



### a recommender system



Edwin Henneken Smithsonian Astrophysical Observatory



Introduction
Machinery
Example
Closing Remarks

### The Literature Universe





Introduction
 Machinery
 Example
 Closing Remarks

### The Literature Universe





Hachinery
 Achinery
 Example
 Closing Remarks

### The Literature Universe Discovery Tools







Haroduction
 Machinery
 Example
 Closing Remarks

## Recommender Systems...

- are a technological proxy for a social process
- are a way of suggesting like or similar items and ideas to a users specific way of thinking
- try to automate aspects of a completely different information discovery model where people try to find other people with similar tastes and then ask them to suggest new things

## Goal(s):

- + help users find information that was previously unknown to them
- for those commercially inclined: improve sales



Hachinery
 Machinery
 Example
 Closing Remarks

## **Key Elements**

### • **\*** "proximity" concept:

- I. Item  $\mathcal{X}$  is close to item  $\mathcal{Y}$
- **2.** Person  $\mathcal{P}$  is close to item  $\mathcal{X}$

• \* "ansatz": person with similar preferences in the past are likely to have similar preferences in the future

#### ▼ <u>Approach</u>

- Build a "space" in which documents and persons can be placed
- Determine document clustering ("thematic map")



Hachinery
 Machinery
 Example
 Closing Remarks

## Key Ingredients

- assign labels to documents ("topic vector")
- assign labels to persons ("interest vector")

How? Persons read documents

Keywords



• Antroduction • Antr

	More theoretical		
Extra-galactic astronomy		Galactic astronomy	
	Nore experimental		







Introduction
Machinery
Example
Closing Remarks

## Finding Recommendations





Introduction
Machinery
Example
Closing Remarks

## Finding Recommendations





Antroduction
 Machinery
 Example
 Closing Remarks

## Finding Recommendations

Fact: document X belong in cluster C

Definition: group G is the group of 40 documents closest to X (within C)

- I. paper read the most before reading a paper from G
- 2. paper read the most after reading a paper from G

Find all the persons who read a paper from G:

- 3. the most read paper by these persons in the last few months
- 4. the most recent paper in the top 100 of the most also-read list

From citations:

5. paper which cites the most papers in group G



Antroduction
 Machinery
 Example
 Closing Remarks

## Example: Physics Literature

data used to build topic space and clusters:

journals: Physical Review (A,B,C,D,E,Letters), Physics Letters (A,B),

Reviews of Modern Physics, Nuclear Physics (A,B),

Journal of Chemical Physics, Applied Physics Letters

time range: 1995 - 2009

resulting data set: 144,110 papers

available keywords: 6,349

usage data: period July - December 2009



• Introduction • Machinery • Example • Closing Remarks





• Antroduction • Antr

ads	The Sm	ithsonian/N	IASA Astrophy	sics Da	ita System
Home	Help	Sitemap		2010arXiv100	2.3050R Search
Free Fulltext /	Article	Citations	Find Similar Articles	Full	I record info
Why LHC?					Related Articles
eprint arXiv:10	02.3050				2009PhRvD79k4026K Kisslinger, Mixed heavy quark hybrid mesons, decay puzzles, and RHIC
I discuss LHC physics in the historical perspective of the progress in particle physics. After a recap of the Standard Model of particle physics. I discuss the production					1996PhRvD54.3216B Braaten et al, Helicity decomposition for inclusive J/ψ production
high energy colliders leading up to LHC and their role in the discovery of these SM particles. Then I discuss the two main physics issues of LHC, i.e. Higgs mechanism and Supersymmetry. I briefly touch upon Higgs and SUSY searches at LHC along with their cosmological implications.				2001PhRvA63c21010 Olsen, Electroproduction of relativistic positronium	
					2009PhHvL.103/2001A Abazov et al, Observation of Single Top-Quark Production
Keywords: Higi	h Energy Physics - Ph	enomenology			2009PhRvL.103i2002A Aaltonen et al, Observation of Electroweak Single Top- Quark Production
					2009arXiv0908.1964W Weinberg, Effective Field Theory, Past and Future
					2010PhRvD81c4012S Shu et al, Explorations of the top quark forward- backward asymmetry at the Tevatron
					2009arXiv0912.3259M Morrissey et al, New Physics at the LHC
					2008PhLB6671P Particle Data Group et al, Review of Particle Physics
					2009PhRvD. 79a5014A Arkani-Hamed et al, A theory of dark matter
👽 💖 The	ADS is Operated by th	e Smithsonian Astropl	nysical Observatory under NA	ASA Grant NN	X09AB39G



Haroduction
Machinery
Example
Closing Remarks

### So, is this something new?

JANUARY, 1934

#### THE PSYCHOLOGICAL REVIEW

Vol. 41, No. 1

#### THE VECTORS OF MIND<sup>1</sup> BY L. L. THURSTONE The University of Chicago

Under the title of this address, 'The Vectors of Mind,' I shall discuss one of the oldest of psychological problems with the aid of some new analytical method. I am referring to the old problem of classifying the temperaments and personality types and the more recent problem of isolating the different

mental abilities. Until very recently the only attempt to solve this problem in a quantitative way seems to have been the work of Professor Spearman and his students. Spearman has formulated methods for dealing with the simplest case, in which all of the variables that enter into a particular study can be regarded as having only one factor in common The factor theory that I shall describe starts without this limitation, in that I shall make no restriction as to the number of factors that are involved in any particular problem. The resulting factor theorems are quite different in form and in their underlying assumptions, but it is of interest to discover that they are consistent with Spearman's factor theory, which turns out to be a special case of the present general factor theory.

In this paper I shall first review the single-factor theory of Spearman. Then I shall describe a general factor theory. Those who have only a casual interest in the theoretical aspects of this problem will be more interested perhaps in the applications of the new factor theory to a number of psychohadress of the preident before the American Psychological Association, Chicago meeting, September, 1933.

Louis L. Thurstone (1934), The Psychological Review, 41, 1

#### Classification Space: A Multivariate Procedure For Automatic? Document Indexing And Retrieval

Author: Peter G. Ossorio

Affiliation: \* University of Colorado. DOI: 10.1207/s15327906mbo104\_6 Publication Frequency: 6 issues per year Published in: I full walthwarate Behavioral Research, Volume 1, Issue 4 October 1966 pages 479 - 524 Formats available: PDF (English)

Article Requests: Order Reprints : Request Permissions Single Article Purchase: US\$30.00 - buy now = add to cart +

[ show other buying options ] Sign In ⇔ Online Sample 🛱

View Article: 🖾 View Article (PDF)

#### Abstract

A conceptual approach to linguistic data processing problems is sketched and empirical illustrations are presented of the major software components- indexing, storage, and retrieval-of a document processing system which offers, in principle, the advantages of complete automation, unlimited cross- indexing, effective sequential retrieval, sub-documentary indexing reflecting heterogeneity of subject matter within a document, and a procedure for automatically identifying retrieval requests which would be inadequately handled by the system. The indexing schema, designated as a "Classification Space" consists of a Euclidean model for mapping subject matter similarity within a given subject matter domain. A schema of this kind is empirically derived for certain fields of Engineering and Chemistry. A set of five related empirical studies provide convincing evidence that when appropriate experimental procedures are followed a very stable C-Space for a given content domain can be constructed on a surprisingly small data base. Other empirical studies demonstrate specific computational procedures for effective automatic indexing of documents in a C-Space, using a relatively small system vocabulary. One study demonstrates that a C-Space maps subject matter relevance as well as subject matter similarity, and thereby pro- motes effective sequential retrieval ; this result is also shown under conditions of automatic indexing. Negative results are found in an attempt to use the structural linguistic distinction of subject and object as a means of improving techniques for automatic indexing.

Peter G. Ossorio (1966) Multivariate Behavioral Research, 1, 479

#### Chapter 3

Advice from the Oracle: Really Intelligent Information Retrieval

#### Michael J. Kurtz

Harvard-Smithsonian Center for Astrophysics 60 Garden Street, Cambridge, MA 02138 (USA) Email: kurtz%cfazwi.decnet@cfa.harvard.edu

#### 3.1 Introduction

What is "intelligent" information retrieval? Essentially this is asking what is intelligence. In this article I will attempt to show some of the aspects of human intelligence, as related to information retrieval. I will do this by the device of a semi-imaginary Oracle. Every Observatory has an oracle, someone who is a distinguished scientist, has great administrative responsibilities, acts as mentor to a number of less senior people, and as trusted advisor to even the most accomplished scientists, and knows essentially everyone in the field.

In an appendix I will present a brief summary of the Statistical Factor Space method for text indexing and retrieval, and indicate how it will be used in the Astrophysics Data System Abstract Service.

#### 3.2 Advice from the Oracle

#### 1. The Oracle sometimes answers without being asked.

Our Oracle walks the hallways, looks into offices, and offers unsolicited advice, often. For example a conversation about the proper definition of galaxy photometry for a particular project was occurring in my office; the Oracle appeared out of nowhere and said 'the amplifier on the new chip will not do any better than ten electrons readout

A. Heek and F. Murtagh (eds.), Intelligent Information Retrieval: The Case of Astronomy and Related Space Sciences, 21–28. © 1993 Kluwer Academic Publishers. Printed in the Netherlands.

Michael J. Kurtz (1993),

in: A. Heck and F. Murtagh (eds.), "Intelligent Information Retrieval: The Case of Astronomy and Related Space Sciences", Astronomy and Space Science Library, 182, 21



Introduction
Machinery
Example
Closing Remarks

### **Future Developments**

- Stability analysis
- Testing, Testing and more Testing



Introduction
Machinery
Example
Closing Remarks

The ADS is Operated by the Smithsonian Astrophysical Observatory under NASA Grant NNX09AB39G

### Thank you! Questions?



"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO,"