that match the request and returns the various links that match. The inverted index is quick at retrieval; however, building the index is slower.

Even with these characteristics, the inverted index is not suitable with respect to any of the above definitions of creativity. We are only able to search over the contents of the document as they are, with *no understanding of their meaning*. As such, being able to implement pataphysical themes like clinamen or syzygy would be very challenging.

It is possible to apply these concepts to a traditional search engine architecture by modifying the user's search request. Hendlér and Hugill (2011) suggest that by using 'panalogies' we can model patadata and as such apply pataphysical constructs to requests. In the proposal that is outlined, the system would be applied to work on the open Web, using results from commercial search engines, as well as domain specific systems such as the British Library.

However, there is a limitation to such a system. Whilst we can modify the initial request to something with a more creative twist, the system cannot make decisions based on the underlying content of the results. As such, the quality of the results is limited by the quality of the indexer and not the search algorithm. Whilst this could be argued to be true in any search architecture, the index is built up of data that we wish to access directly, i.e. searching over the content to find a document that matches based on certain rules. With respect to creative search, it makes more sense that we look at how different parts of the document relate to each other, and other documents based upon underlying meaning, and not pure text. Even with this in mind, such a system would be adventurous from a creative standpoint over current search engines, and would provide an interesting insight into how people would respond to such a system and how important the user interface would be in such a system.

4.1 Semantic search engines

Semantic search engines would therefore seem to be a more logical fit to a pataphysics-inspired creative search engine as they will allow the creation of links between different documents based on more than the exact words used.

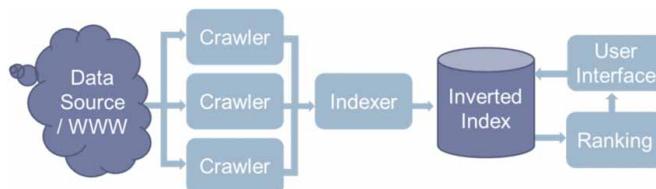


Figure 1. A traditional architecture for a web search engine. © Fania Raczinski and James Sawle. Reproduced with permission.

The key difference between the architectures of traditional IR systems is the way that the data is stored, and hence the indexing process. The majority of different semantic search systems use *Resource Description Framework (RDF) triples* as a way of storing data, based on semantic web ontologies. In RDF, each entry to the data store has the following attributes: <<object>> <<relation>> <<value>>. For example, a blue balloon would be <<balloon>> <<hasColour>> <<blue>>. This is not meant to represent the syntax of RDF; however, for this example the relation of ‘hasColour’ and the concept of blue being a colour would have to already have been defined. However, trying to represent the concept of ‘twelve blue balloons’ requires even more relations and concepts to be defined; therefore, if we end up with a large amount of loosely related data, the number of concepts and relations defined will explode. However, if the data are tightly related, the number of relations and concepts is much more concise and cogent.

With the data stored in this format, *inference logic* and/or *fuzzy set theory* can be used to carry out searches over the data set to return concepts that relate. These inference searches are slow and tend not to relate directly to documents, instead returning a list of different concepts which can then be linked back to the documents. This is usually done with an inverted index using traditional methods to return documents that match numerous combinations of the concepts.

With this method, the trickiest part of the system is *indexing*, as a document must be related back to concepts that exist within the system already. If they do not exist in the system

already, the concept must be found in an *ontology* that has already been defined, which then leads to the problem of ontology merging, or creating them from scratch.

Whilst this clearly allows more for the concepts that we have defined for results to be creative, there are a number of issues that arise. For example, once a document has been added to the system, the concepts are set in stone. Whilst the RDF store will evolve over time and hence change the concept results that emerge, the document’s classifications are set in stone. This is not just a problem for a creative search engine, but for all semantic search engines as well.

Also, within an ontology, different concepts tend to be linked with relations that are descriptive, such as <<isAn>>, <<hasColour>>, <<produces>>, etc. However, these types of relations are not analogous with the pataphysical concepts that have been defined, and as such it is not immediately apparent how one could implement a syzygetic transformation of a search request using an RDF data store.

As can be seen by looking at the two main search architectures, a new IR system architecture is needed; however, instead of defining it from the top-down, the algorithms need to be defined first to allow maximum flexibility in the system to allow the definition of creativity to evolve over time.

5 Pataphysical search algorithms

The conceptual space for our project is ‘pataphysical Web searching’. There are some very simple rules or constraints that form an initial definition of the project. For example, it is clear that we

Hugill et al.

want to search the World Wide Web (rather than a library database), that we want to return a list of search results (and not a pile of books), and that we want the search process and its results to be creative/pataphysical (rather than relevant). In a more technical sense, we have the query term(s), the index (of all web pages that we have crawled) and some pataphysical rules in our conceptual space. How we structure our search system, how we format the index or how we go about finding our results is not in our conceptual space, however. We can explore the space to its limits, and we can transform it if we want to or feel like we need to. Our pataphysical rule set will include methods for transforming the space. By applying pataphysical rules to find results to our query we are *pataphysicalising* the query.

Definitions:

- To pataphysicalise (verb):* apply pataphysical transformations;
- Pataphysicalisation (noun):* the process of pataphysicalising;
- Patadata (noun):* any data which have been pataphysicalised.

The idea of patadata is derived from the idea below:

*Physics → Metaphysics → Pataphysics
Data → Metadata → Patadata*

But what exactly does the process of pataphysicalisation include? The kinds of transformations we are thinking of could be, for example, replacing or adding to the query term(s) with synonyms, antonyms, opposites, syzygies, clinamens, etc. This can be done with the help of thesauri or dictionaries and ontologies. Whether we pataphysicalise our query term(s), the index or the results does not matter at this point. They are all possible and will maybe be done all at the same time (see Figure 2). We can consider the possibility of a *patametric index* rather than a parametric index or a *patasaurus* (pataphysical thesaurus/ontology).

Arguably, few other textual forms will have greater impact on the way we read, receive,

search, access, use and engage with the primary materials of humanities studies than the metadata structures that organize and present that knowledge in digital form
(Drucker 2009, 9).

Patadata will allow us to engage with digital knowledge in a more creative way even. If metadata help us *organise information semantically*, then patadata are for *organising information pataphysically*. If metadata are objective, then patadata are subjective. Drucker also points out that 'many information structures have *graphical analogies* and can be understood as diagrams that organise the relations of elements within the whole' (Drucker 2009, 16, emphasis added). So, maybe patadata could allow us to represent these graphical analogies in some way? An alphabetical list is a typical model for representing text data sets, for example. Or an otherwise ranked list, a tree structure, a matrix, a one-to-many relationship, etc. But is a ranked list really the best way to represent search results? *Ranking* itself seems unpataphysical. It contradicts the philosophy of pataphysics, although we can argue that this contradiction makes it pataphysical again. Maybe this dilemma can be solved simply by adopting another type of graphical analogy to structure the results, such as a tree structure instead of a ranked list.

In a traditional web search, ranking signals contribute to the improvement of the ranking process. These can be content signals or structural signals. *Content signals* are referring to anything that is concerned with the text and content of a page. This could be simple word counts or the format of text such as headings and font weights. The *structural signals* are more concerned about the linked structure of pages. They look at incoming and outgoing links on pages. There are also *web usage signals* that can contribute to ranking algorithms such as the clickstream. This also includes ideas such as the Facebook 'like' button or the Google '+ 1' button, which could be seen as direct user-relevance feedback.

Ranking can be done at different stages of the search process. Depending on how the index is formatted and what information can be

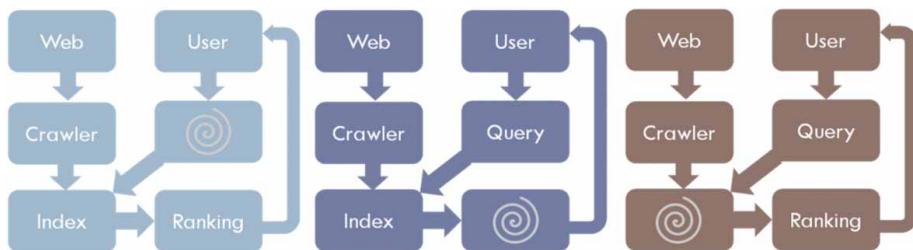


Figure 2. Three possibilities where pataphysicalisation can happen. © Fania Raczinski. Reproduced with permission.

pre-computed at that stage, the ranking algorithm evaluates every web page for *relevance* and returns them in order. There exist lots of different approaches on ranking, including PageRank (Brin and Page 1998) and HITS (Kleinberg 1999), which both analyse the link structure of the World Wide Web. They analyse the incoming and outgoing links on pages. PageRank, for example, assigns a numerical weight to each document, where each link counts as a vote of support in a sense. It is executed at *indexing time*, so the ranks are stored with each page directly in the index. HITS stands for ‘Hyperlink Induced Topic Search’, and its basic features are the use of so-called hubs and authority pages. It is executed at *query time*. Pages that have many incoming links are called authorities and pages with many outgoing links are called hubs.

Given a query term X, what is considered a relevant match, though? Do we simply return a list of web pages where X appears in the heading of each page? It is obviously not that easy. Several ranking signals are combined together; Google states that it uses over 200 signals including PageRank, and it personalises results using signals such as the web history and location (Google, n.d.).

What kinds of ranking signals do we need for our pataphysical web search tool? We could say that a page Y is relevant if it matches the patadata for query X. So, for example, Y would be a relevant result if it is a clinamen or syzygy to X. The more patadata matches there are, the higher the ranking, maybe. We don’t necessarily have to assign a numerical ranking value to each

page. Depending on how we structure our results page, that might not be necessary. Shuffling the results list or the results tree could be an option.

For example, let’s say our patadata are represented by a list of keywords that each stands for a pataphysicalisation of the original query term. This list is added to each item in the index:

Query = ‘Tree’

Patadata = [Tree(equivalent), Car(opposite), Paper(antinomy), Narwhal(anomaly), Book(syzygy), Venus Fly Trap(clinamen)]

Query = ‘Sun God Ra’

Patadata = [Sun God Ra(equivalent), Slave(opposite), Holiday(antinomy), Blue Balloon(antinomy), Pyramid(syzygy), Sphinx(clinamen)]

6 A new architecture for search

It is clear that any of these new algorithms, or ones that follow, will not be suitable for existing system architectures in IR research, and as such a new one will need to be defined. The question becomes whether or not the architecture itself can help enhance the chance of providing creative search results. If so, would it be possible to abstract this so that it can be used in other types of systems to help allow creative computing to flourish in areas where it may not have been possible before? This is a tall order, and one that is not likely to come soon; however, developing an architecture that is as generic as possible can only aid this task.

Hugill et al.

The concept of pataphysicalisation, using pataphysical methods to transform an object/idea, on the search request does appear to be an interesting place to start the search for a new architecture. While it can certainly constrain the possible amount of creative outputs and the general characteristics of such a system, it will give valuable insight into areas that need to be addressed by both the algorithms and architecture.

A *component-based system architecture* is therefore proposed to allow for greater flexibility in search engine development, whilst reducing the coupling between different parts of the system. This coupling tends to mean that for a new concept to be tried, large proportions of the entire search engine need to be redeveloped. The use of standard interfaces, for different types of components, would therefore allow a *generic harness* to handle the communication between these different components and provide a seamless service to the end user. The wiring of these components could be handled by a configuration file, therefore allowing people to build systems without needing any explicit programming skills.

Whilst this architecture itself does not explicitly improve the chance of creative results being returned, it will allow for new components to be tested in a full-scale environment in an *agile* way, and as such should allow for quicker testing. The *harness* is currently being developed, including a number of administration and monitoring tools inbuilt to aid analysis. The aim is to test the new architecture using a standard search engine and the Syzygy Surfer proposed by Hendler and Hugill (2011).

It is interesting to note how such a system could also be used in an educational environment to teach students how search engines work. Students could attempt to build systems using

pre-built components and see how different arrangements of such components affect the outcome. This is very similar to the way that the Massachusetts Institute of Technology has proposed to teach children how to program using the Scratch development environment.² More advanced students could also develop their own components to test out theories and improve their understanding of the base concepts of not just search engines but the various fields that play a role in information retrieval systems.

It is clear that this system could have great advantages outside of being a testbed for new ideas, allowing for the easy development of search engines to suit the needs of all new types of problems, without the need for specific development of every component. Whilst this idea is still within its infancy, the potential is strong and will be explored over time.

7 Discussion

Whilst developing a system that returns creative results to the end user has numerous advantages, the assumptions that are made about and the decisions we take for the user must still be considered. For example, presume that the user inputs a search request ‘The Cat in the Hat’ after reading a Dr Seuss book to their child, and the system employs an anomalous method on the query and searched ‘sunglasses’. Whilst there is logic to the new search request, it is anomalous to the initial request; if the user receives these results without being told what method was used, the results will appear *random*, and therefore are likely to be detrimental to the user. Therefore, the level of *interaction* the user has with the system and the *feedback* the system gives to the user on decisions it is making will have a large

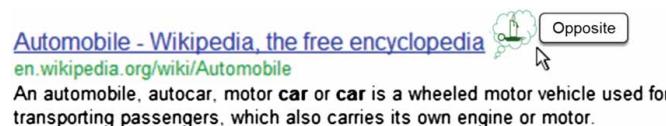


Figure 3. Example of an icon explaining the pataphysicalisation of a search result. © Fania Raczinski. Reproduced with permission.

influence on the overall effectiveness and appreciation of the search tool. A quick and simple solution to this problem would be to add an icon to the side of each search result which displays how the original query was pataphysicalised.

The image in Figure 3 shows an example of how this could be implemented. The little green candle (a reference to pataphysics in itself by the way) shows a pop-up note when hovered over with the mouse pointer. In this case, the original query could have been ‘tree’, and ‘car’ was returned as an opposite to that.

In the end, it comes to a point of being able to identify which of these factors will affect how the user perceives the results and which do not, and therefore give the system greater flexibility. This in itself is a huge undertaking, with which large quantities of empirical data will be required, and is therefore left for future work on the project.

8 Conclusion

Current information retrieval systems might be used for creative purposes. However, they do not directly provide creative results to their users; instead they focus on precise and relevant results only. Therefore, we argue that a new style of system is required. It is clear that the fundamental problem in this is that standard algorithms are not suited for these problems, with them considering a document to be groupings of words in traditional IR systems, and that an entire document falls under the same classifications in semantic IR systems.

The proposed concept for a pataphysical algorithm requires precise data structures to represent the transformations that have taken place during the pataphysicalisation, such as the patadata. The system’s index has to be adapted to accommodate this new type of data structure. It also needs to be flexible enough to allow algorithms to fit in at different stages or locations of the system; for example, the inverted index, ranking functions or query itself.

Whilst this new style of algorithm has been proposed, current architectures are not capable of supporting it. As such, a new, flexible

component-based software architecture has been proposed which will allow for a range of different style systems to be developed with little overhead, thereby improving the chance of creative outcomes occurring in a different way.

We have introduced the motivation and concept for a creative web search tool and discussed some of the major challenges a project like this faces. With web search being a major research and learning tool nowadays, it is imperative to think about *how* such a tool could be (ab)used. Ethical issues that arise through the provision of unexpected results, and the misunderstandings this could lead to, will be discussed in future work. Nevertheless, we believe that creative web search can facilitate inspirational learning through an exploratory search journey, and we hope our tool will provide just that.

Acknowledgements

We would like to thank Professor Jim Hendler at the Rensselaer Polytechnic Institute for his valuable thoughts on this work.

Notes

¹ Although note how the perplexing apostrophe that sometimes appears before the word ‘pataphysics undermines too literal an interpretation of this construction. Jarry only ever used the apostrophe on a single occasion, specifying that he did so ‘in order to avoid a simple pun’ (Jarry 1996, 21). What that pun might be has never been fully explained.

² See <http://scratch.mit.edu/>.

References

- Baeza-Yates, D. R., and D. B. Ribeiro-Neto. 2011. *Modern Information Retrieval: The Concepts and Technology Behind Search*. London: Addison Wesley.
- Baudrillard, J. 2007. *Pataphysics*. Translated by Drew Burk. CTheory. Accessed December 25, 2012. <http://www.ctheory.net/articles.aspx?id=569>.
- Benzécri, J. P. 1973. *L’Analyse des Données*. Paris: Dunod.

- Hugill et al.
- Boden, M. A. 2003. *The Creative Mind: Myths and Mechanisms*. London: Routledge.
- Bök, C. 2002. *'Pataphysics: The Poetics of an Imaginary Science*. Evanston, IL: Northwestern University Press.
- Borges, J. L. 2000. "The Analytical Language of John Wilkins." In *Selected Non-Fictions*, edited by E. Weinberger, 229–232. London: Penguin Books.
- Brin, S., and L. Page. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks and ISDN Systems* 30 (1–7): 107–117. doi: 10.1016/S0169-7552(98)00110-X.
- Chen, B., and X. Tai. 2009. "Semantic Retrieval Using Ontology and Document Refinement." *Proceedings of the Second Symposium International Computer Science and Computational Technology*, 302–309. Oulu, Finland: Academy Publisher.
- Colton, S. 2008. "Computational Creativity." *Artificial Intelligence and the Simulation of Behaviour Quarterly* 126: 6–7.
- Cooper, W. S. 1968. "Expected Search Length: A Single Measure of Retrieval Effectiveness Based on the Weak Ordering Action of Retrieval Systems." *American Documentation* 19: 30–41.
- Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. 1990. "Indexing by Latent Semantic Analysis." *Journal of the American Society for Information Science* 41 (6): 391–407.
- Dominich, S. 2000. "A Unified Mathematical Definition of Classical Information Retrieval." *Journal of the American Society for Information Science* 51: 614–624.
- Drucker, J. 2009. *SpecLab: Digital Aesthetics and Projects in Speculative Computing*. Chicago, IL: University of Chicago Press.
- Drucker, J., and B. Nowviskie. 2007. "Speculative Computing: Aesthetic Provocations in Humanities Computing". Chap. 29 in *A Companion to Digital Humanities*, edited by S. Schreibman, J. Unsworth, and R. Siemens, 431–447. Oxford: Blackwell. doi: 10.1002/9780470999875.ch29.
- Furnas, G. W., T. K. Landauer, L. M. Gomez, and S. T. Dumais. 1987. "The Vocabulary Problem in Human–System Communication." In *Communications of the ACM*, 30 (11): 964–971.
- Google. n.d. "Ten Things We Know to be True." Google Company Overview. Accessed December 25, 2012. <https://www.google.com/intl/en/about/company/philosophy/>
- Heilman, K. M., S. E. Nadeau, and D. O. Beversdorf. 2003. "Creative Innovation: Possible Brain Mechanisms." *Neurocase* 9 (5): 369–379.
- Hendler, J., and A. Hugill. 2011. "The Syzygy Surfer: Creative Technology for the World Wide Web." In *Proceedings of the 3rd International ACM Conference on Web Science*. June 14–17 2011, Koblenz, Germany. http://www.websci11.org/fileadmin/websci/Posters/3_paper.pdf
- Hendler, J., N. Shadbolt, W. Hall, T. Berners-Lee, and D. Weitzner. 2008. "Web Science: An Interdisciplinary Approach to Understanding the Web." In *Communications of the ACM* 51 (7): 60–69.
- Indurkha, B. 1997. "Computers and Creativity." Unpublished manuscript based on the keynote speech "On Modeling Mechanisms of Creativity" delivered at Mind II: Computational Models of Creative Cognition, Dublin, September 15–17.
- Jarry, A. 1996. *Exploits and Opinions of Dr Faustroll, Pataphysician*. Cambridge, MA: Exact Change.
- Kleinberg, J. M. 1999. "Authoritative Sources in a Hyperlinked Environment." *Journal of the ACM* 46 (5): 604–632.
- Koestler, A. 1964. *The Act of Creation*. London: Penguin.
- Manning, C. D., P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval*. New York: Cambridge University Press.
- Maron, M. E., and J. L. Kuhns. 1960. "On Relevance Probabilistic Indexing and Information Retrieval." *Journal of the ACM* 7 (3): 216–244.
- Motte, W. 1998. *Oulipo: A Primer of Potential Literature*. Lincoln, NE: University of Nebraska Press.
- Newell, A., J. G. Shaw, and H. A. Simon. 1963. *The Process of Creative Thinking*. New York: Atherton.
- Oulipo. 1973. *La Littérature Potentielle*. Paris: Gallimard.
- Oxford Dictionaries. n.d. Oxford University Press. Accessed December 25, 2012. <http://oxforddictionaries.com/definition/english/animal>.
- Poletti, N. 2004. "The Vector Space Model in Information Retrieval – Term Weighting Problem." Trento: Department of Information and Communication Technology, University of Trento.
- Salton, G., and C. Buckley. 1987. "Term Weighting Approaches in Automatic Text Retrieval." Technical Report. Ithaca, NY: Cornell University.

- Sternberg, R. J. 1988. *The Nature of Creativity*. Cambridge: Cambridge University Press.
- Sternberg, R. J. 1999. *Handbook of Creativity*. Cambridge: Cambridge University Press.
- Wong, S. K. M., and V. V. Raghavan. 1984. "Vector Space Model of Information Retrieval: A Reevaluation." In *Proceedings of the 7th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 167–185. Swinton: Bristish Computer Society.

Andrew Hugill is Director of Creative Computing at Bath Spa University. He is a composer and transdisciplinary researcher. He is the author of *Pataphysics: A Useless Guide* (MIT Press, 2012) and *The Digital Musician* (Routledge, 2008). He is the former Director of the Institute of Creative Technologies at De Montfort University.

Hongji Yang is Deputy Technical Director of the Software Research Technology Laboratory at De Montfort University. He is a Golden Core Member of IEEE Computer Society. He has been organiser for many international conferences, such as a Programme Co-Chair at IEEE International Conference on Software Maintenance 1999 (ICSM '99) and as the Programme Chair at IEEE Computer Software and Application Conference 2002 (COMPSAC '02). His research inter-

ests are software engineering, internet computing and creative computing. He has published intensively in the area of software evolution. He is currently developing a new journal, the *International Journal of Creative Computing* (IJCrC), to supply a unique forum for scholars to discuss how computing can support creativity.

Fania Raczinski is currently working on her PhD about pataphysical search algorithms, having completed a BSc in computer science (Europe) at the University of Leicester and an MSc in creative technologies from De Montfort University in Leicester. She has interests in written (programming) and spoken (natural) languages, creative computing, Semantic Web, pataphysics, culture and arts.

James Sawle is also working on his PhD and is looking into a new system architecture for search engines. He received his BSc (Hons) in computer science and mathematics from Durham University in 2009 before completing his MSc in software engineering in the Software Technology Research Laboratory at De Montfort University, Leicester. His research interests include IR systems, Semantic Web, computational creativity and pataphysics.



Creative Search Using Pataphysics

Fania Raczinski, Hongji Yang, Andrew Hugill
Institute of Creative Technologies

Pataphysics

- is the science of the **particular**
- examines the laws governing **exceptions**
- explains a universe supplementary to this one
- is the science of **imaginary** solutions
- is to metaphysics as metaphysics is to physics

"To understand 'pataphysics is to fail to understand 'pataphysics"
(Andrew Hugill's Useless Guide to 'Pataphysics)

2 / 10

Pataphysics

- | | |
|--|--|
| <ul style="list-style-type: none">• the clinamen or the chance• particular over the general• exceptions over the ordinary• the paradox or antinomy• simultaneous existence of mutually exclusive opposites | <ul style="list-style-type: none">• subjective• symmetry, duality• absurd anomalies• transcended reality or the absolute• epiphenomenalism• unexpected alignment, syzygy |
|--|--|

Definition of “animal”

- those that belong to the Emperor,
• embalmed ones,
- those that are trained,
• suckling pigs,
• mermaids,
• fabulous ones,
• stray dogs,
• those included in the present
classification,
- those that tremble as if they were
mad,
• innumerable ones,
• those drawn with a very fine
camelhair brush,
• others,
• those that have just broken a
flower vase,
• those that from a long way off look
like flies.

4 / 10

Creative Search Using Pataphysics

Aim: surprising, novel, inspiring, humorous search

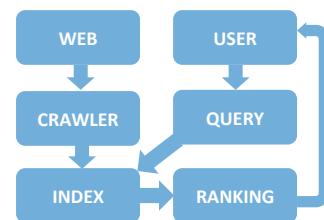
- Exploratory search, not information lookup
- Creative computing by using pataphysical algorithms
- Generate creative search results rather than relevant ones

Pataphysicalisation = applying pataphysical techniques to data

5 / 10

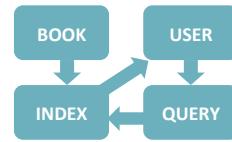
Information Retrieval

1. Information need
2. Query formulation
3. Retrieval process
4. Ranking
5. Results evaluation



Prototype

- Local search in book
- Text only, no images
- Tiny index
- No ranking
- Three algorithms



Exploits and Opinions of Dr Faustroll, Pataphysician (Jarry 1907)

7 / 10

Algorithms

Clinamen clear - leaf

Deliberate spelling errors

Syzygy clear – disappear – vanish

Query → synonyms → hypo/hyper/holonyms

Antinomy clear – allow light to pass through – opaque

Antonyms of query & antonyms of synonyms

8 / 10

Experimental Results

QUERY	CLINAMEN	SYZYGY	ANTINOMY
clear	altar, leaf, pleas, cellar	vanish, allow, bare, pronounce	opaque
solid	sound, valid, solar, slide	block, form, matter, crystal, powder	liquid, hollow
books	boot, bones, hooks, rocks, banks	dialogue, authority, record, fact	-
troll	grill, role, tell	wheel, roll, mouth, speak	-
live	love, lies, river, wave, size, bite	breathe, people, domicile, taste, see, be	recorded, dead

Conclusion

- Algorithms
- Larger Index
- Web
- Multimedia
- Evaluation
- Interpretation
- Apparent randomness
- Transparency

a pataphysical approach to making a creative exploratory search tool

10 / 10

References

- Bök, C. '*Pataphysics: The Poetics of an Imaginary Science*', Evanston, Illinois: Northwestern University Press, 2002
- Borges, J.L. The Analytical Language of John Wilkins in *Selected Non-Fictions*, Weinberger, E. (Ed.), London: Penguin Books, 2000
- Brotchie, A., Chapman, A., Foulc, T. and Jackson, K. (Eds.), '*Pataphysics: Definitions and Citations*', London Institute of 'Pataphysics, London: AtlasPress, 2003
- Brotchie, A. *Alfred Jarry: A Pataphysical Life*, Cambridge, MA: MIT Press, 2011
- Damerau, F.J. A Technique for Computer Detection and Correction of Spelling Errors in *Communications of the ACM*, 7(3), 1964, pp.171-176
- Foucault, M. (1966). *The Order of Things*. France: Editions Gallimard.
- Hendler, J. and Hugill, A. The Syzygy Surfer : Creative Technology for the World Wide Web, in *Proc. WebSci 2011*, ACM Press (2011).
- Hugill, A. '*Pataphysics: A Useless Guide*', Cambridge, MA: MIT Press, 2012
- Jarry, A. *Exploits and Opinions of Dr Faustroll, Pataphysician*, Cambridge, MA: Exact Change, 1997.
- Sutcliffe, A. and Ennis, M. Towards a cognitive theory of information retrieval, *Interacting with Computers*, vol. 10, pp. 321–351, 1998.

11 / 10

Creative Search Using Pataphysics

Fania Raczinski, Hongji Yang, Andrew Hugill

Institute of Creative Technologies

De Montfort University

Leicester LE1 9BH, UK

{fania, hyang, ahu}@dmu.ac.uk

ABSTRACT

This paper looks at defining, analysing and practicing how creativity can be applied to search tools. It defines creativity with respect to search and discusses how these concepts could be applied in software engineering using principles from the pseudo-philosophy of pataphysics. The aim of the proposed tool is to generate surprising, novel, humorous and provocative search results instead of purely relevant ones, in order to inspire a more creative interaction between a user, their information need and the application. A proof-of-concept prototype is described to justify the ideas presented before implications and future work are discussed.

Author Keywords

Creative computing; pataphysics; information retrieval; creative technologies.

ACM Classification Keywords

F.3.m. Logics and Meanings of Programs: Miscellaneous
H.3.3. Information Search and Retrieval: Search Process
J.5. Arts and Humanities: Literature

General Terms

Algorithms; Design, Experimentation.

INTRODUCTION

Imagine a web search engine that does not quite return the results you expect. For example, imagine you search for “animal” and the top three results are a list of animals in the Emperor’s possession, followed by instructions about embalming animals and information on a society for animal training. Google’s top search results for this query on the other hand return the webpage of an action sports lifestyle brand, the Wikipedia article and a BBC (British Broadcasting Corporation) page about animal videos. While there is certainly nothing wrong with Google’s results, they are simply not very inspiring. The first example of search

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honoured. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

C&C '13, June 17 - 20 2013, Sydney, NSW, Australia
Copyright 2013 ACM 978-1-4503-2150-1/13/06...\$15.00.

results is adapted from Jorge Luis Borges’s *Chinese Encyclopaedia* [4] which lists several creative definitions of the term “animal”. Whilst they might not provide the kind of information we were initially seeking (if we even had a clear idea of the kind of answers we wanted), they are still perfectly valid results for the query and might even provoke a smirk upon their encounter. These are the kind of search results we are aiming for; strange, creative, surprising, inspiring and possibly funny (which some would call irrelevant) yet perfectly valid.

Pataphysics can provide some useful techniques that are very suitable for creative computing. Hendl and Hugill first suggested the use of three of its principles: clinamen, syzygy and anomaly, in their “Syzygy Surfer” [15].

“The ambiguity of experience is the hallmark of creativity, that is captured in the essence of pataphysics. Traversing the representations of this ambiguity using algorithms inspired by the syzygy, clinamen and anomaly of pataphysics, using a panalogical mechanism applied to metadata, should be able to humanize and even poeticize the experience of searching the Web.” [15]

In the rest of this paper we will introduce creativity and pataphysics and explain how they are used for our algorithms and the general philosophy during the development. We then discuss some of the implementation details for our proof-of-concept prototype and speculate on users and uses of the tool. We conclude the paper with a short discussion on further work.

CREATIVITY AND PATAPHYSICS

Creativity

We define creativity as “the ability to use original ideas to create something new and surprising of value”. Here, we generally speak of creative *ideas* rather than *products*, since we believe creative products merely provide evidence of a creative process that has already taken place. Creativity is often divided into two types, one is a personal everyday type of creativity (P-creativity [2] or mini-c/little-c creativity [21]) and the other is a more eminent historical type (H-creativity [2] or Pro-c/Big-C creativity [21]). Margaret Boden further divides creativity into three categories [2, with some additional descriptions from 17, 21, 22], the concepts of which are also described in Kaufman & Beghetto’s Four-C model [21].

- Combinational creativity: making unfamiliar combinations of familiar ideas; juxtaposition of

dissimilar; bisociation; deconceptualisation, interpretive process of constructing and understanding

- Exploratory creativity: exploration of conceptual spaces; noticing new things in old spaces, interpretive process of constructing and understanding
- Transformative creativity: transformation of space; making new thoughts possible by altering the rules of old conceptual space, transformative learning

Boden also argues that creative ideas are surprising because they go against expectations and she believes that constraints support creativity and are even essential for it to happen. She says that *constraints map out a territory of structural possibilities which can then be explored, and perhaps transformed to give another one* [2]. This view supports our use of pataphysical concepts or constraints to enable creativity in search tools.

In many cases (especially of P-creative or mini-c types), both the originality and the value of a creative idea are evaluated using subjective or intrapersonal criteria [21]. Pataphysics, which represents an extreme form of subjectivity, is therefore a highly appropriate framework within which to encourage and enable creative thinking and operations.

Pataphysics

"To understand pataphysics is to fail to understand pataphysics." [16]

Pataphysics was invented by a group of French schoolboys in France in the 1880s. One of their number was the author and playwright Alfred Jarry (1873-1907) [20], who later developed the concept both in his celebrated *Ubu* plays and in his novels and speculative writings. In short (there are over 100 equally correct definitions [6]) it can be defined as follows:

- Pataphysics is the science of imaginary solutions,
- Pataphysics is the science of the particular,
- Pataphysics is the science of the laws governing exceptions and contradictions,
- Pataphysics is to metaphysics as metaphysics is to physics, and
- Pataphysics describes a universe supplementary to this one.

We argue that pataphysics can facilitate creative computing. A pataphysical grammar can consist of exceptions, syzygies, anomalies, clinamen, antinomies, contradictions, equivalents and imaginaries. Such concepts or constraints can influence the ways in which we may navigate and transform our conceptual space. Pataphysical concepts are likely to cause surprise and could therefore be considered unconventional and provocative.

The concept of the clinamen can be understood as an unpredictable swerve which Bök called *the smallest possible aberration that can make the greatest possible difference* [3]. One of the most famous examples of a clinamen is Jarry's *merdre* (the first word in his *Ubu* plays). He squeezed an extra 'r' into the French word *merde* (meaning *shit*) and translates into something like *pshit*.

A syzygy both surprises and confuses. The concept originally comes from the field of astronomy where it denotes the alignment of three celestial bodies. In a pataphysical context it usually describes a conjunction of things, something unexpected and surprising. Unlike serendipity, a simple chance encounter, the syzygy has a more scientific purpose. A typical instance is the pun, which Jarry called the *syzygy of words* [20]. Next to being intentionally funny, puns demonstrate a clever use (or abuse) of grammar, syntax, pronunciation and/or semantics, often taken to a quite scientific level, such that without understanding of what is said and what the intended meaning is, the humour of the pun might be lost.

The antinomy, in a pataphysical sense, is the mutually incompatible or paradox. Mutually contradictory opposites can and do co-exist in the pataphysical universe.

CREATIVE COMPUTING AND SEARCH

Creative Computing

The concept of creative computing has existed for some time but has not yet managed to evolve into a recognised discipline within computer science. Computational creativity, on the other hand, has emerged as a field within artificial intelligence research [18] and overlaps with creative computing ideas to some extent.

It is important to differentiate between the ideas of *creative computing* and *computational creativity*. Intuitively the former is about doing computations in a creative way, while the latter is about achieving creativity through computation. You can think of the latter falling into the artificial intelligence category (using formal computational methods to mimic creativity as a human trait, see also [18]) and the former being a more poetic endeavour of how the computing itself is done, no matter what the actual purpose of the program is.

As a good example of creative computing, consider the *International Obfuscated C Code Contest* [19]. The competition revolves around writing compilable/runnable code, while visually appearing as obfuscated as possible. They value unusuality, obscurity and creativity but expect contestants to follow the strict rules and constraints of the C programming language.

Examples of computational creativity are Simon Colton's *Painting Fool* [9] or Harold Cohen's *AARON* [8]; both are computer programs that paint pictures. Kurzweil's *Cybernetic Poet* [23] is a classic example of a program that produces poetry.

Our search tool can be seen from both perspectives and therefore somewhat lies in-between. We want to use creative techniques to come up with refreshing results to provide a counter-inspiration for the relevant results provided by Google or other mainstream Web search engines. We (are trying to) use creative techniques to build something that (hopefully) also has a creative purpose and value.

Search

In simple terms, a typical search process can be described as follows. A user is looking for some information so she or he types a search term or a question into the text box of a search engine. The system analyses this query and retrieves any matches from the index, which is kept up to date by a web crawler. A ranking algorithm then decides in what order to return the matching results and displays them for the user. In reality of course this process involves many more steps and levels of detail, but it provides a sufficient enough overview.

From the users' point of view the search process can be broken down into four activities [31] reminiscent of classic problem solving techniques [29]:

1. Problem identification = information need (IN),
2. Need articulation = IN in natural language terms,
3. Query formulation = translate IN into query terms, and
4. Results evaluation = compare against IN.

Searching can be thought of in two ways, information lookup (searching) and exploratory search (browsing) [11, 24]. A situation where an information need cannot easily be articulated or in fact is not existent (the user is not looking for anything specific) can be considered a typical case of exploratory search and describes the kind of search that is most suited to our proposed tool.

Most big search engines like Google, Baidu or Bing focus on usefulness and relevance of their results. [13, 1, 26] Google uses over 200 signals [14] that influence the ranking of web pages including their original PageRank algorithm [5]. We can only speculate whether these signals also take into account any creative factors due to their secrecy. Other search engines like YossarianLives (currently in alpha release) [32] concentrate on purely abstract concepts like metaphors for their search algorithms.

Any information retrieval process is constrained by factors like subject, context, time, cost, system and user knowledge [25]. Such constraints should be taken into consideration in the development of any search tool. A web crawler needs resources to crawl around the Web, language barriers may exist, the body of knowledge might not be suitable for all queries, the system might not be able to cater for all types of queries (e.g. multi-word queries), or the user might not be able to understand the user interface, and many more. It is therefore imperative to eliminate certain constraining

factors (for example by targeting a very specific audience or filtering the amount of information gathered by a crawler from web pages).

PROPOSED SYSTEM

The general concept of the project described in this paper is *pataphysical web searching* and the following three points summarize its main aims:

- search the Web for suitable answers to a given query,
- return results as a list or a mixture of data structures, and
- present pataphysical results (rather than relevant ones).

Principles

The essence of the proposed search tool lies in its algorithms which make the difference to traditional search engines. The philosophical ideology behind the tool is fundamentally different. Our system will still consist of the main components typically found in Web search engines (crawler, index and ranking) but they will have slightly different inner workings and target a different audience of users.

To link back to some of the creative, pataphysical concepts we have discussed earlier, let us put some of the ideas for our tool into perspective. The constraints for our conceptual space are the pataphysical rules that we want to apply to our data. We use those rules to explore, combine and transform our space; giving us the flexibility and freedom we need to find interesting results.

We developed the idea of *pataphysicalising* data as the process of applying such pataphysical rules in order to produce creative search results. This pataphysicalisation process forms a central component of our system (see Figure 1) and influences all areas of the search tool.

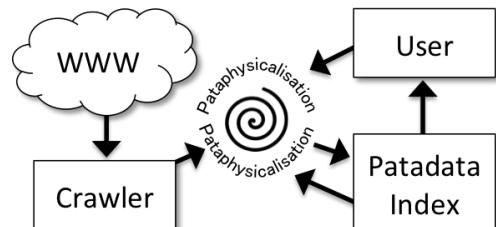


Figure 1. Pataphysicalisation as a central component

Our index will contain what Handler and Hugill have called *patadata* [15]. Patadata is to metadata as metadata is to data - inspired by one of the definitions of pataphysics: *that which is above that which is after physics* [20]. This suggests that patadata provides another layer of information above information. If metadata helps us organise information semantically then patadata is for organising information pataphysically. If metadata is objective then patadata is subjective and that is precisely what pataphysics is for.

Prototype

The prototype described here (see Figure 2) was developed as a proof-of-concept tool to demonstrate some example search results using pataphysical algorithms. In this case the results are limited to the text of Alfred Jarry's *Exploits and Opinions of Dr. Faustroll, Pataphysician* [20] and only the main algorithmic functionality of this prototype is discussed here.

Pataphysical Search!

Clinamen - 19 pataphysicalised reverberations found for: "clear"

...s webbed feet . Pitiful **pleas** swim up against the stream...
...was impossible to enter the **cellar** due to the flooding thereof...
...the length of a cabbage **leaf** , paying no attention to...
...with Mendacious in the **lead** . Since the episcopal nature...
...gigantic , black , mass **altar** . At the blunt point...
...blue diadem . He was **clad** , too , in sky...
...TO ARTICLE 819 In this **year** Eighteen Hundred and Ninety - ...
...route ; but such a **leap** is not within everyone '...
...any need for Faustroll to **fear** that his scalp hair...
...which followed us and those **near** us which crossed our path...
..., maple , oak , **cedar** , sorb wood and poplar...
...yellow skin , his face **clean** - shaven , apart from...
...exclaims : Never , I **swear** , shall I forget the...
...the pale forehead , the **dear** face , this terrible little...
...LAURENT TAILHADE We could already **hear** bells - as loud as...
...content to be black . **Fear** , turning away its head...
...- Nage 's right **ear** and four of his teeth...
...doctor informed me , discern **clearly** through these mirrors those ultraviolet...
...WITCH Her hump to the **rear** , belly to the fore...

Syzygy - 22 pataphysicalised reverberations found for: "clear"

...French language , he could **pronounce** fairly correctly a few words...
...as his tonsure , laying **bare** the optic nerve and the...
...few quarter - centuries will **determine** their periods . Soon ,...
...allows air and steam to **pass** through but is impermeable to...
...zero , if these dimensions **vanish** on both sides of our...
...the Mayor , who did **certificate** the original thereto ; within...
...- four hours , to **pay** to the claimant into my...
...choice of the two asphyxiating **make** - ups called white hanged...
...hereunder . The sale will **take** place on whatever day shall...
...web , leaves the holes **empty** - the number of which...
...guide had given him absolutely **free** ; one represented realistically ,...
...'s day ... serene **countenance** ... supreme image , so...
...usual example of water , let us reflect , in this...
...our dead drunk credits and **gain** , without wasting our talent...
...as far as I could **judge** , understood these prodigies very...
...examine any disturbances which this **change** in size might involve in...
...except perhaps in the **country** , he will rarely see...
...the Snout , to the **clear** anxiety of those present....
...found by experiment that the **benefit** extends only to those whose...
...meshes are wide enough to **allow** the passage of a large...
...the globe by attraction , **permit** me , I pray
.... And I ' ll **declare** He ' s mooning up...

Antinomy - 1 pataphysicalised reverberations found for: "clear"

...colors were locked in an **opaque** box ; until he was...

Figure 2. Screenshot of search results for the query "clear"

In short, the prototype's workflow can be described as follows:

- 1) tokenise text and remove stopwords to build index,
- 2) query triggers the three pataphysical functions,
- 3) each function finds matches for query as described above,

- 4) retrieve some words before/after match for context, and
- 5) return list of resulting sentences.

The three functions inspired by pataphysics (clinamen, syzygy and antinomy) are described in more detail in the next section. Figure 2 shows a screenshot of the resulting list of results for the query *clear*. The specific results for each of the three methods are simply a few words surrounding the pataphysicalised query term from within the book, which does not necessarily represent complete sentences but simply provides some context for the result.

The same principles and algorithms can be applied to different types of media, for example images or video and even sound. The complete tool would include a mixture of different types of media in its results with various styles of displaying them.

Algorithms

The clinamen function uses the Damerau-Levenshtein algorithm [10], which measures the distance between two strings (with 0 indicating equality), to find words that are similar but not quite the same. The distance is calculated using insertion, deletion, substitution of a single character, or transposition of two adjacent characters. We are basically asking the program to return matches (v) that are of distance two or one to query term t , meaning they have two or one spelling errors in them (see Equation 1). While we only return matches that actually appear in the book (i.e. they exist in the index), and by doing so eliminate the introduction of new words like Jarry's *merdre*, the swerve or aberration is still evident.

$$\text{clinamen}(t) = \{ v : 0 < \text{dameraulevenshtein}(t, v) \leq 2 \}, \text{ for } v \in V \quad (1)$$

For the syzygy function, we made use of the WordNet lexical database [30] using the NLTK python library [27] to find suitable results. Specifically, as shown in Equation 2, the algorithm fetches the set of synonyms (synsets) for query term t first and then finds any hyponyms, hypernyms or holonyms for each of those (each of which denotes a sort of relationship or membership with its parent synonym). We then return a list of all of those related terms if they appear in the original vocabulary of the text (index V). This approach mimics the syzygy alignment of three words in a line mentioned earlier (query \rightarrow synonym \rightarrow hypo/hyper/holonym).

$$\begin{aligned} \text{syzygy}(t) &= \{ h : h \in \text{union}(t) \wedge \exists h \in V \} \\ \text{union}(t) &= \text{hypo}(t) \cup \text{hyper}(t) \cup \text{holo}(t) \\ \text{hypo}(t) &= \{ h : h \in \text{hyponyms}(s) \} \\ \text{hyper}(t) &= \{ h : h \in \text{hypernyms}(s) \} \\ \text{holo}(t) &= \{ h : h \in \text{holonyms}(s) \} \\ \text{syno}(t) &= \{ s : s \in \text{synonyms}(t) \} \\ \text{for } s \in \text{syno}(t) \end{aligned} \quad (2)$$

For the antinomy function we simply made use of WordNet's antonyms (opposites) (see Equation 3). We first get all synonyms for query term t , find any antonyms for

those and return any that also appear in the index V . Naturally, not all words have an opposite, especially given WordNet's limited vocabulary of around 150,000 words, but a pataphysical antinomy should still be able to find a match. This is a big shortcoming of our prototype at this point. A better thesaurus or a larger index (e.g. based on more than one book – or, of course, the Web) could improve this function drastically.

$$\begin{aligned} \text{antinomy}(t) &= \{ h : h \in \text{anto}(t) \text{ and } \exists h \in V \} \\ \text{anto}(t) &= \{ h : h \in \text{antonyms}(s) \} \\ \text{syno}(t) &= \{ s : s \in \text{synonyms}(t) \} \end{aligned} \quad (3)$$

for $s \in \text{syno}(t)$

Table 1 shows some example matches produced by the three algorithms described above. While the syzygy and antinomy methods both work in a semantic manner, the clinamen function is purely syntactical, which becomes very obvious when seeing the different results side by side as in the table. Relying on WordNet's limited vocabulary means less matches can be found mostly because the text of *Faustroll* uses a very specific language and not always matches that found in the thesaurus. On the other hand, it illustrates the breadth of vocabulary used by Jarry in his writing nicely, as only those results are returned that actually appear in the book.

	clinamen	syzygy	antinomy
clear	altar, leaf, pleas, cellar	vanish, allow, bare, pronounce	opaque
solid	sound, valid, solar, slide	block, form, matter, crystal, powder	liquid, hollow
books	boot, bones, hooks, rocks, banks	dialogue, authority, record, fact	-
troll	grill, role, tell	wheel, roll, mouth, speak	-
live	love, lies, river, wave, size, bite	breathe, people, domicile, taste, see, be	recorded, dead

Table 1. Example search results. Queries are shown in column one, algorithms used in row one.

POSSIBLE APPLICATIONS

In this section we consider the possible uses and applications for the proposed creative search tool.

Our target audience is not quite as broad as that of a general search engine like Google. Instead, we aim to specifically cater for users who can appreciate creativity or users in need of creative inspiration. Users should generally be educated about the purpose of the search tool so they are not discouraged by what might appear to be *nonsensical* results. Potential users could include artists, writers or poets

and anybody who is looking for out-of-the-box inspirations or a refreshingly different search engine to the norm.

Uses

There are many ways a pataphysical search tool could be used across disciplines.

In literature, for example, it could be used to write or generate poetry, practically or as a simple aid for inspiration. We are not limited to poetry either; novels, librettos or plays could benefit from such pataphysicalised inspirations. One can imagine tools using this technology that let you explore books in a different ordering of sentences (a sort of pataphysical journey of paragraph hopping), tools that re-write poems or mix and match them together. Even our simple prototype shows potential in this area and could be even more powerful if we extended it to include more source texts, for example the whole set of books contained in Faustroll's library ([20] and also [12]). A richer body of texts (by different authors) would also produce a much larger index which would then possibly find many more matches through WordNet and end in a more varied list of results.

From a computer science perspective this pataphysical approach could be added to the many algorithms used by traditional search engines for purposes like query feedback or expansion (e.g. “did you mean ...” or “you might also be interested in ...”). Depending on how creative we want the search engine to be, the higher we would rank the importance of this particular algorithm. One of the concepts related to the search tool, namely patadata, could have an impact on the development of the Semantic Web. Just as the Semantic Web is about organizing information semantically through objective metadata, patadata could be used to organize information pataphysically in a subjective way.

Our prototype tool is already being used in the creation of an online opera, provisionally entitled *from [place] to [place]*, created in collaboration with The Opera Group¹, an award-winning, nationally and internationally renowned opera company, specialising in commissioning and producing new operas. In particular, it is being used to create the libretto for one of the virtual islands whose navigation provides the central storyline for the opera. The opera will premiere in 2013, and will continue to develop thereafter, deploying new versions of the tool as they appear.

Evaluation

Evaluating creative software is not an easy task and there are no standard approaches. Pease and Colton [28] divide it into two notions:

¹ www.theoperagroup.co.uk

- whether an idea or artefact is valuable or not, and
- whether a system is acting creatively or not.

Following this approach, we would need to investigate each individual search result in terms of its value and creativity. This could be done by user ratings or satisfaction questionnaires. Rather than measuring the success of individual results though, we could also look at evaluating them as one set instead, similar to the blind side-by-side comparisons by the Bing search engine or the selective omission of a certain number of results by search engine MillionShort².

The way we display and label results produced by the tool can influence how the user perceives them. Our current prototype for example separates the results into its three components but we could have equally just mixed them all together. It is not always clear how each result connects to the initial query, even if we identify through which algorithm a result has been obtained. These keywords (syzygy, clinamen and antinomy) might not be helpful to users unfamiliar with the concept of pataphysics anyway and might therefore appear rather nonsensical. Whilst there is a clear logic to each search result, they might appear anomalous to the user's expectations if he received these results without knowing the philosophy of the search tool. The results could possibly appear random then, and could therefore likely be detrimental to the user. The level of interaction between the user and the system and the feedback the tool gives to the user on its internal processes will have a large influence on the overall effectiveness, perception and appreciation of the tool.

The less obvious the processes in the background are for the user, the more difficult it might be to appreciate the search results. On the other hand, too much transparency could spoil much of the experience. After all, explaining a joke kills it. The issue therefore becomes a question of finding just the right level of transparency to satisfy curious users while at the same time not spoiling the seemingly serendipitous experience of others.

FURTHER WORK AND CONCLUSION

We are just beginning to understand the potential of such a creative search tool and its implications. There is much research left to be done, specifically in developing more and different kinds of search algorithms and evaluating the results we obtain. We could try to implement different algorithms or different pataphysical concepts within our existing prototype or built a different system altogether. We could also try to implement a fully functioning Web search engine using the algorithms described in this paper and then compare the two different types of results. It could be interesting to investigate how users perceive and use search

² www.bington.com and www.millionshortiton.com

results produced in either the book based search or an open Web based search.

Before we go into further development and programming though, it might be worth studying, evaluating and interpreting the results produced by the prototype presented in this paper. An evaluation framework for pataphysical search results is under development. A study of user's reactions to the prototype could be very interesting as well and will be part of future work in this project.

Finally, to summarise, in this paper we have introduced a new approach for a creative search tool that uses pataphysics as an underlying philosophy. We have explained how pataphysics can be applied to search algorithms in order to produce interesting results with a humorous twist. Our initial experiments within a limited domain have shown that the generated results could indeed be interpreted as being novel, surprising and useful. We have also briefly discussed ideas for applications of the tool and issues that may trigger possible further research in the field of creative computing. We have also presented some thoughts on evaluation of our tool and future work.

ACKNOWLEDGMENT

Jim Hendler at the Rensselaer Polytechnic Institute has provided valuable advice for this project.

REFERENCES

1. Baidu *The Baidu Story* [online] available at <http://ir.baidu.com/phoenix.zhtml?c=188488&p=irol-homeprofile> [Accessed 12 May 2013]
2. Boden, M.A. *The Creative Mind: Myths and Mechanisms*, London: Routledge, 2003.
3. Bök, C. *'Pataphysics: The Poetics of an Imaginary Science*, Evanston, Illinois: Northwestern University Press, 2002.
4. Borges, J.L. The Analytical Language of John Wilkins in *Selected Non-Fictions*, Weinberger, E. (Ed.), London: Penguin Books, 2000.
5. Brin, S. and Page, L. The anatomy of a large-scale hypertextual Web search engine in *Computer Networks and ISDN Systems*, 30(1-7), 1998, pp.107-117.
6. Brotchie, A., Chapman, A., Foulc, T. and Jackson, K. (Eds.), *'Pataphysics: Definitions and Citations*, London Institute of 'Pataphysics, London: AtlasPress, 2003.
7. Brotchie, A. *Alfred Jarry: A Pataphysical Life*, Cambridge, MA: MIT Press, 2011.
8. Cohen, H. *AARON* [online] available at <http://www.kurzweilcyberart.com/aaron/history.html> [Accessed 12 May 2013]
9. Colton, S. *The Painting Fool* [online] available at <http://www.thepaintingfool.com/> [Accessed 12 May 2013]
10. Damerau, F.J. A Technique for Computer Detection and Correction of Spelling Errors in *Communications of the ACM*, 7(3), 1964, pp.171-176.

11. de Vries, E. Browsing vs Searching, *OCTO report* 93/02, 1993.
12. Fisher, B. *The Pataphysician's Library*, Liverpool University Press, 2000.
13. Google Company Overview [online] available at <https://www.google.com/intl/en/about/company/> [Accessed 12 May 2013]
14. Google Ten things we know to be true [online] available at <https://www.google.com/intl/en/about/company/philosophy/> [Accessed 12 May 2013]
15. Hendler, J. and Hugill, A. The Syzygy Surfer : Creative Technology for the World Wide Web, in *Proc. WebSci 2011*, ACM Press (2011).
16. Hugill, A. *'Pataphysics: A Useless Guide*, Cambridge, MA: MIT Press, 2012.
17. Indurkhy, B. *Computers and Creativity*, Unpublished Manuscript, Based on the key note speech On Modeling Mechanisms of Creativity, delivered at Mind II: Computational Models of Creative Cognition, Dublin, Ireland, September 15-17, 1997.
18. International Conference of Computational Creativity [online] available at <http://www.computationalcreativity.net/iccc2013/> [Accessed 12 May 2013]
19. International Obfuscated C Code Contest [online] available at <http://www.ioccc.org/> [Accessed 12 May 2013]
20. Jarry, A. *Exploits and Opinions of Dr Faustroll, Pataphysician*, Cambridge, MA: Exact Change, 1997.
21. Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology*, 13(1), 1–12.
22. Koestler, A. *The Act of Creation*, London: Penguin Books, 1964.
23. Kurzweil, R. *Cybernetic Poet* [online] available at http://www.kurzweilcyberart.com/poetry/rkcp_overview.php [Accessed 12 May 2013]
24. Marchionini, G. From finding to understanding, *Communications of the ACM*, vol. 49, no. 4, pp. 41–46, 2006.
25. Marchionini, G. and Shneiderman, B. Finding facts vs. browsing knowledge in hypertext systems, *Computer*, vol. 21, no. 1, pp. 70–80, 1988.
26. Microsoft *Bing Fact Sheet* October 2012[online] available at <http://www.microsoft.com/en-us/news/download/presskits/bing/docs/MSBingAll-UpFS.docx> [Accessed 12 May 2013]
27. NLTK Project, 2012, *Natural Language Toolkit* [online] available at <http://nltk.org/> [Accessed 12 May 2013].
28. Pease, A. and Colton, S. On impact and evaluation in Computational Creativity: A discussion of the Turing Test and an alternative proposal in *Proc. AISB 2011 Symposium on AI and Philosophy*.
29. Polya, G. *How To Solve It*, 2nd ed. Princeton, New Jersey: Princeton University Press, 1957.
30. Princeton University, 2010, *About WordNet* [online] available at <http://wordnet.princeton.edu> [Accessed 12 May 2013]
31. Sutcliffe, A. and Ennis, M. Towards a cognitive theory of information retrieval, *Interacting with Computers*, vol. 10, pp. 321–351, 1998.
32. YossarianLives *The Metaphorical Search Engine* [online] available at <http://www.yossarianlives.com/> [Accessed 12 May 2013]

TDC Talk Andrew Hugill

A Framework for Creativity in Search Results

James Sawle, Fania Raczinski, and Hongji Yang
Institute of Creative Technologies
De Montfort University
Leicester, UK
{jsawle, fania, hyang}@dmu.ac.uk

Abstract—Although trying to define creativity has been a hot area of research in many fields, the field of information retrieval has remained under developed. Over the report we attempt to define a structural definition of creativity which could be applied to search results in order to aid users in their creative endeavours. After defining creativity for search, we have then devised a simple metric based upon it, to show that there is a need for this research. The results, whilst positive, could be interpreted as a poor definition of creativity, and as such this is a sounding paper for future work.

Index Terms—information retrieval; computational creativity

I. INTRODUCTION

Over the past decade search has been focused on returning the smallest number of results which correlate to the user's information need. This has been a logical trend to pursue, as 92% of people use the internet as their first port of call when looking for everyday information [1].

However, this has meant that the creativity inspired by 'surfing the web' has over time slowly diminished. This research is not advocating the end of document based search; however, we propose that a new search engine architecture, which aims to inspire the creativity of its users, can only be beneficial to the landscape of the world wide web.

Over the course of the paper, we define what we mean by the creativity of a search result, with respect to a single result as well an entire set. The concepts presented in this paper, are inspired by 'Pataphysics, a pseudo-philosophy defined as "the science of imaginary solutions, which symbolically attributes the properties of objects, described by their virtuality, to their lineaments" [2].

The rest of the paper is organised as followed. Section 2 explores definitions of creativity from both computer science and psychology. In Section 3, we outline a general definition of creativity in search, which can be used to create a metric. Section 4 will see a simplistic metric to be used for the purpose of evaluating the concept as well as some experimental data.

II. DEFINITION OF CREATIVE SEARCH

A. A Framework to Base Creativity Upon

Creativity is a subjective topic, with different people defining the creative worth of a piece of information differently; however, Newell, Shaw and Simon [3] devised a definition based upon four criteria to categorise the creativity of a given solution or answer.

- 1) The answer is novel and useful (either for the individual or for society)
- 2) The answer demand that we reject ideas we had previously accepted
- 3) The answer results from intense motivation and persistence
- 4) The answer comes from clarifying a problem that was originally vague

Each of these criterion for creativity approach a definition from a different perspective. Whilst trying to relate this to information retrieval, it should be simple to see that criterion 1 relates to the goal of the search, whilst criterion 4 relates to the information need, or starting point. What may be less obvious however, is that criterion 3 relates to the scale of the search and hence the number of dead ends that may be encountered and that criterion 2 suggests which search paths should be avoided whilst looking for creative results.

Whilst this framework gives us a very high level definition of creativity, it is hard to apply it in its current form. Through applying some of the more prevalent techniques used in the field of computational creativity, we can attempt to reduce this down into a more precise definition.

B. P-Creativity and H-Creativity

Boden [4] defines that there are two forms of creativity, P-creativity and H-creativity. P-creativity or 'psychological' creativity, is an idea or solution that is new to the person who came up with it. An idea that represents 'historical' creativity, H-creativity, on the other hand, is one which has not been thought of by anybody before and can therefore be deemed a historical-sociological category [5]. H-creativity is subsequently a special case of P-creativity which many people consider to be the more important of the two, as this is what drives forward human knowledge.

When we relate these concepts to search results we end up with some interesting outcomes.

1) *Single Search Result*: A single search result is most likely to be P-creative or neither. This is because, for it to be H-creative, there must be some logic in the document that nobody else has noticed, or drawn the same conclusions from different information. For the single result to be neither P-creative or H-creative, the user must have a thorough understanding of the topic, and the result must add no new information.

2) *Set of search results*: A set of search results is most likely to be P-creative. It is highly unlikely that a user would

have explored every possible creative avenue over a set of results, unless the set is not of a trivial size. But by the same logic, if a large range of ideas are contained, it is unlikely that the set will be H-creative, as somebody is likely to have linked them together.

The question becomes, is there a link between maximising the chance of something being P-creative and H-creative or is the link more subtle. Or is it enough for a search engine to try and improve the chances of P-creativity for a user.

C. Exploratory and Transformational Creativity

Boden [4] goes on to define the concepts of exploratory and transformational creativity. She defines exploratory creativity to be the exploration of a space of partial and complete possibilities. This therefore suggests that there are rules that confine this space. If we were therefore to alter the rules that define the space, and subsequently alter the space that we are exploring, this is defined as transformational creativity [6].

Whilst this does give us a nice slant to look at creativity, comparing the trade-off of traditional problem spaces compared to augmented ones, this is very difficult to model, combined with the fact that the solutions found by tweaking the rules that confine the space can easily rule out the solution in the traditional space [7].

D. Bisociation

Bisociation makes a distinction between the routine skills of thinking on a single ‘plane’, and the creative act, which operates on more than one plane [8]. This means, with Koestler’s definition, that we must define creativity as a set of results such that they are simultaneously associated with two habitually incomparable contexts.

It is clear to see how this model extends from that of Boden’s theory of exploratory and transformational creativity. The fact that more than one ‘plane’ must be considered will force a transformational process to occur. However, unlike transformational creativity, both processes must be considered, the exploratory and transformational. Subsequently, we should not end up with a solution that can’t exist within the rules defined by the original problem, even if we transcend into transformational creativity, as long as we finish the process in the plane that we started in.

E. Conceptual Blending

The idea of combining different thought processes, whilst more elegant than transformational creativity, does not give us a nice definition that applies to search results as well as tying in with our underlying philosophy. Conceptual blending is a step closer. This general theory of cognition, formally called Conceptual Integration Networks [9], allows us to look at a number of different dataspaces, and attempt to ‘blend’/merge them in such a way that the new dataspace tries to simulate how we use large amounts of information and bring it together to form new ideas.

F. Combinatorial Creativity

Both of the above concepts fall into the general category of combinatorial creativity. This is a logical assumption of modelling creativity, as people tend to come up with solutions by first looking at new combinations of currently existing ideas. This therefore allows us to consider the idea of creativity as a search process through the space of all possible combinations, therefore this fits into the idea of search engines.

Whilst conceptual blending explores the idea of combining different thought processes and bisociation, looking at different planes of creative thought; let us consider the idea of placing the data itself into different concepts, enabling us to get the following areas of combinatorial creativity to explore with respect to creative search based upon philosophical.

- Placing a familiar object in an unfamiliar setting or placing an unfamiliar object into a familiar setting.
- Blending two superficially different objects or concepts
- Comparing a familiar object to a superficially unrelated and semantically distant concept
- Searching through a number of different concepts that are related to each other but could be considered as swerving away from the original concept. This is based upon Epicurus’s theory of clinamen from his doctrine of atomism [10].

III. DEFINITION

The above definitions, allow us to define creativity in search results with pre-existing concepts agreed by the academic community.

It is clear, that in the case of search results, we still have the issue of a group of results providing greater creative inspiration to one user than another. This tends to be a problem with most metrics, the problem of objectiveness vs subjectiveness. With subjectiveness being a quality that is important, it means that we have a problem getting repeatable results. We therefore need to build a definition that is as objective as possible, whilst not overlooking some of the dynamic properties that it may be possible to model.

At this stage it is important to stress that this is not an attempt to model the creative process, but to give a model for how useful a set of results might be in inspiring creativity.

A. A Single Result

It is intuitive for us to start with a single result. Whilst maximising the possibility for a single result being H-creative, it is very unlikely that this will be the case with a full set of results. The issue becomes, measuring how P-creative an individual result is to a search result.

It seems sensible to assume, that if a result has no relevance to the search request, then the result will have no chance of inspiring P-creativity. The more information about the search request a single result has, increases the chance of a result inspiring P-creativity, therefore using relevance metrics.

B. Set of Results

To improve the chances of inspiring creativity, a group of related results which discuss a number of different areas of the topic would logically improve the quality of the results. As stated, if we maximise the breadth of information of a single result it would improve creativity, we should therefore attempt to do the same across the entire set.

The issue however, is that the majority of users do not look past the top 10 search results [11]. Whilst this is unlikely to be the case for people using a search engine targeted at inspiring creativity, it does make sense to try to reduce the overall amount of data provided. We must therefore penalise repetition in the results provided, forcing a more diverse set of results.

This can be taken a step further by only considering a certain number of results and ignoring the ordering, because there is no simple way to define how ordering affects the creative process. With a lack of defined ordering, it means that having endless results would be tedious and counter-productive. Whilst we have no strong view on the exact number of results that should be considered, we believe that it should not be substantially greater than 10, for the reason discussed above.

The way that each result is provided to the user will affect how the user perceives the results. A diversity of different document types, e.g., text, images, sound, we believe would improve the quality of creativity inspiration.

C. Results as a Set of Sets

We could extend this concept to the next logical step of returning results as a set containing multiple related sets of results. In this analogy, each of the inner sets could relate to an individual concept related to the information need, and a clearer relationship between concepts, how they relate to each other and how the results represent the concept they are contained within would exist.

The question becomes how we measure the creative quality of this type of result. Due to the structure of the results, we can attempt to model the creativity in different levels allowing us to try and abstract the problem as much as possible.

Due to the fact that this is not a method that is currently used to return search results, we shall not explore it further at this point in time. However, we believe that this would be a logical way to return results in the future.

IV. EXAMPLE METRIC

As the above definition is meant as a guideline for defining creativity, this section attempts to give a real world example. The metric defined below is a contrived example to show how it could be applied with current search results.

A. Algebraic Definition

Taking the definition defined in Section III-B, we have derived the following abstract metric.

Let us define a query as q , a set of results as r and an individual result as d . As such $r = \{d_1, d_2, d_3, \dots, d_i\}$ where i is the number of results examined.

For the quality of a single result, we shall define $P(q, d)$ as a measure between 0 and 1, where 1 is the optimal value.

To reduce the amount of data duplication in the returned results, we shall define $D(r)$ which has to return a value between 0 and 1, where 0 means that no data is duplicated.

Let us define $T(r)$ as a way to weigh the final outcome of the metric to ensure that a diverse set of document types are returned. This metric will return 1 if a satisfactory balance is returned, and 0 if only a single document type is returned.

We can therefore compile these measures into a single metric, the Search Creativity Metric or SCM:

$$SCM = T(r) \cdot \frac{1}{i} \sum_{j=1}^i P(q, d_j) \cdot (1 - D(r))$$

As such, this metric will always return a value between 0 and 1, with 1 being the optimal value.

B. Fleshy Out the Metric

To enable us to apply any experimental data to the metric, we must first give definitive definitions to each of the functions provided above, $P(q, d)$, $D(r)$ and $T(r)$.

1) $D(r)$: As this measures the number of duplicate results in a return set, we can easily define it as the number of results that have a majority of information that is contained within another article. This allows the following definition

$$D(r) = \frac{\text{Number of results with data in previous results}}{\text{Number of results}}$$

As we relate each result to the previous results in the list, the results must always be $0 < D(r) \leq 1$. This makes sense, as even if all of the results are identical, there may still be some creative inspiration contained in the first result. This also allows us to penalise results heavily for leaning too much on one area of information.

2) $T(r)$: As with $D(r)$, we need to define this measure so that we penalise for a lack of diversity, but do not eradicate all results, as this would not reflect the possible creative quality of the information returned.

For this definition, we will need to leverage on the definitions provided earlier. Let i is the number of results within the result set r . We can therefore define n to be the number of different result types that are returned, and σ to be the standard deviation of the number of results for each media type. It is interesting to note that $0 \leq \sigma < \frac{i}{2}$, such that $\sigma = \frac{i}{2}$ means that the results are biased to only one result.

$$T(r) = \begin{cases} 1 - \frac{2 \cdot \sigma}{i} & : n > 2 \\ 0.1 & : n \leq 2 \end{cases}$$

For the case of this sample measure, we have defined that for a result set to be considered to be broad enough, that it must contain at least 3 different media types. This measure has no empirical backing.

TABLE I: Example Results

URL	$P(q, d)$	Reason
www.unicorn-darts.com	0.3	Company called Unicorn due to the single point on a dart
en.wikipedia.org/wiki/Unicorn	1.0	Contains mythology as well as related animals
www.unicorn-grocery.co.uk	0.0	No relation to unicorns
www.unicorn-theatre.com	0.25	Uses the mythology of unicorns to draw children into theatre
http://katemckinnon.files.wordpress.com/2008/07	0.7	Image of a unicorn but purely as a distraction from the rest of the article
http://www.unicorn-centre.co.uk/	0.9	Unicorn mythology about the soul applied to a spiritual ideal including image
http://31st-and-chi.blogspot.com/2010/07/bunch-of-pictures-of-unicorns.html	0.9	Large array of unicorn pictures. One is identical to result 5.
http://disgrasian.com/2010/09/unicorns-really-do-exist-and-theyre-asian/	0.9	Picture of unicorn and asian 2 horn unicorn.
http://www.youtube.com/watch?v=QSim0Ssyus	0.4	Comedy cartoon video about unicorns.
http://www.youtube.com/watch?v=v25MaXwopNI	0.8	Music for a cartoon character.

3) $P(q, d)$: With respect to the relevance of an individual result compared to the information need, there are a number of different methods that could be used. For example, keyword analysis in text documents and image recognition in images, it is clear that a separate method would be needed for each media type that is returned.

With this in mind, for the example below, the individual relevance of a given result will be manually determined and a brief explanation given. The focus will be more on the relevance of the result to the information need, with some weighting given if there is a creative link.

C. Experimental Data

To show this metric in practice we will need to get real world data about a topic. We have used Google to search for results on the following creative need - unicorns from Greek mythology.

The search term input into Google on Thursday 12th May was 'unicorn'. Below is a table of a url to each result, their assigned P rating and a brief description of the reason why. We have taken the top 10 results including the first 4 images and videos.

Due to the repeated result in result 4 and 6, $D(r) = 0.1$ and $T(r) = 1 - \frac{2 \cdot 0.94}{10} = 0.812$. If we then feed these results into the SCM metric we get.

$$SCM = 0.812 \cdot \frac{1}{10} \cdot 6.15 \cdot (1 - 0.1) = 0.449$$

A 0.449 result for us represents a set of results that contain some creative merit, but which also could be improved. This

result could be enhanced, based on this metric, if four of the results were to have been replaced with more relevant results.

We still need to understand whether the low result is due to the fact that the results are not inspiring creativity as we presume, or that the definition that we have provided is not complete and that we need to extend it further. It is planned, that we take this research further to answer the question using in-depth empirical analysis.

V. CONCLUSION AND FUTURE WORK

Over the course of the report, we have attempted to define what we mean by creativity with respect to search engine results using the concepts from computational creativity. The definition is focused more on the structure and relationship between the results returned than the content of the results themselves. This will allow us to define this separately after carrying out further experiments.

This is evident from the metric that we generated to show how the definition could be used. We believe that the low result shows that the return set does not have a high creative merit; however, more testing will be needed to check whether this is the case, or whether the definition needs to be redefined.

We believe, that whilst this paper has little empirical backing, it has highlighted a short fall in the information retrieval domain, namely that of creative search. Even from the simple test that was conducted, it is apparent, that even when we reach a metric for measuring the creative quality of results, a new form of search engine will be required to achieve top quality results consistently.

The next stage of the research will focus on applying what we have learnt and combine quantitative and qualitative analysis to try and develop a new metric with a strong empirical backing. This means that our definition of creativity will likely need to be adapted over time; however, this could allow us to develop a metric that evolves over time to adapt to what the users consider to be creative search.

REFERENCES

- [1] D. Fallows, "The Internet and daily life," *Pew Internet and American Life Project*, 2004.
- [2] A. Jarry, *Selected works of Alfred Jarry*, 1st ed., R. Shattuck and S. W. Taylor, Eds. Grove Press, 1965.
- [3] A. Newell, J. G. Shaw, and H. A. Simon, *The process of creative thinking*. New York: Atherton, 1963, pp. 63 – 119.
- [4] M. Boden, *The Creative Mind: Myths and Mechanisms*, 2nd ed. Weidenfeld & Nicolson, 1992.
- [5] A. Brannigan, *The Social Basis of Scientific Discoveries*. Cambridge University Press, 1981.
- [6] G. A. Wiggins, "A preliminary framework for description, analysis and comparison of creative systems," *Knowledge-Based Systems*, vol. 19, pp. 449–458, 2006.
- [7] M. A. Boden, *Computer Models of Creativity*. Cambridge University Press, 1999, pp. 351–372.
- [8] A. Koestler, *The Act of Creation*. London: Hutchinson & Co, 1964.
- [9] G. Fauconnier and M. Turner, "Conceptual integration networks," *Cognitive Science*, vol. 22, no. 2, pp. 133–187, 1998.
- [10] Epicurus, *The Epicurus Reader*, B. Inwood and L. P. Gerson, Eds. Hackett, 1994.
- [11] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a very large web search engine query log," *SIGIR Forum*, vol. 33, no. 1, pp. 6–12, 1999.

Bibliography

- 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 (cit. on p. 78).
- Baeza-Yates, Ricardo and Berthier Ribeiro-Neto (2011). *Modern Information Retrieval: The Concepts and Technology Behind Search*. Addison Wesley (cit. on pp. 70, 73, 74, 76–79, 98, 99).
- Baidu (2012). *Baidu About* (cit. on p. 70).
- Bao, Shenghua et al. (2007). 'Optimizing Web Search Using Social Annotations'. In: *Distribution*, pp. 501–510 (cit. on p. 78).
- 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 (cit. on p. 78).
- Baudrillard, Jean (2007). *Pataphysics* (cit. on pp. 29, 116).
- Beghetto, Ronald A. and James C. Kaufman (2007). 'Toward a broader conception of creativity: A case for 'mini-c' creativity.' In: *Psychology of Aesthetics, Creativity, and the Arts* 1.2, pp. 73–79 (cit. on p. 44).
- Bharat, Krishna and George Mihaila (2000). 'Hilltop: A Search Engine based on Expert Documents'. In: *Proc of the 9th International WWW*. Vol. 11 (cit. on p. 77).
- Bird, Steven, Ewan Klein and Edward Loper (2009). *Natural Language Processing with Python*. Sebasopol, CA: O'Reilly Media (cit. on p. 81).
- Boden, Margaret (2003). *The Creative Mind: Myths and Mechanisms*. London: Routledge (cit. on pp. 42, 45–47, 53, 102, 113, 116–118, 158).
- Boek, Christian (2002). *'Pataphysics: The Poetics of an Imaginary Science*. Evanston, Illinois: Northwestern University Press (cit. on pp. 10, 22, 29, 30, 32, 33, 133).
- Borges, Jorge Luis (1964). *Labyrinths - Selected Stories and Other Writings*. New York: New Directions (cit. on p. 35).
- (1999). *Collected fictions*. Trans. by Andrew Hurley. Penguin (cit. on p. 35).

- Borges, Jorge Luis (2000). ‘John Wilkins’ Analytical Language’. In: *Selected Non-Fictions*. Ed. by Eliot Weinberger. London: Penguin Books, pp. 229–232 (cit. on pp. 5, 6, 35).
- (2010). *La biblioteca de Babel*. Reclam (cit. on p. 35).
- Borges, Jorge Luis and L.S. Dembo (2010). ‘Interview with Borges’. In: *Contemporary Literature* 11.3, pp. 315–323 (cit. on p. 35).
- Borges, Jorge Luis and Margarita Guerrero (1957). *Book of Imaginary Beings*. Trans. by Andrew Hurley. Viking (cit. on p. 35).
- 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 (cit. on pp. 70, 76).
- (1998b). ‘The PageRank Citation Ranking: Bringing Order to the Web’. In: *World Wide Web Internet And Web Information Systems*, pp. 1–17 (cit. on pp. 70, 76, 124).
- Brotchie, Alastair (2011). *A supplement*. UK: Atlas Press (cit. on pp. 24–27, 31).
- Brotchie, Alastair and Stanley Chapman, eds. (2007). *Necrologies*. London: Atlas Press (cit. on p. 27).
- Brotchie, Alastair, Stanley Chapman et al., eds. (2003). *'Pataphysics: Definitions and Citations*. London: Atlas Press (cit. on pp. 21, 22).
- Brotchie, Alistair, ed. (1995). *A True History of the College of 'Pataphysics - 1*. Trans. by Paul Edwards. London: Atlas Press (cit. on p. 30).
- Burdick, Anne et al. (2012). *Digital Humanities*. Cambridge, Massachusetts: MIT Press (cit. on pp. 61–64, 67).
- Candy, Linda (2012). ‘Evaluating Creativity’. In: *Creativity and Rationale: Enhancing Human Experience by Design*. Ed. by J.M. Carroll. Springer (cit. on pp. 38, 103–106, 108, 109, 154).
- Colton, Simon (2008a). ‘Computational Creativity’. In: *AISB Quarterly*, pp. 6–7 (cit. on pp. 54, 55, 101).
 - (2008b). ‘Creativity versus the perception of creativity in computational systems’. In: *In Proceedings of the AAAI Spring Symp. on Creative Intelligent Systems* (cit. on pp. 53, 101, 102).
- Colton, Simon, Alison Pease and Graeme Ritchie (2001). *The Effect of Input Knowledge on Creativity* (cit. on p. 101).
- Colton, Simon and Geraint A Wiggins (2012). ‘Computational Creativity: The Final Frontier?’ In: *Proceedings of the 20th European Conference on Artificial Intelligence*. Montpellier, France: IOS Press, pp. 21–26 (cit. on pp. 55, 56).
- Corbyn, Zoe (2005). *An introduction to 'Pataphysics* (cit. on p. 29).
- Cruickshank, Douglas (nd). *Why Anti-Matter Matters* (cit. on p. 29).
- Cutshall, James Anthony (1988). ‘The Figure of the Writer - Alfred Jarry’. Thesis. University of Reading, p. 258 (cit. on p. 27).

- Damerau, Fred J (1964). 'A Technique for Computer Detection and Correction of Spelling Errors'. In: *Communications of the ACM* 7.3, pp. 171–176 (cit. on p. 134).
- Daumal, Rene (2012). *Pataphysical Essays*. Trans. by Thomas Vosteen. Cambridge, Massachusetts: Wakefield Press (cit. on p. 30).
- De Bra, Paul, Geert-jan Houben et al. (1994). 'Information Retrieval in Distributed Hypertexts'. In: *Techniques* (cit. on p. 77).
- 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 (cit. on p. 77).
- (1994b). 'Searching for Arbitrary Information in the WWW: the Fish Search for Mosaic'. In: *Mosaic A journal For The Interdisciplinary Study Of Literature* (cit. on p. 77).
- Dean, Jeffrey, Luiz Andre Barroso and Urs Hoelzle (2003). 'Web Search for a Planet: The Google Cluster Architecture'. In: *Ieee Micro*, pp. 22–28 (cit. on p. 79).
- Deerwester, Scott et al. (1990). 'Indexing by Latent Semantic Analysis'. In: *Journal of the American Society for Information Science* 41.6, pp. 391–407 (cit. on p. 75).
- 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 (cit. on p. 78).
- Drucker, Johanna (2009). *SpecLab: Digital Aesthetics and Projects in Speculative Computing*. University of Chicago Press (cit. on pp. 10, 59, 60, 120, 123).
- Drucker, Johanna and B Nowviskie (2007). 'Speculative Computing: Aesthetic Provocations in Humanities Computing'. In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, John Unsworth and Ray Siemens. Oxford: Blackwell Publishing. Chap. 29 (cit. on pp. 60, 61, 119, 120).
- 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 (cit. on p. 78).
- Dubbelboer, Marieke (2009). "UBUSING' CULTURE". Thesis. Rijksuniversiteit Groningen, p. 233 (cit. on p. 28).
- Eden, Amnon H. (2007). 'Three Paradigms of Computer Science'. In: *Minds and Machines* 17.2, pp. 135–167 (cit. on pp. 55, 57).
- Elton, Matthew (1995). 'Artificial Creativity: Enculturing Computers'. In: *Leonardo* 28.3, pp. 207–213 (cit. on p. 52).
- Foucault, Michel (1966). 'The Order of Things - Preface'. In: *The Order of Things*. France: Editions Gallimard. Chap. Preface, pp. xv–xxiv (cit. on p. 35).

- Garcia-Molina, Hector, Jan Pedersen and Zoltan Gyongyi (2004). ‘Combating Web Spam with TrustRank’. In: *In VLDB*. Morgan Kaufmann, pp. 576–587 (cit. on p. 78).
- Gelernter, David (1994). *The Muse in the Machine*. London: Fourth Estate Limited (cit. on p. 51).
- 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 (cit. on p. 78).
- Google (2012). *Google Ranking* (cit. on pp. 10, 70).
- 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 (cit. on p. 78).
- Heilman, Kenneth M, Stephen E Nadeau and David O Beversdorf (2003). ‘Creative innovation: possible brain mechanisms.’ In: *Neurocase* 9.5, pp. 369–79 (cit. on p. 118).
- Hendler, Jim and Andrew Hugill (2011). ‘The Syzygy Surfer : Creative Technology for the World Wide Web’. In: *ACM WebSci 11* (cit. on p. 6).
- (2013). ‘The syzygy surfer: (Ab)using the semantic web to inspire creativity’. In: *International journal of Creative Computing* 1.1, pp. 20–34 (cit. on pp. 6, 10, 115).
- 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 (cit. on p. 78).
- 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 (cit. on p. 78).
- Hugill, Andrew (2012a). *Lineaments of 'Pataphysics*. Self (cit. on pp. 23, 31–33).
- (2012b). *'Pataphysics: A Useless Guide*. Cambridge, Massachusetts: MIT Press (cit. on pp. 10, 20, 22).
- (2013). ‘Introduction: transdisciplinary learning for digital creative practice’. In: *Digital Creativity* 24.3, pp. 165–167 (cit. on pp. 55, 114).
- Hugill, Andrew and Hongji Yang (2013). ‘The creative turn: new challenges for computing’. In: *International journal of Creative Computing* 1.1, pp. 4–19 (cit. on pp. 57–59).
- Hugill, Andrew, Hongji Yang et al. (2013). ‘The pataphysics of creativity: developing a tool for creative search’. In: *Digital Creativity* 24.3, pp. 237–251 (cit. on p. 10).
- Indurkhya, Bipin (1997). ‘Computers and creativity’. Unpublished manuscript. Based on the keynote speech ‘On Modeling Mechanisms of Creativity’ delivered at Mind II: Computational Models of Creative Cognition (cit. on pp. 49, 118).

- Jarry, Alfred (1996). *Exploits and Opinions of Dr Faustroll, Pataphysician*. Cambridge, MA: Exact Change (cit. on pp. 6, 21, 29, 32, 33, 128, 134).
- Jeh, Glen and Jennifer Widom (2002). ‘SimRank: A Measure of Structural Context Similarity’. In: *In KDD*, pp. 538–543 (cit. on p. 78).
- Jordanous, Anna Katerina (2011). ‘Evaluating Evaluation : Assessing Progress in Computational Creativity Research’. In: *Proceedings of the Second International Conference on Computational Creativity* (cit. on pp. 100, 157).
- 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 (cit. on pp. 40, 53, 102, 103).
- Jorn, Asger (1961). ‘Pataphysics - A Religion In The Making’. In: *Internationale Situationniste* 6 (cit. on p. 29).
- Jurafsky, Daniel and James H Martin (2009). *Speech and Language Processing*. London: Pearson Education (cit. on p. 81).
- Kamps, Jaap, Rianne Kaptein and Marijn Koolen (2010). *Using Anchor Text , Spam Filtering and Wikipedia for Web Search and Entity Ranking*. Tech. rep. ? (Cit. on p. 78).
- Kaufman, James C. and Ronald A. Beghetto (2009). ‘Beyond big and little: The four c model of creativity’. In: *Review of General Psychology* 13.1, pp. 1–12 (cit. on pp. 43, 158).
- Kleinberg, Jon M (1999). ‘Authoritative sources in a hyperlinked environment’. In: *journal of the ACM* 46.5, pp. 604–632 (cit. on pp. 77, 124).
- Kleinberg, Jon M et al. (1999). ‘The Web as a graph : measurements, models and methods’. In: *Computer* (cit. on p. 77).
- Koestler, Arthur (1964). *The Act of Creation*. London: Hutchinson and Co (cit. on pp. 41, 42, 118).
- Levenshtein, Vladimir I (1966). ‘Binary codes capable of correcting deletions, insertions, and reversals’. In: *Soviet Physics Doklady* 10.8, pp. 707–710 (cit. on p. 134).
- Luo, Fang-fang, Guo-long Chen and Wen-zhong Guo (2005). ‘An Improved ‘Fish-search’ Algorithm for Information Retrieval’. In: *2005 International Conference on Natural Language Processing and Knowledge Engineering*, pp. 523–528 (cit. on p. 78).
- Macdonald, Craig (2009). ‘The Voting Model for People Search’. In: *Philosophy* (cit. on p. 75).
- Manning, Christopher, Prabhakar Raghavan and Hinrich Schuetze (2009). *Introduction to Information Retrieval*. Cambridge UP (cit. on pp. 74, 81).
- Marchionini, Gary (2006). ‘From finding to understanding’. In: *Communications of the ACM* 49.4, pp. 41–46 (cit. on pp. 112, 113).

- Marchionini, Gary and Ben Shneiderman (1988). ‘Finding facts vs. browsing knowledge in hypertext systems’. In: *Computer* 21.1, pp. 70–80 (cit. on pp. 70, 112).
- 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 (cit. on p. 88).
- 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 pp. 37, 156).
- McBride, Neil (2013). *Robot Ethics: The Boundaries of Machine Ethics*. Leicester (cit. on p. 68).
- Microsoft (2012). *Bing Fact Sheet* (cit. on p. 70).
- Miller, George A. (1995). ‘WordNet: a lexical database for English’. In: *Communications of the ACM* 38.11, pp. 39–41 (cit. on p. 137).
- Minsky, Marvin (1980). ‘K-Lines : A Theory of Memory’. In: *Cognitive Science* 33.4, pp. 117–133 (cit. on pp. 50, 52).
- (1988). *The Society of Mind*. Simon and Schuster, p. 336 (cit. on pp. 50, 52).
- Miyamoto, Sadaaki (1988). *Information Retrieval based on Fuzzy Associations* (cit. on p. 75).
- (2010). *Fuzzy Sets in Information Retrieval and Cluster Analysis (Theory and Decision Library D)*. Springer, p. 276 (cit. on p. 75).
- 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 (cit. on p. 75).
- Motte, Warren (2007). *Oulipo, A primer of potential literature*. London: Dalkey Archive Press (cit. on pp. 116, 117).
- Newell, A, J. G. Shaw and H. A. Simon (1963). *The Process Of Creative Thinking*. New York: Atherton (cit. on p. 118).
- Nick, Z.Z. and P. Themis (2001). ‘Web Search Using a Genetic Algorithm’. In: *IEEE Internet Computing* 5.2, pp. 18–26 (cit. on p. 78).
- Nicolescu, Basarab (2010). ‘Methodology of Transdisciplinarity - Levels of Reality, Logic of the Included’. In: *Transdisciplinary journal of Engineering and Science* 1.1, pp. 19–38 (cit. on p. 17).
- Partridge, Derek and Jon Rowe (1994). *Computers and Creativity*. Oxford: Intellect (cit. on pp. 42, 50).
- 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 (cit. on p. 102).

- Pease, Alison, Daniel Winterstein and Simon Colton (2001). 'Evaluating Machine Creativity'. In: *Proceedings of ICCBR Workshop on Approaches to Creativity*, pp. 129–137 (cit. on pp. 47, 53, 101, 102).
- 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 (cit. on pp. 102, 154).
- Poincare, Henri (2001). *The Value of Science*. Ed. by Stephen Jay Gould. New York (cit. on pp. 42, 58).
- Polya, George (1957). *How To Solve It*. 2nd. Princeton, New Jersey: Princeton University Press (cit. on pp. 43, 58, 112).
- Raczinski, Fania, Hongji Yang and Andrew Hugill (2013). 'Creative Search Using Pataphysics'. In: *Proceedings of the 9th International Conference on Creativity and Cognition*. Sydney, Australia: ACM New York, NY, USA, pp. 274–280 (cit. on p. 10).
- Ramesh, V., Robert L. Glass and Iris Vessey (2004). 'Research in computer science: an empirical study'. In: *journalttitle of Systems and Software* 70.1-2, pp. 165–176 (cit. on p. 15).
- Rhodes, Mel (1961). 'An analysis of creativity'. In: *The Phi Delta Kappan* 42.7, pp. 305–310 (cit. on p. 40).
- 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 (cit. on pp. 53, 102).
- (2007). 'Some Empirical Criteria for Attributing Creativity to a Computer Program'. In: *Minds and Machines* 17.1, pp. 67–99 (cit. on pp. 53, 101, 102).
- Sawle, James, Fania Raczinski and Hongji Yang (2011). 'A Framework for Creativity in Search Results'. In: *The Third International Conference on Creative Content Technologies*. Rome, pp. 54–57 (cit. on pp. 10, 98).
- Schuetze, Hinrich (1998). 'Automatic Word Sense Discrimination'. In: *Computational Linguistics* (cit. on p. 75).
- Schuetze, Hinrich and Jan Pedersen (1995). *Information Retrieval Based on Word Senses* (cit. on p. 75).
- Shattuck, Roger (1959). *The Banquet Years*. London: Faber (cit. on p. 27).
- Shu, Bo and Subhash Kak (1999). 'A neural network-based intelligent meta-search engine'. In: *Information Sciences* 120 (cit. on p. 78).
- Singh, Push (2005). 'EM-ONE: An Architecture for Reflective Commonsense Thinking'. PhD thesis. Massachusetts Institute of Technology (cit. on p. 6).
- Srinivasan, P (2001). 'Vocabulary mining for information retrieval: rough sets and fuzzy sets'. In: *Information Processing and Management* 37.1, pp. 15–38 (cit. on p. 75).

- Stahl, Bernd Carsten, Marina Jirotnka and Grace Eden (2013). 'Responsible Research and Innovation in Information and Communication Technology: Identifying and Engaging with the Ethical Implications of ICTs'. In: *Responsible Innovation*. Ed. by Richard Owen. John Wiley and Sons. Chap. 11, pp. 199–218 (cit. on p. 158).
- Sternberg, Robert J (1999). *Handbook of creativity*. Cambridge University Press, p. 490 (cit. on pp. 40, 44, 47, 118).
- (2006). 'The Nature of Creativity'. In: *Creativity Research journal* 18.1, pp. 87–98 (cit. on p. 44).
- Sutcliffe, Alistair and Mark Ennis (1998). 'Towards a cognitive theory of information retrieval'. In: *Interacting with Computers* 10, pp. 321–351 (cit. on p. 112).
- Taye, Mohammad Mustafa (2009). 'Ontology Alignment Mechanisms for Improving Web-based Searching'. PhD thesis. De Montfort University (cit. on p. 78).
- Thomas, Sue et al. (2007). 'Transliteracy: Crossing divides'. In: *First Monday* 12.12 (cit. on p. 61).
- 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 (cit. on pp. 53, 102).
- Vian, Boris (2006). '*Pataphysics? What's That?* Trans. by Stanley Chapman. London: Atlas Press (cit. on p. 30).
- Vries, Erica de (1993). 'Browsing vs Searching'. In: *OCTO report* 93/02 (cit. on pp. 112, 113).
- Wallas, Graham (1926). *The Art of Thought*. Jonathan Cape (cit. on pp. 42, 58).
- Walsh, Dave (2001). *Absinthe, Bicycles and Merdre* (cit. on p. 25).
- 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 (cit. on pp. 75, 78).
- Wiggins, Geraint A (2006). 'A preliminary framework for description, analysis and comparison of creative systems'. In: *Knowledge Based Systems* 19.7, pp. 449–458 (cit. on pp. 53, 102).
- Yang, Hongji (2013). 'Editorial'. In: *International journal of Creative Computing* 1.1, pp. 1–3 (cit. on p. 57).

Glossary

bisociation

Two self-consistent but habitually incompatible frames of reference intersecting to give rise to a new creative idea.. [41](#)

holonym

The relationship between a term denoting the whole and a term denoting a part of, or a member of, the whole. That is, 'X' is a holonym of 'Y' if Ys are parts of Xs, or 'X' is a holonym of 'Y' if Ys are members of Xs. For example, 'tree' is a holonym of 'bark', of 'trunk' and of 'limb.' Holonymy is the opposite of meronymy.. [91](#)

hypernym

A hyponym shares a type-of relationship with its hypernym. For example, pigeon, crow, eagle and seagull are all hyponyms of bird (their hypernym); which, in turn, is a hyponym of animal.. [91](#)