

More Like This: Approaches to Recommending Similar Items Using Subject Headings

Kevin Beswick
NCSU Libraries Fellow
code4lib 2014 conference, Raleigh, NC

Agenda

- What?
- Why?
- How?
- Evaluation. Are these approaches any good?
- Where are we going from here?

Recommendation Systems

- A system that presents a set of related items that would interest a particular user
- Collaborative filtering - look at user behavior
 - eg. full record page view data, circulation data, etc
- Content-based filtering - look at properties of content itself
 - eg. call numbers, subject headings, etc.

Motivations

- Many popular web services offer this functionality
 - eg. Facebook, Netflix, Amazon, etc.
 - Users coming to expect it
 - Encourages use & makes it easier to use our service
- Also...



bookBot

- Most of Hunt's collection is stored in an ASRS
 - No physical browsing
 - Need to explore methods for serendipitous discovery



A Brief History of Browse @ NC State

- Virtual Browse team with members from many library departments
- Previous Projects:
 - “Browse Shelf” feature in library catalog
 - Virtual Browse kiosk @ Hunt Library

Browse Shelf

Browse Shelf [Start Over](#)

Browse items from

[Cover View](#) | [List View](#) [Ask Us chat now](#)

List (0)

The image displays a grid of 10 book covers from a digital library. The books are arranged in two rows of five. The top row includes 'Dive Into Python' by Mark Pilgrim, 'Dive Into Python [electronic resource]' by Mark Pilgrim, 'Dive Into Python 3' by Mark Pilgrim, 'Python Cookbook' by Mark Meeker and David Ascher, and another 'Python Cookbook' by Mark Meeker and David Ascher. The bottom row includes 'PYTHON HOW TO PROGRAM' by Mark Summerfield, 'Mobile Python' by Dagner Schiele, 'Pro Python System Administration' by Rytis Sileika, 'Programming in Python 3' by Mark Summerfield, and another 'Programming in Python 3' by Mark Summerfield. The covers feature various designs, including abstract patterns, animal illustrations (a kangaroo and a scorpion), and text-based layouts. Navigation arrows are visible on the left and right sides of the grid.

Advantages of Subject Heading Based Recommendation

- Vs. Call Number Browse
 - Can recommend more than items that are shelved next to each other
 - A lot of our e-books don't have call numbers
- Vs. Collaborative Filtering
 - Hard to collect reliable circulation data for electronic resources

Four Algorithms/ Approaches

Most Subject Headings

North Carolina. Department of Agriculture -- Bibliography -- Catalogs

Agriculture -- North Carolina -- Bibliography -- Catalogs

Administrative law -- North Carolina

First Subject Headings

Highest Weighted

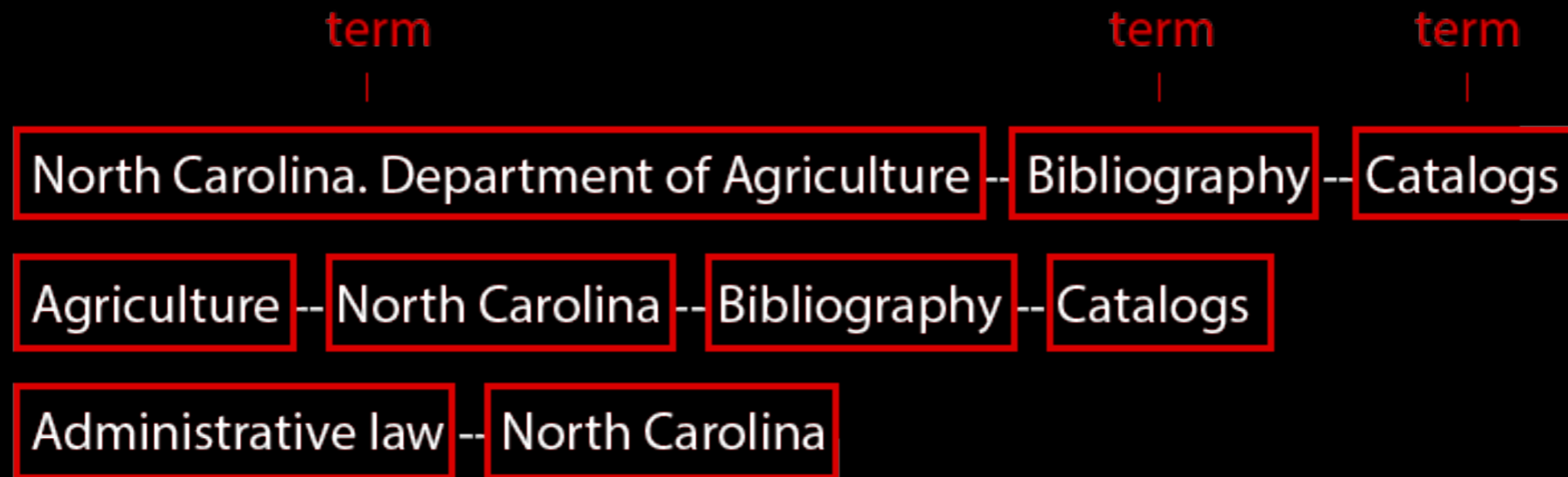
1 North Carolina. Department of Agriculture -- Bibliography -- Catalogs

2 Agriculture -- North Carolina -- Bibliography -- Catalogs

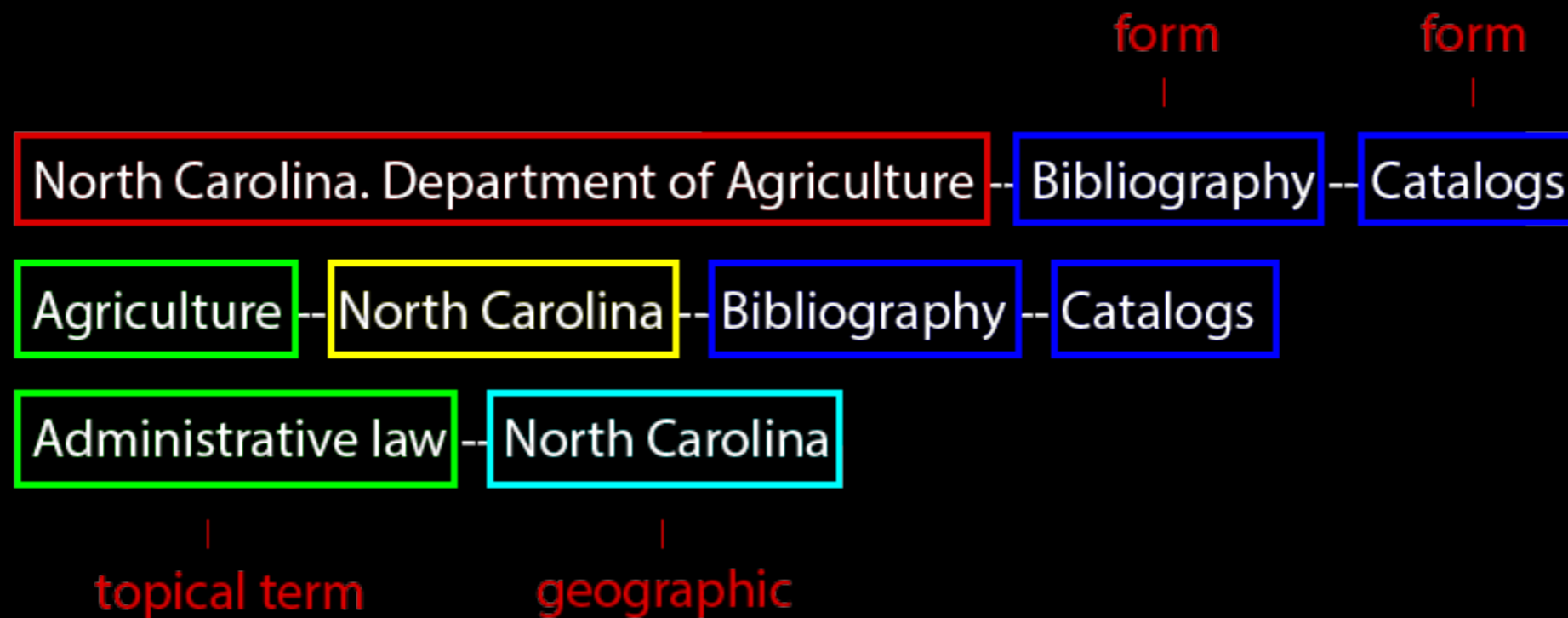
3 Administrative law -- North Carolina

Lowest Weighted

Most Subject Terms



Weighted Subject Terms



- + general subdivision
- + chronological subdivision

Implementation

- Quick & simple implementation
 - Python / Flask - handle requests, provide testing interface
 - Solr / SolrMARC - handle the actual work

Python / Flask App

- Handles requests / responses
 - Accepts a bibliographic ID & algorithm type as input
 - Sends a different query to Solr depending on algorithm
 - Uses SolrPy library
 - Returns a list of recommendations in JSON
 - Also an HTML testing & evaluation interface

Solr / SolrMARC

- Indexed fields with SolrMARC:
 - Entire subject headings
 - Each subject heading term
 - Each topical, general, geographical, chronological, form subdivision
- Lean on Solr to do the heavy lifting in terms of returning the most related items

FindSimilar Results Most Headings Algorithm

Results similar to: "[Perception beyond Gestalt : progress in vision research](#)"

Subject Headings

Gestalt psychology

Perception

[The psychology of perception; a philosophical examination of Gestalt theory and derivative theories of perception](#) **9.460139**

[The world in your head : a gestalt view of the mechanism of conscious experience / Steven Lehar](#) **7.5681114**

[The Legacy of Solomon Asch : essays in cognition and social psychology / edited by Irvin Rock](#) **7.5681114**

[Indirect perception / edited by Irvin Rock ; with a foreword by Stephen E. Palmer](#) **7.5681114**

[Reading and the psychology of perception](#) **7.5681114**

[On perceived motion and figural organization \[electronic resource\] / Max Wertheimer ; edited by Lothar Spillmann ; with essays by Viktor Sarris, Robert Sekuler, and Lothar Spillman ; with contributions by Michael Wertheimer ... \[et al.\]](#) **7.5681114**

How Well Do These
Algorithms Perform?

Preliminary Observation

- Most Headings & Most Terms algorithms looked to be producing decent recommendations a lot of the time
- First Headings algorithm - too few results in a lot of cases
- Weighted Terms algorithm
 - Weighting differs based on subject or user's interests
 - We don't want user input

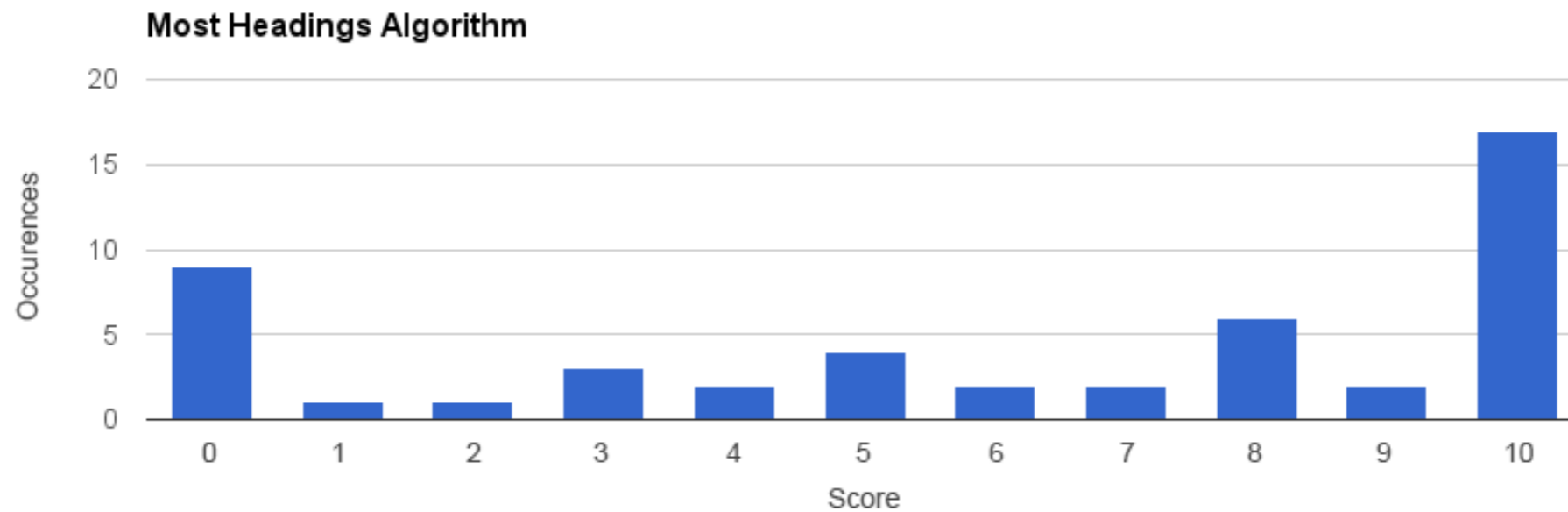
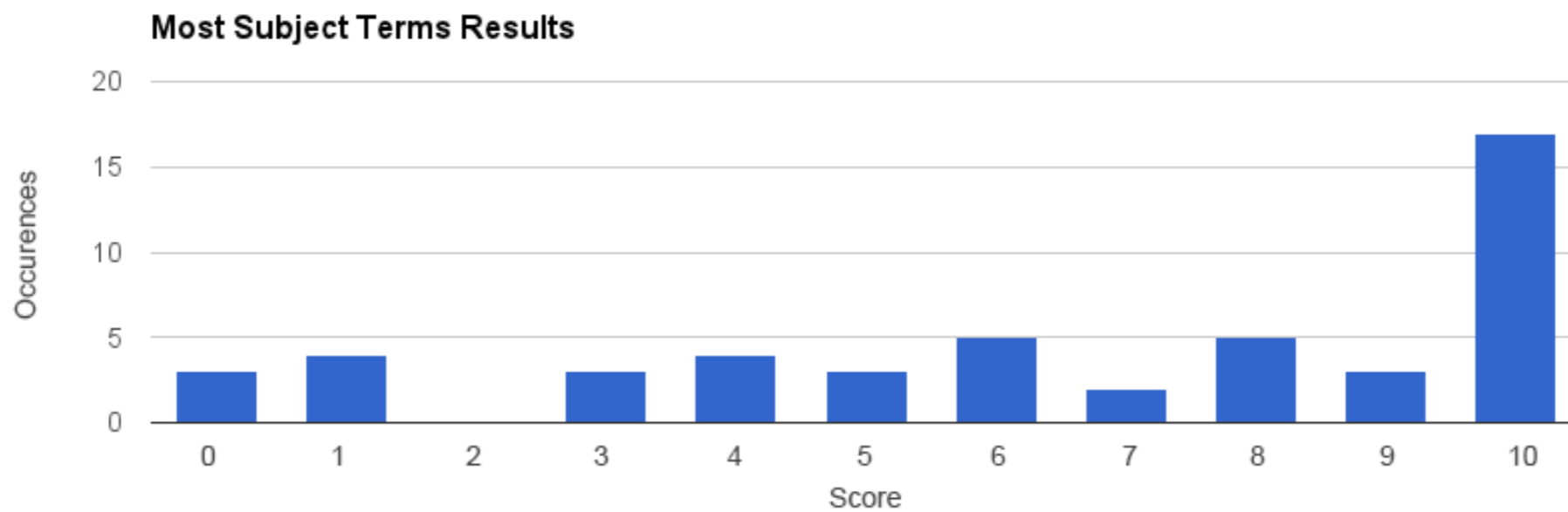
Testing the Algorithms

- Manually test 50 titles on Most Headings & Most Terms algorithms
 - Is either reliable enough & worth implementing?
- 30 hand picked titles
 - representing different subject areas, item formats, lengths & amounts of subject headings
- 20 random titles

Testing the Algorithms

- Blind testing - algorithm unknown
- 10 recommended titles per item
- Rank result set out of 10, 1 point for each relevant work
- Qualitative comments for each result set

Results - Distribution of Scores



Results

- Most headings algorithm performs slightly better for shorter (less subdivisions) & fewer subject headings
- Terms algorithm performs significantly better for longer (more subdivisions) & higher numbers of subject headings
- Found that Gov. Docs & Fiction have interesting thematic recommendations that we can't achieve with shelf browse

Observations

- Duplicate titles
 - Older vs. newer editions
 - Print vs. Electronic
- Format
 - Incorporate a higher weighting on format of recommended items

Observations

- Poorly assigned subject headings responsible for a lot of the poor recommendations
 - General vs. Specific recommendations
 - Automate review/assignment of subject headings to our collection?

Interface Considerations

- Inline “cover-flow” style presentation on full record page
 - Catches eye of user
- Title - “Similar Titles” or “Related Items” etc.
- 5 or so recommendations per title

Library Catalog

Search for words:

Search within results [Start Over](#)

Expand Your Search

- Search for this title at Triangle research libraries
- Search for this title at Libraries worldwide

[Request](#) [Text](#) [Email](#) [Print](#) [RefWorks](#) [List \(0\)](#) [Ask Us chat now](#)



Before dinner : philosophy and ethics of food

by Michiel Korthals ; translator, Frans Kooymans.

Author: **Korthals, Michiel.**

Published: Dordrecht, Netherlands : Springer ; Norwell, MA : Sold and distributed in North, Central, and South America by Springer, c2004.

Description: xiii, 213 p. : ill. ; 25 cm.

Format: Book

Summary There was an error retrieving the data.

[\(more\)](#)



[Browse Shelf](#)

Browse Related Subjects

[\(show more subjects\)](#)

- Food habits -- Philosophy
- Food habits -- Moral and ethical aspects
- Food -- Philosophy
- Food consumption -- Philosophy
- Food consumption -- Moral and ethical aspects

Location **Details** **Contents** **Marc Record**

D.H. Hill Library

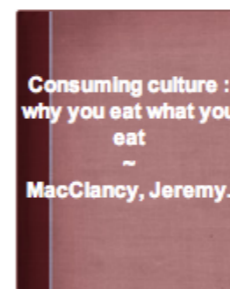
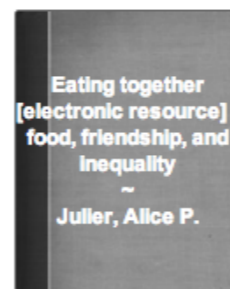
Call Number

Stacks (3rd floor)

GT2850 .K6713 2004

Available

Similar Items



Takeaways

- Overall, the algorithms perform decently for our collection, but could still be improved in a number of ways
- Your mileage may vary - all collections are different
 - Very dependent on quality & coverage of subject headings

Steps Forward

- Use either Most Terms algorithm by itself, or a hybrid of Most Terms & Most Headings
- Still under active development
 - Explore & implement fixes for issues discussed earlier to improve performance

Thank you!

- Kevin Beswick, NCSU Libraries Fellow (IT & Digital Library Initiatives)
- kdbeswic@ncsu.edu
- @kbeswick on Twitter