

Information Retrieval CS 834 : Assignment 5

Miranda Smith
msmit213@odu.edu

December 15, 2017

Abstract

Exercise questions 10.3, 10.5, 10.6, 10.11, 11.5 completed. Spring 2017.

1 Problem 10.3

Compute five iterations of HITS (see Algorithm 3) and PageRank (see Figure 4.11) on the graph in Figure 10.3. Discuss how the PageRank scores compare to the hub and authority scores produced by HITS.

1.1 Solution

I calculated the HITS and page rank score by hand and with excel 103.xlsx got the following results. I did it by hand instead of programming it or using a library so I could better understand how the formulas were working. I did not include iterations 1-3 because they are included in the book and mine are identical.

Iteration 4	Node	Hub	Auth	PageRank
	1	0.14	0.24	0.03
	2	0	0.32	0.05
	3	0.45	0	0.02
	4	0	0.44	0.06
	5	0.21	0	0.02
	6	0.21	0	0.02
	7	0	0	0.02

Iteration 5	Node	Hub	Auth	PageRank
	1	0.15	0.24	0.03
	2	0	0.31	0.05
	3	0.45	0	0.02
	4	0	0.45	0.06
	5	0.2	0	0.02
	6	0.2	0	0.02
	7	0	0	0.02

Page Rank only needed 3 iterations to converge to a number. The numbers are a lot lower and it seems to be harsher than HITS. For example Node 3 ended up with a hub score of 0.45, the highest of all 7 but its page rank is 0.02, one of the lowest. Page rank doesn't take into consideration the importance of being a hub and ranks them without consideration to a page pointing to good pages.

2 Problem 10.5

Find a community-based question answering site on the Web and ask two questions, one that is low-quality and one that is high-quality. Describe the answer quality of each question.

2.1 Solution

I chose stackoverflow because it is the Q & A website I am most familiar with. I chose two questions that I had come across and had difficulty solving in a different class information visualization that have value as a question and only varied the quality with which I asked the question. These questions focus on programming with d3. It is the hope that the similar post time of the question and the similar topics would make the results of the questions as similar as possible in regard to the people who might have answered and remove as much bias as possible.

2.1.1 High quality question

I created this question by taking two graphs in d3 to be displayed on the same page and broke it by naming the variables the same instead of renaming everything. I included all the code that I had written and highlighted the important parts and what I had tried as thoroughly as possible, but left out explicitly mentioning the few extra variables I needed to rename to fix it.

d3 javascript display multiple charts on the same page

I am trying to display two charts, one bar and one line, one the same simple html page. Both charts display properly when displayed in different html files, but when I put them together they break.

Per the advice from this answer on [creating multiple divs](#) I created two div objects to hold my graphs and named them differently.

```
<!--Create svg elements to hold d3 object -->
<div id="Chart1"></div>

<div id="Chart2"></div>
```

Attaching to the first div

```
// append the svg object to the body of the page
// appends a 'group' element to 'svg'
// moves the 'group' element to the top left margin
var svg = d3.select("#Chart2").append("svg")
    .attr("width", width + margin.left + margin.right)
    .attr("height", height + margin.top + margin.bottom)
    .append("g")
    .attr("transform",
        "translate(" + margin.left + "," + margin.top + ")");
```

Attaching to the second div

```
var BarChart = d3.select("#Chart1").append("svg")
    .attr("width", width + margin.left + margin.right)
    .attr("height", height + margin.top + margin.bottom)
    .append("g")
    .attr("transform", "translate(" + margin.left + "," + margin.top + ")");
```

However it only shows the bar chart and the axis for the line chart. The line for the chart is not showing and I do not see any output on the console to indicate errors. [Program Output](#)

The full code is here

```
<html>
<body>

<!--Create svg elements to hold d3 object -->
<div id="Chart1"></div>

<div id="Chart2"></div>

<script>

    // set the dimensions and margins of the graph
    var margin = {top: 50, right: 20, bottom: 80, left: 120},
        width = 960 - margin.left - margin.right,
        height = 900 - margin.top - margin.bottom;

    // parse the date / time
    var parseTime = d3.timeParse("%Y");

    // set the ranges
    var x = d3.scaleTime().range([0, width]);
    var y = d3.scaleLinear().range([height, 0]);

    // define the Line
    var valueline = d3.line()
        .x(function(d) { return x(d.year); })
        .y(function(d) { return y(d.sum); });

    // append the svg object to the body of the page
    // appends a 'group' element to 'svg'
    // moves the 'group' element to the top left margin
    var svg = d3.select("#Chart2").append("svg")
        .attr("width", width + margin.left + margin.right)
        .attr("height", height + margin.top + margin.bottom)
        .append("g")
        .attr("transform",
            "translate(" + margin.left + "," + margin.top + ")");

    // Get the data
    d3.csv("data/YearContributionSum.csv", function(error, data2) {
        if (error) throw error;

        // format the data
        data2.forEach(function(d) {
            d.year = parseTime(d.year);
            d.sum = +d.sum;
        });

        // Scale the range of the data
        x.domain(d3.extent(data2, function(d) { return d.year; }));
        y.domain([0, d3.max(data2, function(d) { return d.sum; })]);

        // Add the valueLine path.
        svg.append("path")
            .attr("class", "line")
            .style("stroke", "steelblue")
            .style("stroke-width", 2)
            .datum(data2)
            .attr("d", valueline(data2));

        // Add the bar chart
        data2.forEach(function(d) {
            var bar = svg.append("rect")
                .attr("x", x(d.year))
                .attr("y", height - y(d.sum))
                .attr("width", x.bandwidth())
                .attr("height", y(d.sum))
                .style("fill", "#87ceeb")
                .style("stroke", "black")
                .style("stroke-width", 1);
        });
    });
</script>
```

```

        // Add the X Axis
        svg.append("g")
            .attr("transform", "translate(0," + height + ")")
            .call(d3.axisBottom(x));

        // text Label for the x axis
        svg.append("text")
            .attr("transform",
                "translate(" + (width/2) + "," +
                    (height + margin.top + 20) + ")")
            .style("text-anchor", "middle")
            .text("Year");

        // Add the Y Axis
        svg.append("g")
            .call(d3.axisLeft(y));

        // text Label for the y axis
        svg.append("text")
            .attr("transform", "rotate(-90)")
            .attr("y", 0 - margin.left)
            .attr("x", 0 - (height / 2))
            .attr("dy", "1em")
            .style("text-anchor", "middle")
            .text("Sum of Contributions");

    });
</script>
<script>

    var margin = {top: 20, right: 30, bottom: 100, left: 80},
        width = 1500 - margin.left - margin.right,
        height = 800 - margin.top - margin.bottom;

    var x = d3.scaleBand()
        .rangeRound([0, width], .1);
    var y = d3.scaleLinear()
        .range([height, 0]);

    var xAxis = d3.axisBottom(x);
    var yAxis = d3.axisLeft(y);

    var BarChart = d3.select("#Chart1").append("svg")
        .attr("width", width + margin.left + margin.right)
        .attr("height", height + margin.top + margin.bottom)
        .append("g")
        .attr("transform", "translate(" + margin.left + "," + margin.top + ")");

    d3.csv("data/PurposeContributionSum.csv", type, function(error, data) {
        if (error) throw error;
        data.sort(function(a, b){ return b.sum - a.sum;});
        x.domain(data.map(function(d) { return d.purpose; }));
        y.domain([0, d3.max(data, function(d) { return d.sum; })]);

        BarChart.append("g")
            .attr("class", "x axis")
            .attr("transform", "translate(0," + height + ")")
            .call(xAxis)
            .selectAll("text")
            .attr("y", 0)
            .attr("x", 9)
            .attr("dy", ".35em")
            .attr("transform", "rotate(90)")
            .style("text-anchor", "start");

        BarChart.append("g")
            .attr("class", "y axis")
            .call(yAxis);

        BarChart.selectAll(".bar")
            .data(data)
            .enter().append("rect")
            .attr("class", "bar")
            .attr("x", function(d) { return x(d.purpose); })
            .attr("y", function(d) { return y(d.sum); })
            .attr("height", function(d) { return height - y(d.sum); })
            .attr("width", x.bandwidth());

    });

    function type(d) {
        d.sum = +d.sum; // coerce to number
        return d;
    }
</script>
</body>
</html>

```

javascript html d3.js

Renamed x and y for one. The line shows up but the x axis is cutting through the middle of the graph

– [Miranda Smith](#) Dec 11 at 22:15

we need more info about the data you are using. I think the issue must be at the time you try to create the line path. Is it possible to have a look to the data you are using? – [torresomar](#) Dec 11 at 22:44

Could you append the path to the g element rather than the svg? You'll need to create a variable to for that, like `p1ot`, then append the line to the plot (adjusted for margins) rather than the svg (not adjusted for margins). – [Ryan Morton](#) Dec 11 at 22:51

1 Answer

The line chart may be having some issues at render time, your code seems to work given a correct input for the line chart, here is a [working JSbin](#).

```
const contributionSum = [ // Had to mock this
  { year: '1990', sum: 1000 },
  { year: '1991', sum: 1090 },
  { year: '1992', sum: 900 },
  { year: '1993', sum: 1300 },
  { year: '1994', sum: 1250 },
  { year: '1995', sum: 1150 }];
// Format data
const formatContributions = contributionSum.map(v => {
  return {
    year: parseTime(v.year),
    sum: +v.sum
  }
});
// Scale the range of the data
x.domain(d3.extent(formatContributions, d => d.year));
y.domain([0, d3.max(formatContributions, d => d.sum)]);
```

I think the issue comes from the input you are providing to the line chart, I also had to create some additional variables to keep things in order at the Bar Chart render:

```
var xBar = d3.scaleBand()
  .rangeRound([0, width], .1);
var yBar = d3.scaleLinear()
  .range([height, 0]);

var xBarAxis = d3.axisBottom(xBar);
var yBarAxis = d3.axisLeft(yBar);
```

answered Dec 11 at 23:28



[torresomar](#)

1,484 2 11 27

Answer Your Question

I had two people respond. The first comment came very close to fixing the issue, they only thing missing was changing the margin along with the axis. The official answer given doesn't change what I was expecting them to but they created a JSbin shown in figure 1 and got the code working for me and created mock data to give it since I didn't include that. They seemed to have put a lot of effort into the answer.

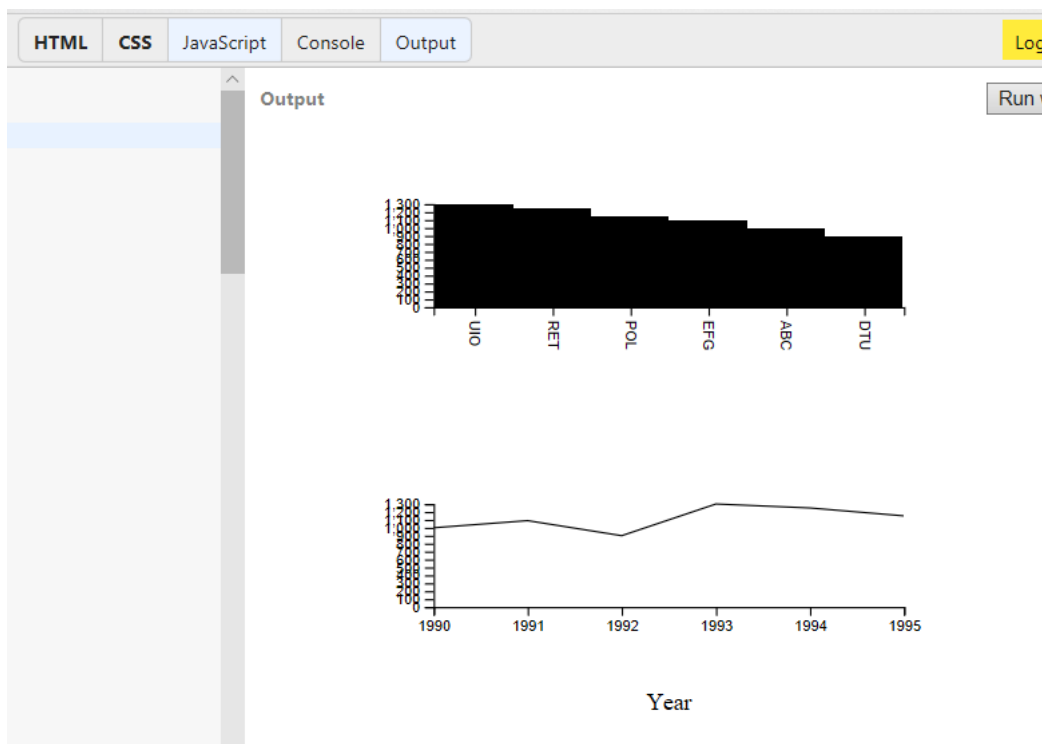


Figure 1: Solution Provided

2.1.2 Low quality question

This question I asked in an ad hoc manner. I provided the minimum amount of effort to get my question across but still had a valid good quality question to be asked. I chose to interpret low quality to mean the effort put into the question and not that its meaning was indecipherable. I provided no code that I had written, only linked to a page showing an example of a grouped bar chart.


Combining color schemes for categorical data with linear ranges

I'm creating a grouped bar chart and trying to create a color scale for it.

I see the example here for [coloring with a scale](#). But let's say the bars in the group could be implicitly contained in smaller groups and should have its own linear color scale.

Let's say I wanted in the example linked the first 3 bars of a group on one linear (blue) color scale and the last 4 bars on a different linear (red) color scale. How would I do this?

javascript d3.js colorbrewer



edited Dec 11 at 23:26

Gerardo Furtado

46.4k 6 28 65

asked Dec 11 at 22:07

Miranda Smith

1 1

1 Answer

This is an **ad hoc** answer, aimed to the example you linked.

One solution is setting the two scales...

```
var scale1 = d3.scaleLinear()
  .range(["lightblue", "darkblue"])
  .domain([0, 2]);

var scale2 = d3.scaleLinear()
  .range(["red", "darkred"])
  .domain([3, 6]);
```

... and using the indices to choose which one to use:

```
.attr("fill", function(d,i) { return i < 3 ? scale1(i) : scale2(i)});
```

Note that I'm using the domains to match the bars:


```
.domain([0, 2])//goes from the first to the third bar
.domain([3, 6])//goes from the forth to the seventh bar
```

Here is the updated bl.ocks:

<https://bl.ocks.org/anonymous/5ba7ad96872b4b5bbc2d212b2ca6e9d0/20e1ed6bb00484f736eab4cad8c2e83bea63d2>

edited Dec 11 at 23:59

answered |

Gerardo

46.4k 6

< >

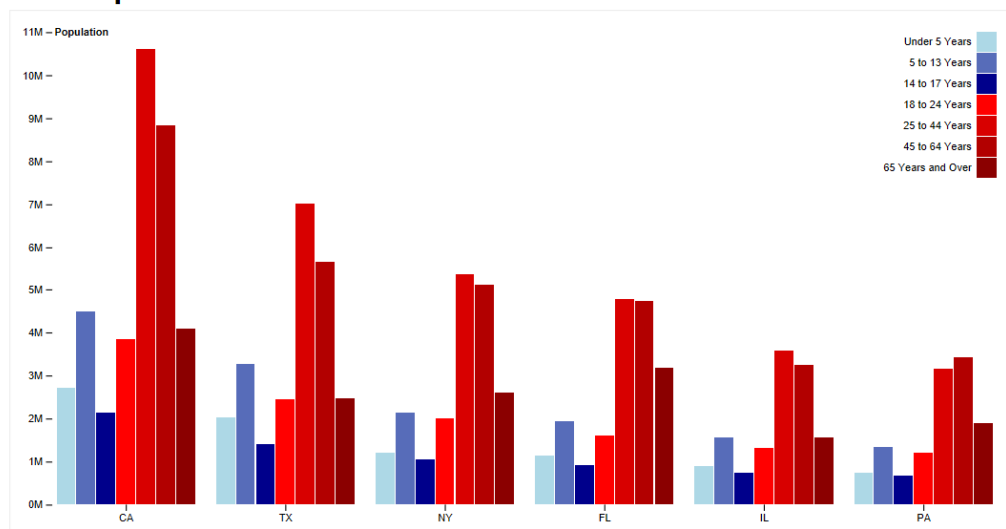
Answer Your Question

I got one answer, and it was exactly what I had wanted and provided a better answer than I had found the first time I had to solve this question. They pointed out exactly what parts of the code to change and created a working example provided in a link to exactly what I was describing shown in Figure 2. The effort seems to have been the same for both questions and provided well working high quality answers.

anonymous's Block 5ba7ad96872b4b5bbc2d212b2ca6e9d0
Updated December 11, 2017

[Popular](#) / [About](#)

Grouped Bar Chart



This grouped bar chart is constructed from a CSV file storing the populations of different states by

[Open](#)

Figure 2: Solution Provided

3 Problem 10.6

Find two examples of document filtering systems on the Web. How do they build a profile for your information need? Is the system static or adaptive?

3.1 Solution

The first example is <https://www.pinterest.com/>. They build a profile for me by using an account and keeping track of what I like and pin to my own boards. The system is adaptive. I have been on pinterest for years and depending on what my most recent browsing history has been, pinterest shows me pins that relate most to recent interests.

The second example is <https://scholar.google.com/>. When I created a profile I created a filter for myself to send emails about articles I would be interested in based off a topic I specify. When I am logged in they also show recommended articles on the search screen. This system is static. It does not update based on interaction activity. If I wanted the content to change I would need to create another filter to replace it or add on to what I'm receiving.

4 Problem 10.11

Suggest how the maximum and minimum resource ranking scores, R_{max} and R_{min} , could be estimated for a given query.

4.1 Solution

To get the R_{max} for a particular resource that has been returned the resources need to be reordered so that the one in question has been returned first and recalculate the statistic that was used to order the resources. The same is true for R_{min} except its position need to be the last possible position

and still be returned. For example if the top 10 resources are being selected out of 100 returned then the position should be 10, not 100.

5 Problem 11.5

How many papers dealing with term dependency can you find in the SIGIR proceedings since 2000? List their citations.

5.1 Solution

I queried google scholar for all papers matching the criteria "term dependency" OR "term dependent" OR "term dependence" source:SIGIR and it had to have been published in the range of 2000-2017. I got back 8 pages with 78 results. I downloaded Publish or Perish software <http://www.harzing.com/pop.htm> to download the results and scrape the pages for all the information I needed.

Title	Venue	Citations
Incorporating term dependency in the DFR framework	... international ACM SIGIR ...	56
Exploiting term dependence while handling negation in medical search	... international ACM SIGIR ...	21
Score-safe term-dependency processing with hybrid indexes	... of the 37th international ACM SIGIR ...	10
Modelling term dependence with copulas	... of the 38th international ACM SIGIR ...	8
Parameterized fielded term dependence models for ad-hoc entity retrieval from knowledge graph	... of the 39th International ACM SIGIR ...	10
Non-compositional term dependence for information retrieval	... International ACM SIGIR ...	7
Utilizing phrase based semantic information for term dependency	... of the 31st annual international ACM SIGIR ...	0
Improving Search using Proximity-Based Statistics	... 38th International ACM SIGIR Conference on Research ...	1
Introduction to probabilistic models in IR	... 33rd international ACM SIGIR conference on Research ...	5

A comparison of various approaches for using probabilistic dependencies in language modeling	... of the 26th annual international ACM SIGIR ...	16
Two-stage query segmentation for information retrieval	... of the 32nd international ACM SIGIR ...	49
Refining term weights of documents using term dependencies	... of the 27th annual international ACM SIGIR ...	9
Query performance prediction in web search environments	... of the 30th annual international ACM SIGIR ...	169
MRF based approach for sentence retrieval	... annual international ACM SIGIR ...	6
Integrating word relationships into language models	... of the 28th annual international ACM SIGIR ...	184
Building and applying a concept hierarchy representation of a user profile	... 26th annual international ACM SIGIR ...	80
A Markov random field model for term dependencies	... of the 28th annual international ACM SIGIR ...	776
Learning to efficiently rank	... of the 33rd international ACM SIGIR ...	74
A study of Poisson query generation model for information retrieval	... of the 30th annual international ACM SIGIR ...	52
Automatic Phrase Indexing for Document Retrieval: An Examination of Syntactic and Non-Syntactic Methods	ACM SIGIR Forum	169
Random walk term weighting for information retrieval	... of the 30th annual international ACM SIGIR ...	35
Social annotation in query expansion: a machine learning approach	... of the 34th international ACM SIGIR ...	39
Latent concept expansion using markov random fields	... of the 30th annual international ACM SIGIR ...	226
Combining concepts and language models for information access	SIGIR Forum	5
A proximity language model for information retrieval	Proceedings of the 32nd international ACM SIGIR ...	69
Learning to reweight terms with distributed representations	Proceedings of the 38th International ACM SIGIR ...	35

Building a web test collection using social media	Proceedings of the 36th international ACM SIGIR ...	3
Structured retrieval for question answering	... international ACM SIGIR ...	106
An exploration of proximity measures in information retrieval	... of the 30th annual international ACM SIGIR ...	245
Positional language models for information retrieval	Proceedings of the 32nd international ACM SIGIR ...	198
Intent-aware search result diversification	... of the 34th international ACM SIGIR ...	98
Sigir 2014 workshop on semantic matching in information retrieval	... of the 37th international ACM SIGIR ...	4
Flat vs. hierarchical phrase-based translation models for cross-language information retrieval	... 36th international ACM SIGIR conference on Research ...	3
Incorporating query term dependencies in language models for document retrieval	... of the 26th annual international ACM SIGIR ...	19
Fielded sequential dependence model for ad-hoc entity retrieval in the web of data	... of the 38th International ACM SIGIR ...	33
Dependence language model for information retrieval	... of the 27th annual international ACM SIGIR ...	294
Axiomatic analysis for improving the log-logistic feedback model	... International ACM SIGIR ...	8
Efficient & Effective Selective Query Rewriting with Efficiency Predictions	... of the 40th International ACM SIGIR ...	0
Query term ranking based on dependency parsing of verbose queries	Proceedings of the 33rd international ACM SIGIR ...	28
CRTER: using cross terms to enhance probabilistic information retrieval	... of the 34th international ACM SIGIR ...	32
Term Proximity Constraints for Pseudo-Relevance Feedback	... International ACM SIGIR ...	0
Modeling higher-order term dependencies in information retrieval using query hypergraphs	... of the 35th international ACM SIGIR conference ...	54
Topic-based index partitions for efficient and effective selective search	SIGIR 2010 Workshop on Large-Scale ...	16

Building simulated queries for known-item topics: an analysis using six european languages	... 30th annual international ACM SIGIR ...	97
Dynamic Factual Summaries for Entity Cards	Proc. of SIGIR	2
Set-based model: A new approach for information retrieval	... international ACM SIGIR ...	38
Query term ranking based on search results overlap	... of the 34th international ACM SIGIR ...	1
Exploiting proximity feature in bigram language model for information retrieval	... the 31st annual international ACM SIGIR ...	6
Exploiting semantics for improving clinical information retrieval	... the 36th international ACM SIGIR ...	21
A frequency-based and a poisson-based definition of the probability of being informative	... of the 26th annual international ACM SIGIR ...	39
Efficient cost-aware cascade ranking in multi-stage retrieval	... International ACM SIGIR ...	3
An adaptive evidence weighting method for medical record search	Proceedings of the 36th international ACM SIGIR ...	14
Using key concepts in a translation model for retrieval	Proceedings of the 38th International ACM SIGIR ...	8
Understanding negation and family history to improve clinical information retrieval	... of the 37th international ACM SIGIR conference ...	6
Nordlys: A Toolkit for Entity-Oriented and Semantic Search	... 40th International ACM SIGIR ...	0
Exploring reductions for long web queries	... international ACM SIGIR ...	74
Learning for search result diversification	... the 37th international ACM SIGIR ...	42
Color retrieval in vector space model	... ACM SIGIR Workshop ...	12
Compact query term selection using topically related text	Proceedings of the 36th international ACM SIGIR ...	38
DBpedia-Entity v2: A Test Collection for Entity Search	... ACM SIGIR ...	1
Embedding-based Query Expansion for Weighted Sequential Dependence Retrieval Model	... of the 40th International ACM SIGIR ...	0

On the cost of phrase-based ranking	Proceedings of the 38th International ACM SIGIR ...	1
Extending BM25 with multiple query operators	Proceedings of the 35th international ACM SIGIR ...	14
Learning for Efficient Supervised Query Expansion via Two-stage Feature Selection	... of the 39th International ACM SIGIR ...	2
Parameterized concept weighting in verbose queries	... of the 34th international ACM SIGIR ...	100
Diversity by proportionality: an election-based approach to search result diversification	Proceedings of the 35th international ACM SIGIR ...	93
Paradox-free formal foundation of vector-space model	... of the ACM SIGIR 2002 workshop on ...	4
Efficient in-memory top-k document retrieval	... of the 35th international ACM SIGIR ...	35
Copulas for information retrieval	... international ACM SIGIR ...	19
Automatic refinement of patent queries using concept importance predictors	... international ACM SIGIR ...	34
Modeling Document Novelty with Neural Tensor Network for Search Result Diversification	... of the 39th International ACM SIGIR ...	10
Term level search result diversification	Proceedings of the 36th international ACM SIGIR ...	36
Learning user reformulation behavior for query auto-completion	... the 37th international ACM SIGIR ...	40
A Probabilistic Model for Information Retrieval Based on Maximum Value Distribution	... 38th International ACM SIGIR Conference on Research ...	3
Enhancing ad-hoc relevance weighting using probability density estimation	... of the 34th international ACM SIGIR ...	9
Pseudo test collections for training and tuning microblog rankers	... international ACM SIGIR ...	17
Introduction to Probabilistic Models for Information Retrieval	... of the 33rd International ACM SIGIR ...	4
MEMBER: Max-Margin Based Embeddings for Entity Retrieval	... the 40th International ACM SIGIR ...	1