CS 886 Advanced Topics in Artificial Intelligence: Multiagent Systems

Rank Aggregation Methods for the Web

Cynthia Dwork Ravi Kumar Moni Naor D. Sivakumar

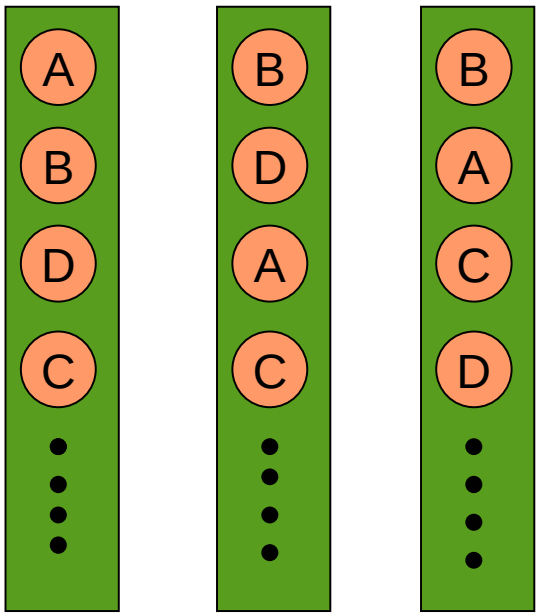
Presented by: Wanying Luo

Outline

- What is rank aggregation problem
- Motivation
- Challenges
- Preliminaries
- First result: spam resistance in meta-search
- Second result: Markov chain methods
- Applications
- Experiments
- Conclusion

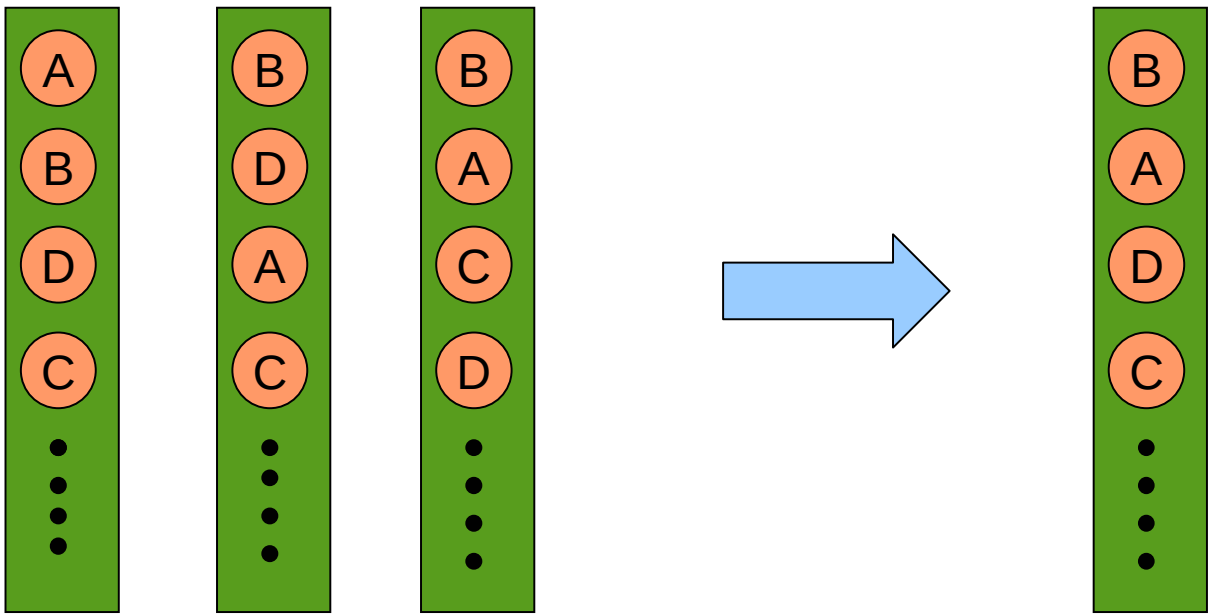
What is rank aggregation problem

Based on different ranking techniques and criteria, we may get different results



What is rank aggregation problem

Need to obtain a ”consensus” ranking of all the individual rankings



Outline

- What is rank aggregation problem
- Motivation
- Challenges
- Preliminaries
- First result: spam resistance in meta-search
- Second result: Markov chain methods
- Applications
- Experiments
- Conclusion

Motivation

- **Provide robust meta search**
  - Examples of meta search engines
    - **Clusty**
    - **Dogpile**
    - **Metacrawler**
  - Spam
    - http://searchenginewatch.com/showPage.html?page=3483601
  - Commercial interests, e.g., sponsored links

web news images wikipedia blogs jobs more »

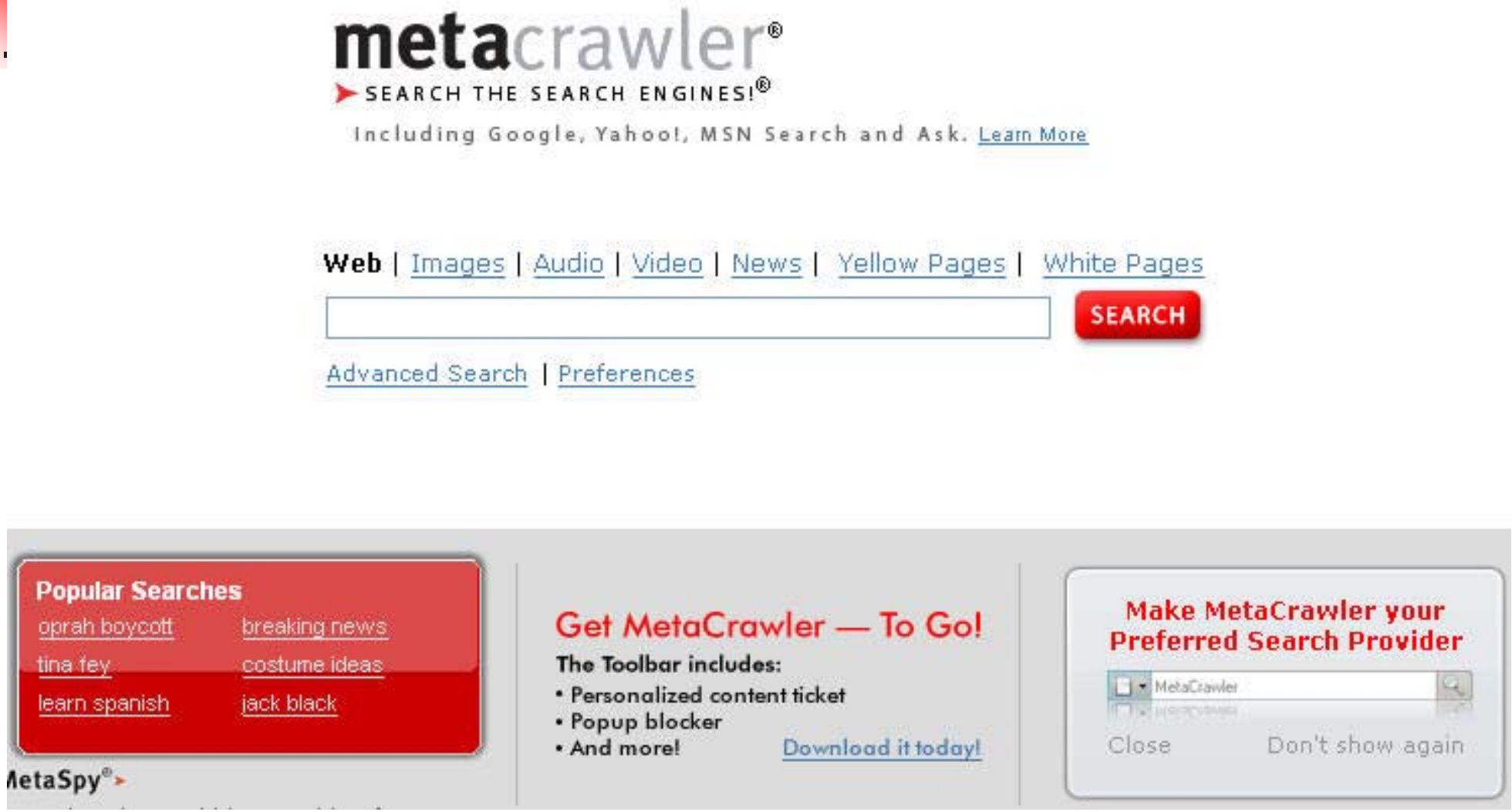
Search

advanced
preferences

Clusty introduces Clustering 2.0

about privacy toolbars site search technology contact us


dogpile®
All the best search engines piled into one.
Web
Images
Audio
Video
News
Yellow Pages
White Pages
Go Fetch!
Google
YAHOO!
Live Search
Ask
Advanced Search
Preferences
Favorite Fetches
aig bailout
school house rock
unconstitutional pants
dow jones
pimp my dorm
constitution day
Quick Starts
Download Desktop Arfie
Add Dogpile To Your Site
Download Toolbar
Laugh Out Loud Pets
SearchSpy
Horoscope


metacrawler®
SEARCH THE SEARCH ENGINES!®
Including Google, Yahoo!, MSN Search and Ask. Learn More
Web | Images | Audio | Video | News | Yellow Pages | White Pages
SEARCH
Advanced Search | Preferences
Popular Searches
oprah boycott
breaking news
tina fey
costume ideas
learn spanish
jack black
MetaSpy®
Get MetaCrawler — To Go!
The Toolbar includes:
• Personalized content ticket
• Popup blocker
• And more!
Download it today!
Make MetaCrawler your Preferred Search Provider
MetaCrawler
Close
Don't show again

Motivation

- **Provide robust meta search**
  - Examples of meta search engines
    - **Clusty**
    - **Dogpile**
    - **Metacrawler**
  - Spam
    - http://searchenginewatch.com/showPage.html?page=3483601
  - Commercial interests, e.g., sponsored links

Motivation

- **Provide robust meta search**
  - Examples of meta search engines
    - **Clusty**
    - **Dogpile**
    - **Metacrawler**
  - Spam
    - http://searchenginewatch.com/showPage.html?page=3483601
  - Commercial interests, e.g., sponsored links

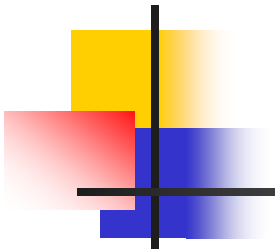
Motivation

- **User may provide a variety of searching criteria**

Outline

- What is rank aggregation problem
- Motivation
- Challenges
- Preliminaries
- First result: spam resistance in meta-search
- Second result: Markov chain methods
- Applications
- Experiments
- Conclusion

Challenges



- **Unrealistic to rank the entire collection of pages on the web**
  - **29.7 billion pages on the World Wide Web as of February 2007 (http://www.boutell.com/)**
- **Most search engines rank only the top few hundred entries**

Challenges

- **Unrealistic to rank the entire collection of pages on the web**
  - **29.7 billion pages on the World Wide Web as of February 2007 (http://www.boutell.com/)**
- **Most search engines rank only the top few hundred entries**

Outline

- What is rank aggregation problem
- Motivation
- Challenges
- Preliminaries
- First result: spam resistance in meta-search
- Second result: Markov chain methods
- Applications
- Experiments
- Conclusion

Preliminaries

- **Ordered list**

  Given a universe U, an ordered list τ with respect to U is an ordering(aka ranking) of a subset S $\subseteq$ U, i.e.

  $\tau=[\chi_1 \geq \chi_2 \geq \ldots \geq \chi_d]$

- **Full list**

  τ contains all the elements in U

- **Partial list**

  $|\tau| < |U|$

Preliminaries

- **Ordered list**

  Given a universe U, an ordered list τ with respect to U is an ordering(aka ranking) of a subset S $\subseteq$ U, i.e.

  $\tau=[\chi_1 \geq \chi_2 \geq \ldots \geq \chi_d]$

- **Full list**

  τ contains all the elements in U

- **Partial list**

  $|\tau| < |U|$

Preliminaries

- **Ordered list**

  Given a universe U, an ordered list τ with respect to U is an ordering(aka ranking) of a subset S $\subseteq$ U, i.e.

  $\tau=[\chi_1 \geq \chi_2 \geq \ldots \geq \chi_d]$

- **Full list**

  τ contains all the elements in U

- **Partial list**

  $|\tau| < |U|$

Preliminaries

- **Rank aggregation approach**
  - **Goal: minimize the total disagreement between several rankings**
  - **Spearman footrule distance**
    - Given two full lists σ and τ, $F(\sigma,\tau)=\sum_{i=1}|\sigma(i)-\tau(i)|$
  - **Kendall tau distance**
    - Given two full lists σ and τ, $K(\sigma,\tau)=|\{(i,j) \mid i<j, \sigma(i)<\sigma(j) \text{ but } \tau(i)>\tau(j)\}|$
  - **These two measurements can be generalized to several lists**
  - **Can also be generalized to partial lists**

Preliminaries

- **Rank aggregation approach**
  - **Goal: minimize the total disagreement between several rankings**
  - **Spearman footrule distance**
    - Given two full lists σ and τ, $F(\sigma,\tau)=\sum_{i=1}|\sigma(i)-\tau(i)|$
  - **Kendall tau distance**
    - Given two full lists σ and τ, $K(\sigma,\tau)=|\{(i,j) \mid i<j, \sigma(i)<\sigma(j) \text{ but } \tau(i)>\tau(j)\}|$
  - **These two measurements can be generalized to several lists**
  - **Can also be generalized to partial lists**

Preliminaries

- **Rank aggregation approach**
  - **Goal: minimize the total disagreement between several rankings**
  - **Spearman footrule distance**
    - Given two full lists σ and τ, $F(\sigma,\tau)=\sum_{i=1}|\sigma(i)-\tau(i)|$
  - **Kendall tau distance**
    - Given two full lists σ and τ, $K(\sigma,\tau)=|\{(i,j) \mid i<j, \sigma(i)<\sigma(j) \text{ but } \tau(i)>\tau(j)\}|$
  - **These two measurements can be generalized to several lists**
  - **Can also be generalized to partial lists**

Preliminaries

- **Rank aggregation approach**
  - **Goal: minimize the total disagreement between several rankings**
  - **Spearman footrule distance**
    - Given two full lists σ and τ, $F(\sigma,\tau)=\sum_{i=1}|\sigma(i)-\tau(i)|$
  - **Kendall tau distance**
    - Given two full lists σ and τ, $K(\sigma,\tau)=|\{(i,j) \mid i<j, \sigma(i)<\sigma(j) \text{ but } \tau(i)>\tau(j)\}|$
  - **These two measurements can be generalized to several lists**
  - **Can also be generalized to partial lists**

Preliminaries

- **Rank aggregation approach**
  - **Goal: minimize the total disagreement between several rankings**
  - **Spearman footrule distance**
    - Given two full lists σ and τ, $F(\sigma,\tau)=\sum_{i=1}|\sigma(i)-\tau(i)|$
  - **Kendall tau distance**
    - Given two full lists σ and τ, $K(\sigma,\tau)=|\{(i,j) \mid i<j, \sigma(i)<\sigma(j) \text{ but } \tau(i)>\tau(j)\}|$
  - **These two measurements can be generalized to several lists**
  - **Can also be generalized to partial lists**

Outline

- What is rank aggregation problem
- Motivation
- Challenges
- Preliminaries
- First result: spam resistance in meta-search
- Second result: Markov chain methods
- Applications
- Experiments
- Conclusion

First result: spam resistance in meta-search

- **Extended Condorcet Criterion (ECC)**
  - If there is a partition (C, C') of S such that for any $x \in C$ and $y \in C'$ the majority prefers x to y, then x must be ranked above y
- **ECC can be used to fight spam in meta-search**
- **How to achieve ECC efficiently**
  - Local Kemenization method

First result: spam resistance in meta-search

- **Extended Condorcet Criterion (ECC)**
  - If there is a partition (C, C') of S such that for any $x \in C$ and $y \in C'$ the majority prefers x to y, then x must be ranked above y
- **ECC can be used to fight spam in meta-search**
- **How to achieve ECC efficiently**
  - Local Kemenization method

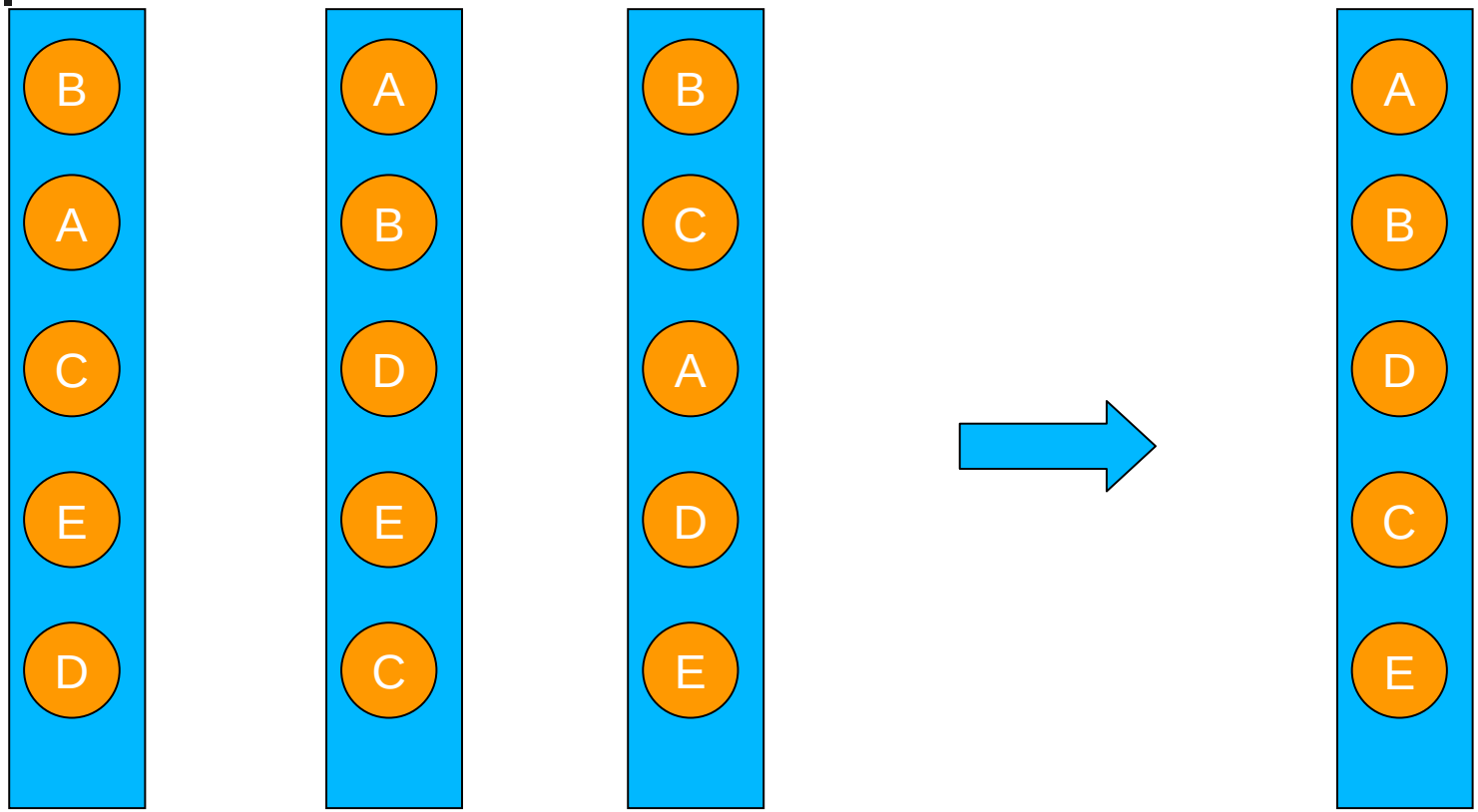
First result: spam resistance in meta-search

- **Extended Condorcet Criterion (ECC)**
  - If there is a partition (C, C') of S such that for any $x \in C$ and $y \in C'$ the majority prefers x to y, then x must be ranked above y
- **ECC can be used to fight spam in meta-search**
- **How to achieve ECC efficiently**
  - Local Kemenization method

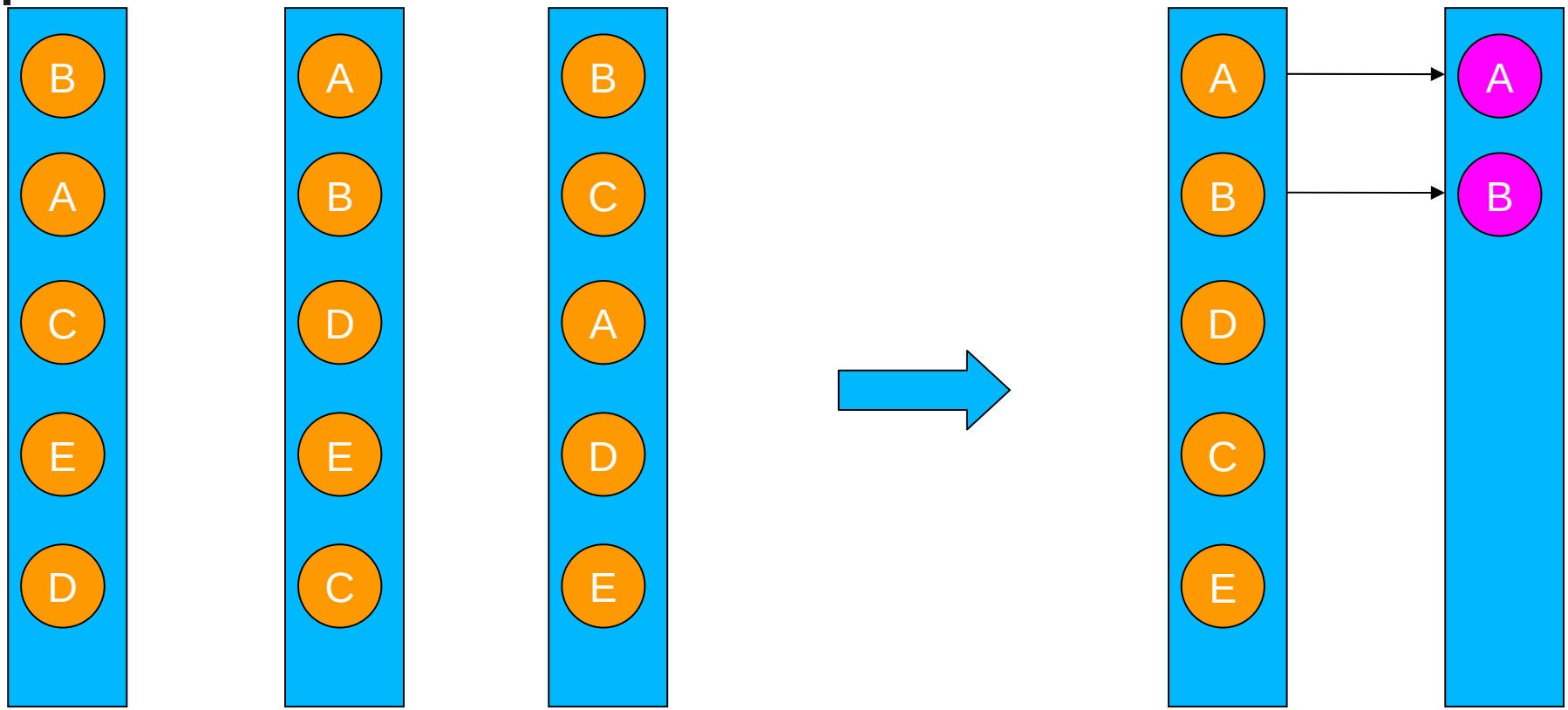
First result: spam resistance in meta-search

- **An example to illustrate Local Kemenization …**

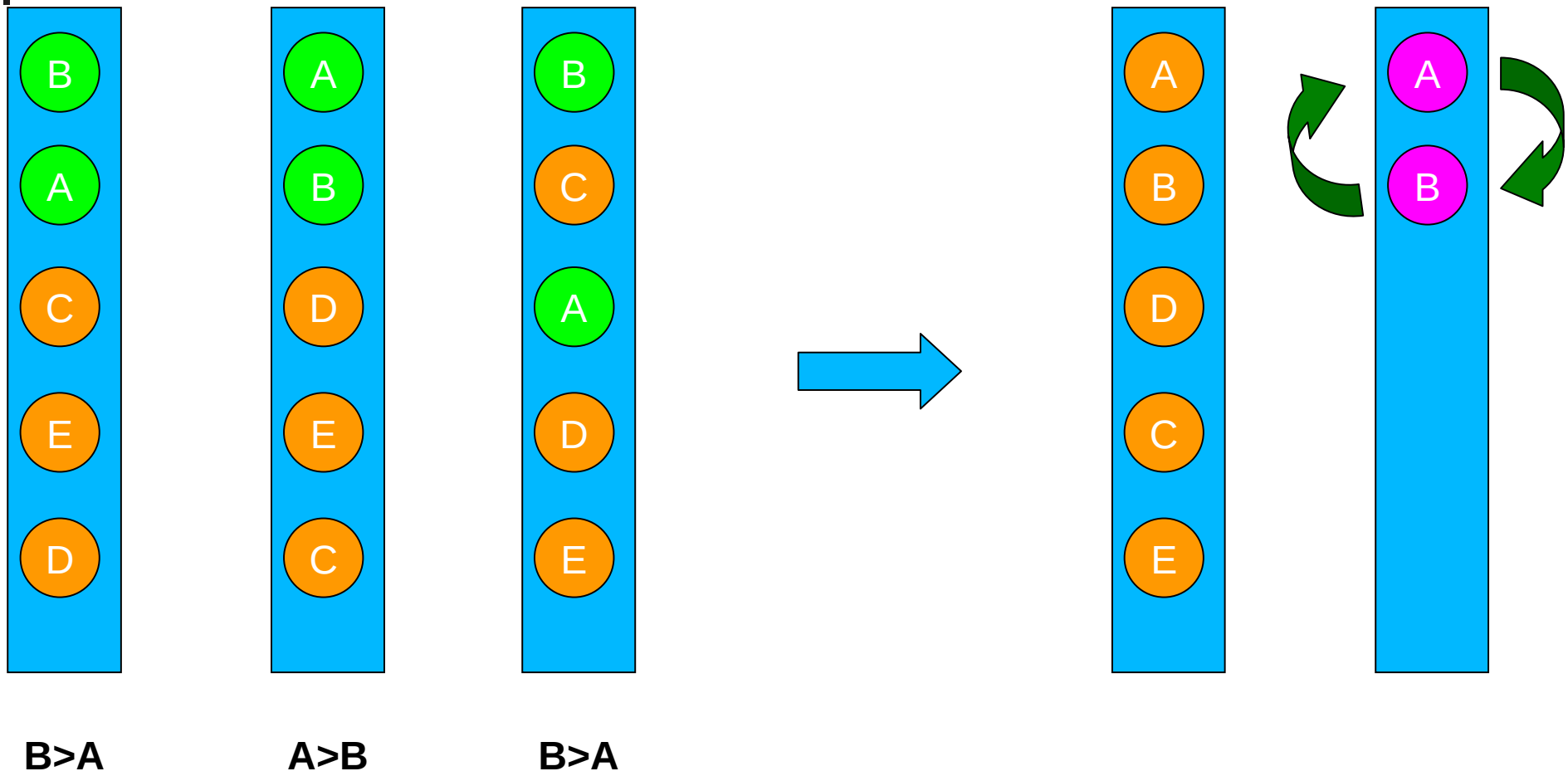
First result: spam resistance in meta-search



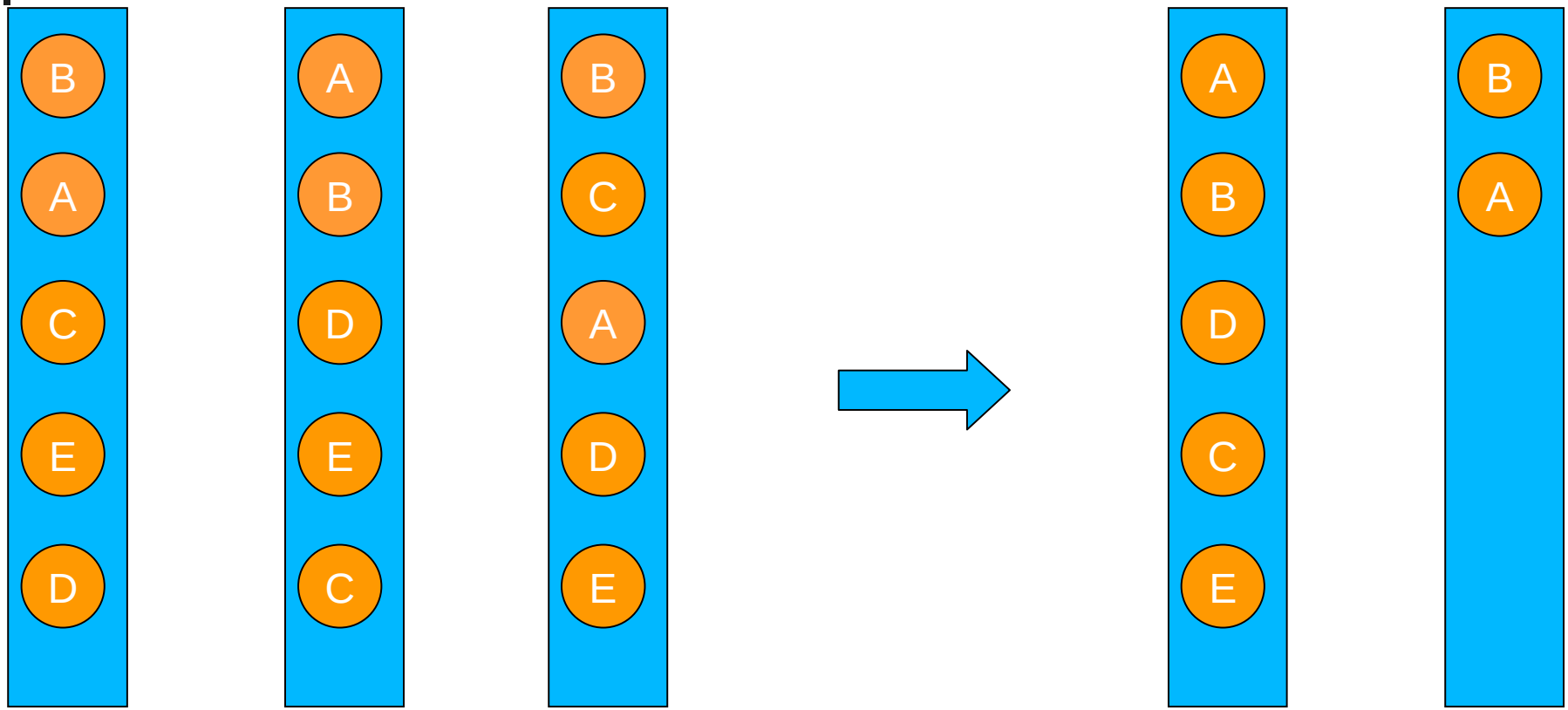
First result: spam resistance in meta-search



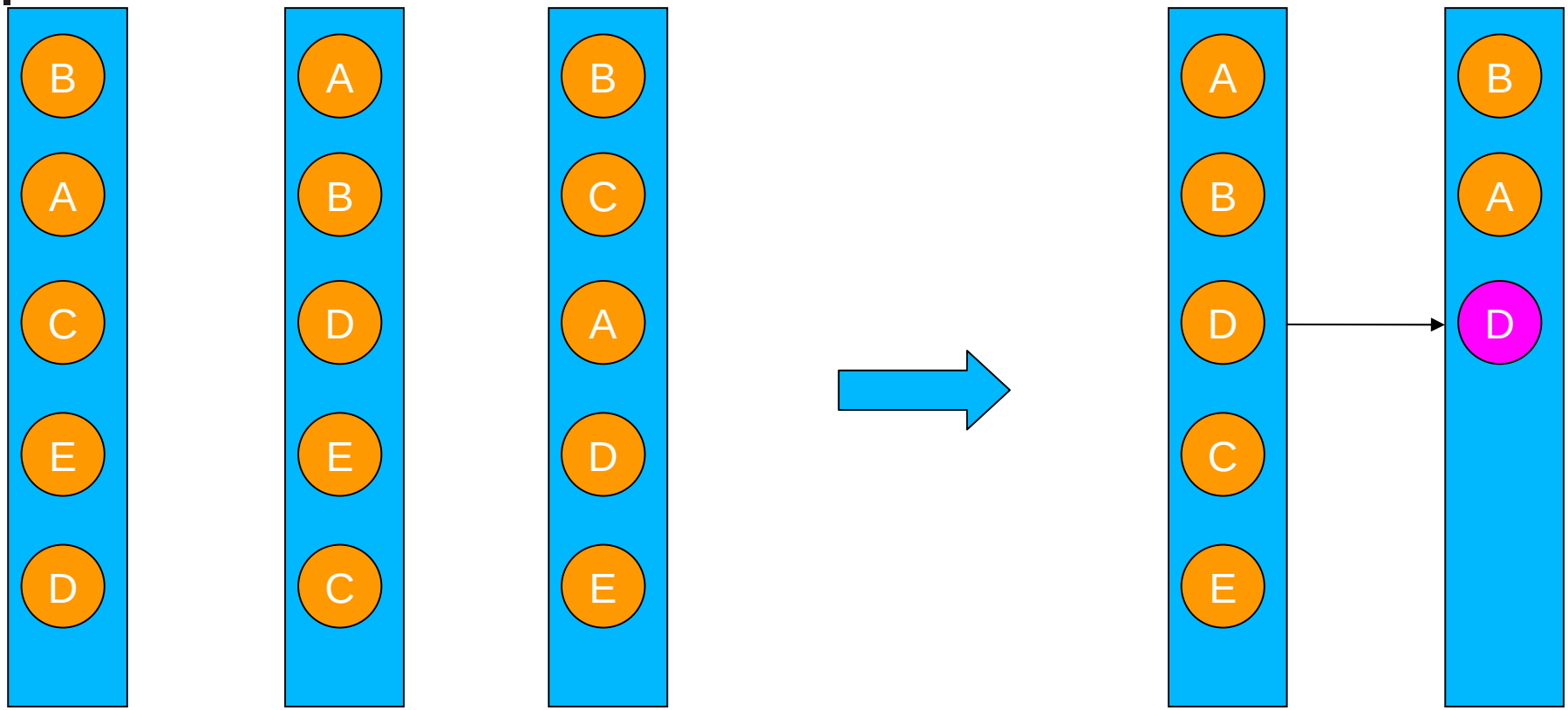
First result: spam resistance in meta-search



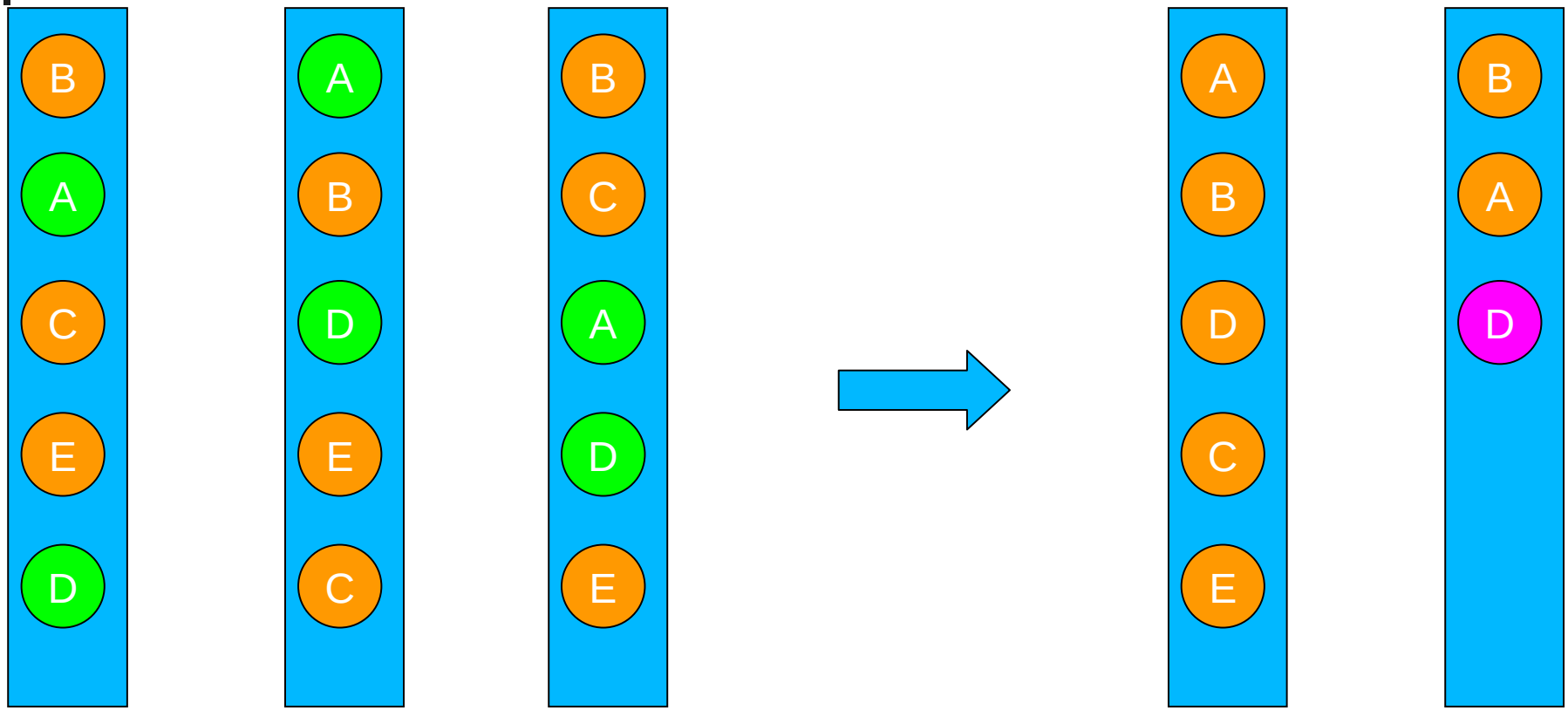
First result: spam resistance in meta-search



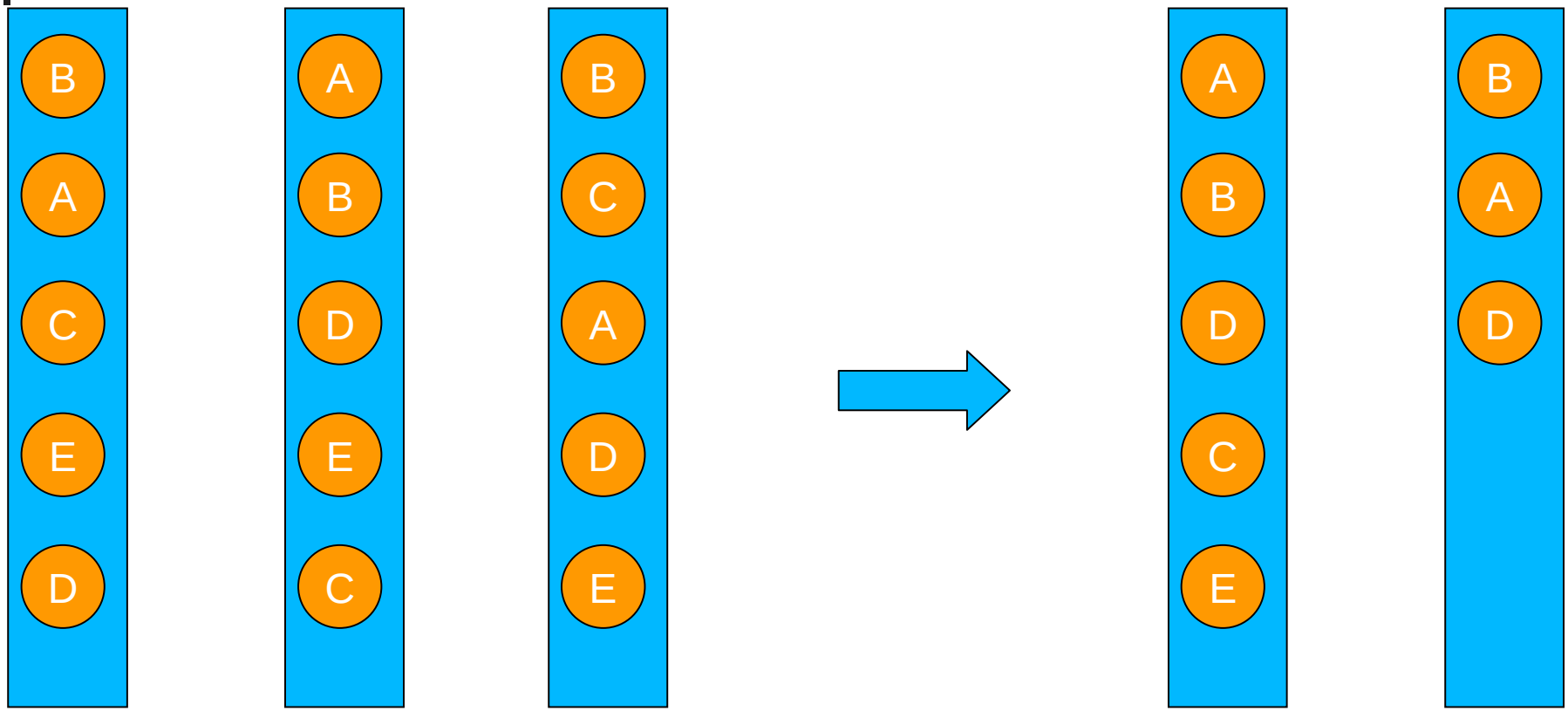
First result: spam resistance in meta-search



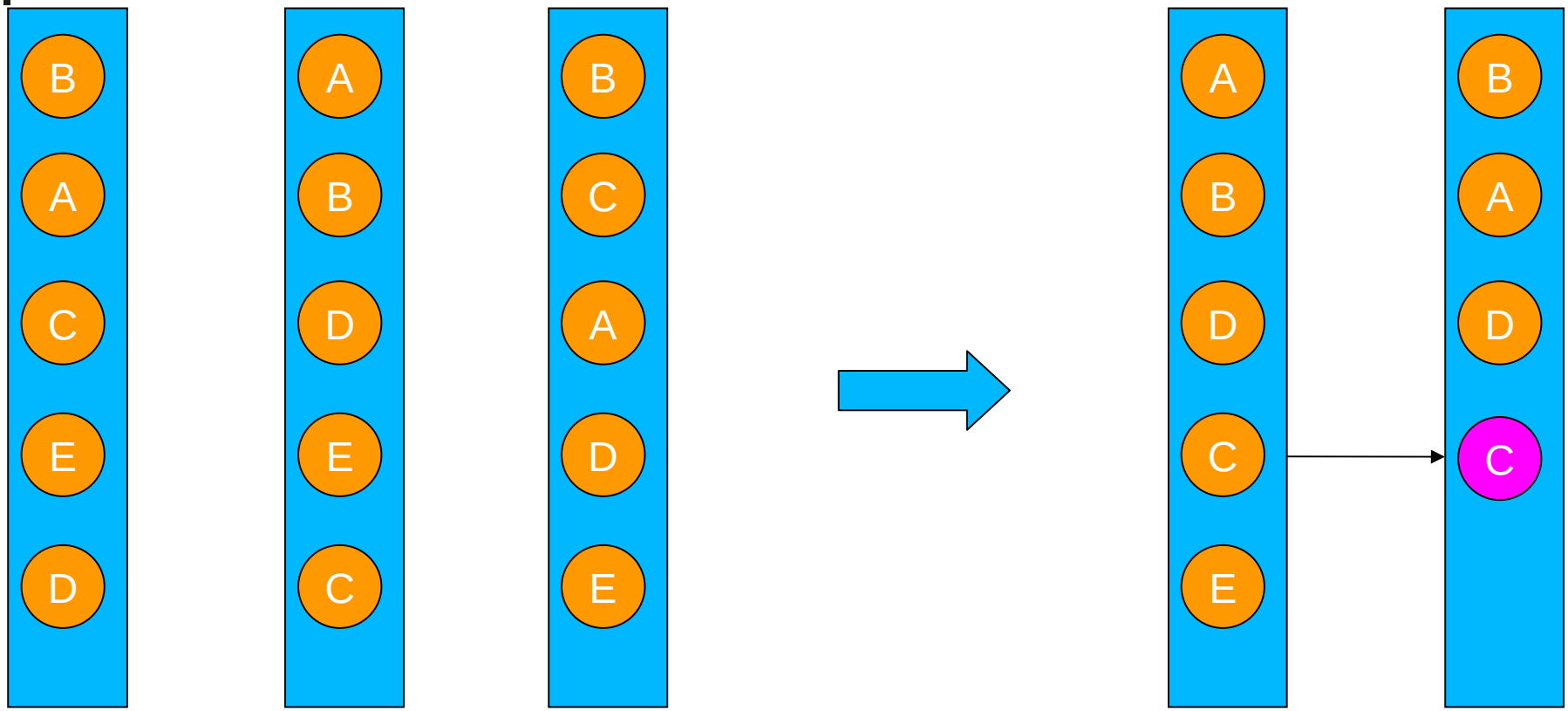
First result: spam resistance in meta-search



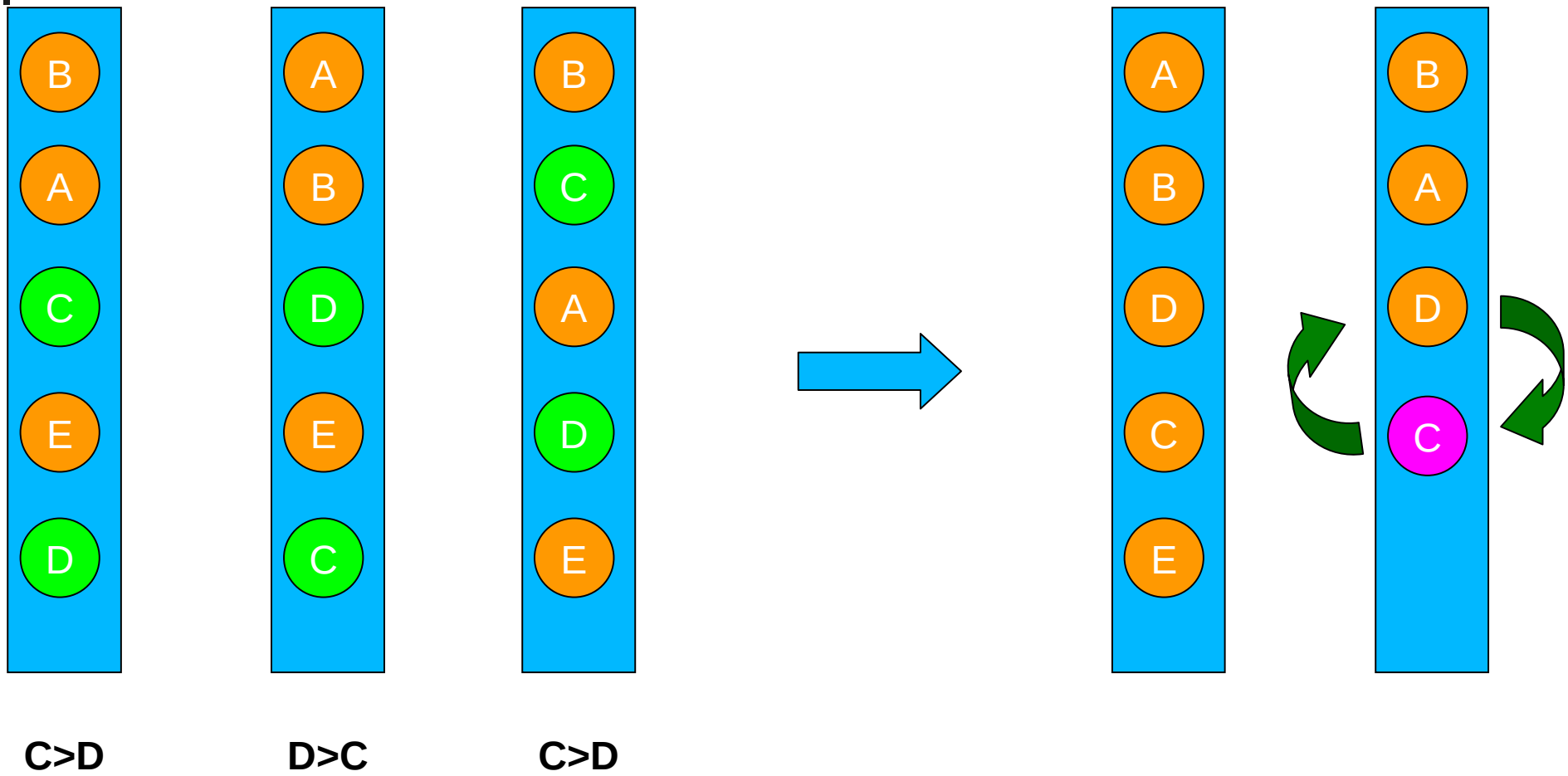
First result: spam resistance in meta-search



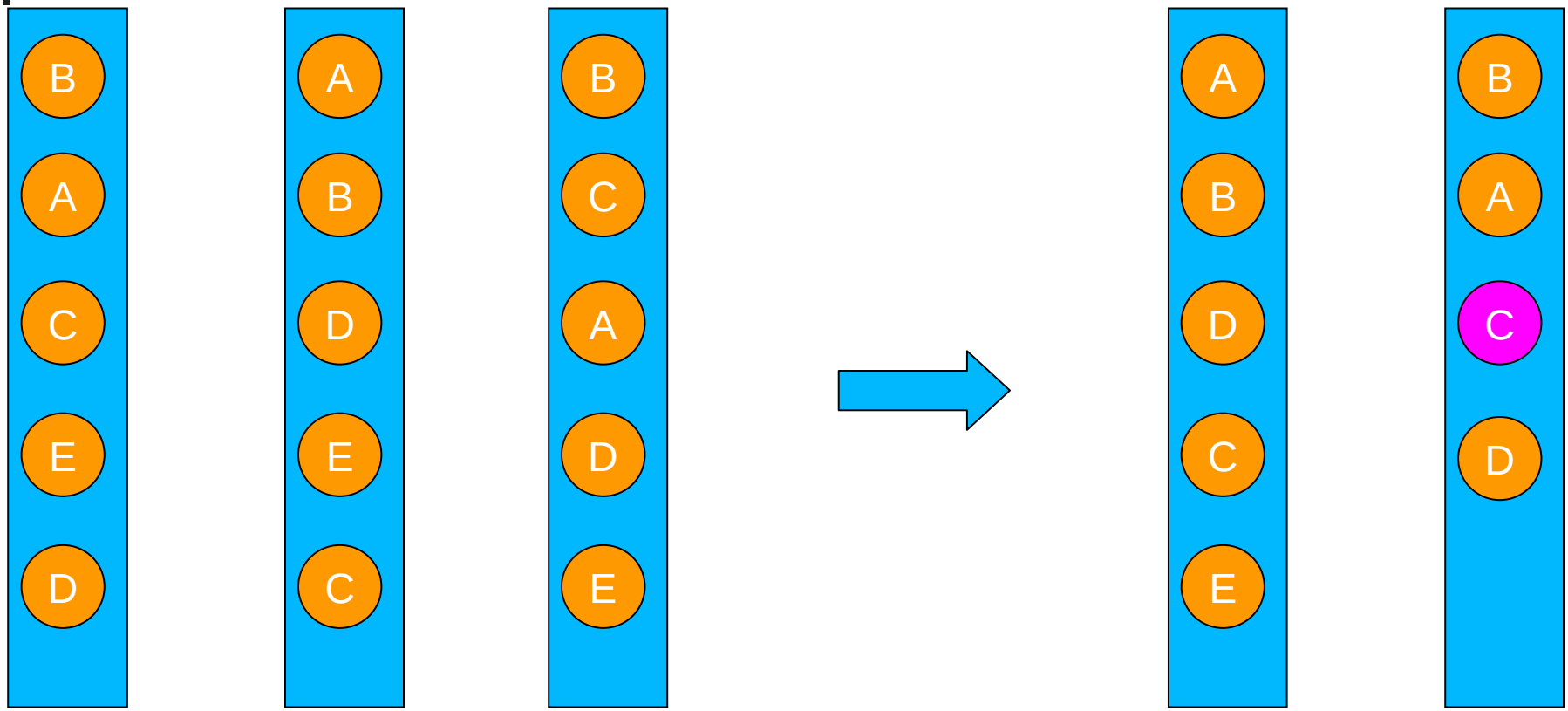
First result: spam resistance in meta-search



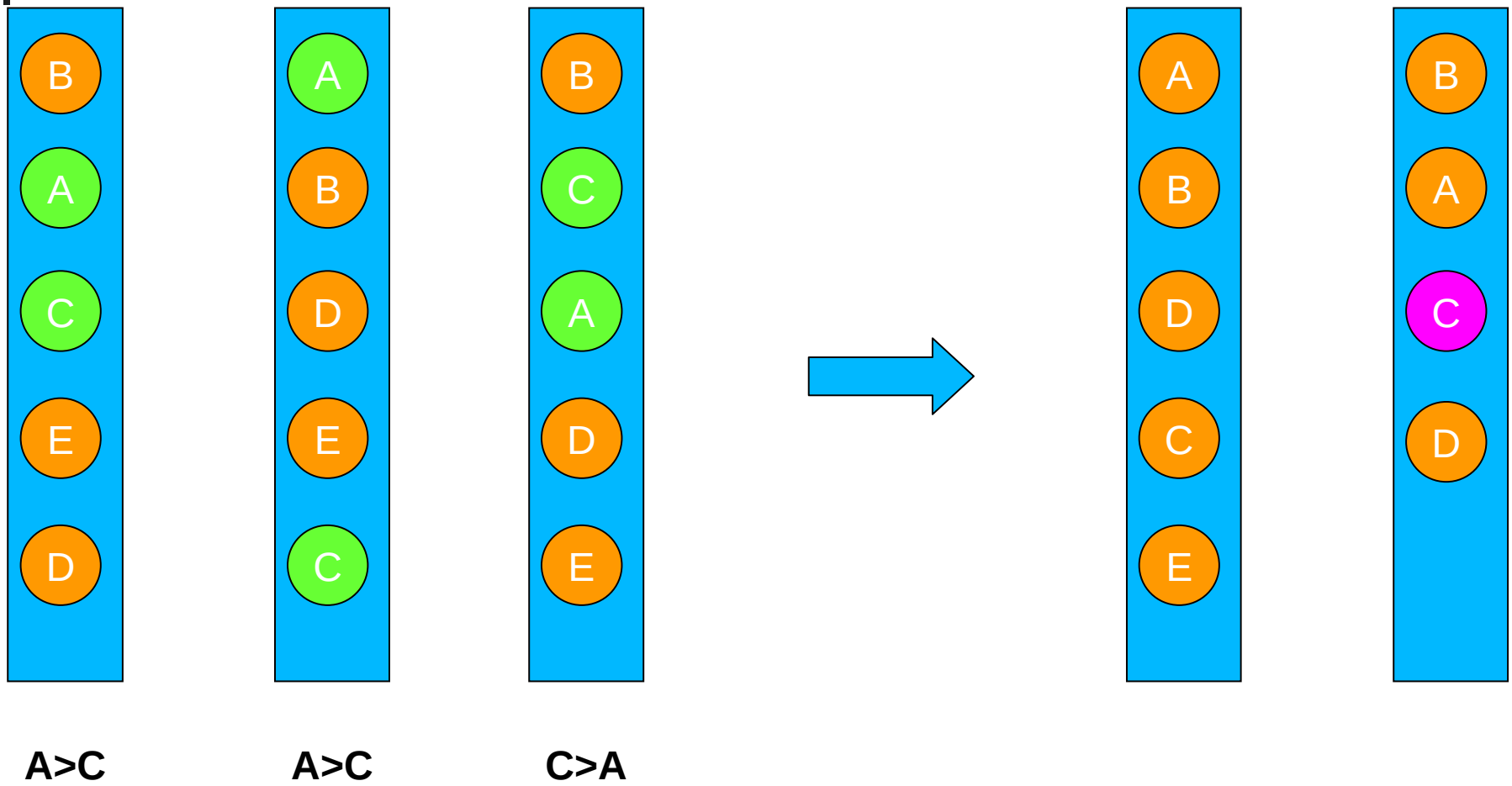
First result: spam resistance in meta-search



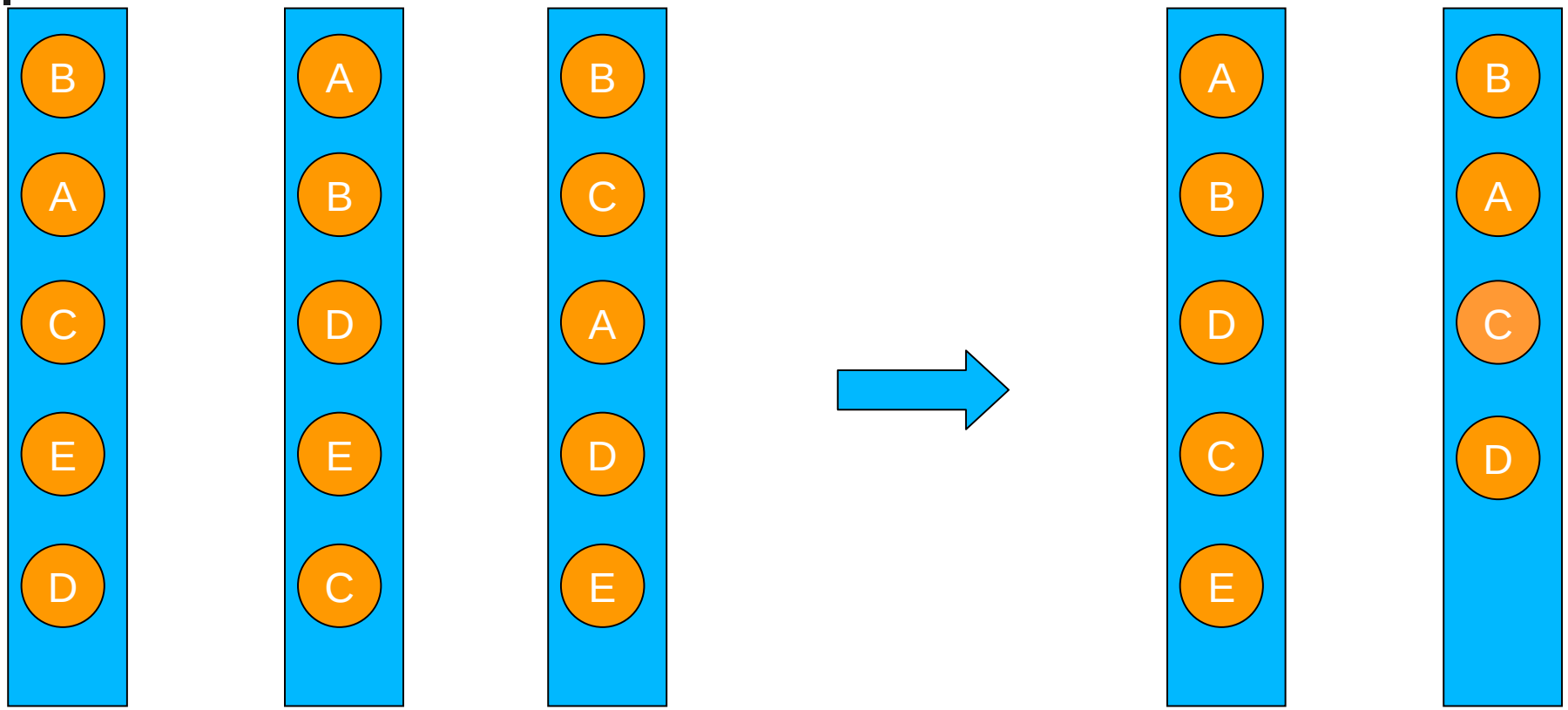
First result: spam resistance in meta-search



First result: spam resistance in meta-search



First result: spam resistance in meta-search



Outline

- What is rank aggregation problem
- Motivation
- Challenges
- Preliminaries
- First result: spam resistance in meta-search
- Second result: Markov chain methods
- Applications
- Experiments
- Conclusion

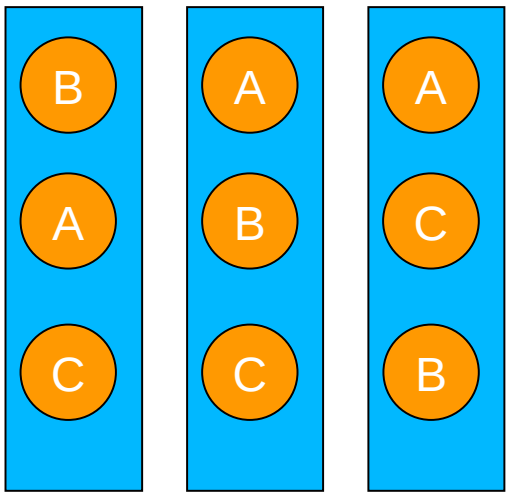
Second result: Markov chain methods

- **Markov chain**
  - **A set of states S={1,2,...,n}**
  - **An n x n matrix M**
  - **Begins with an initial state x**
  - **At each step the system moves from state i to state j with probability M_{ij}**

Second result: Markov chain methods

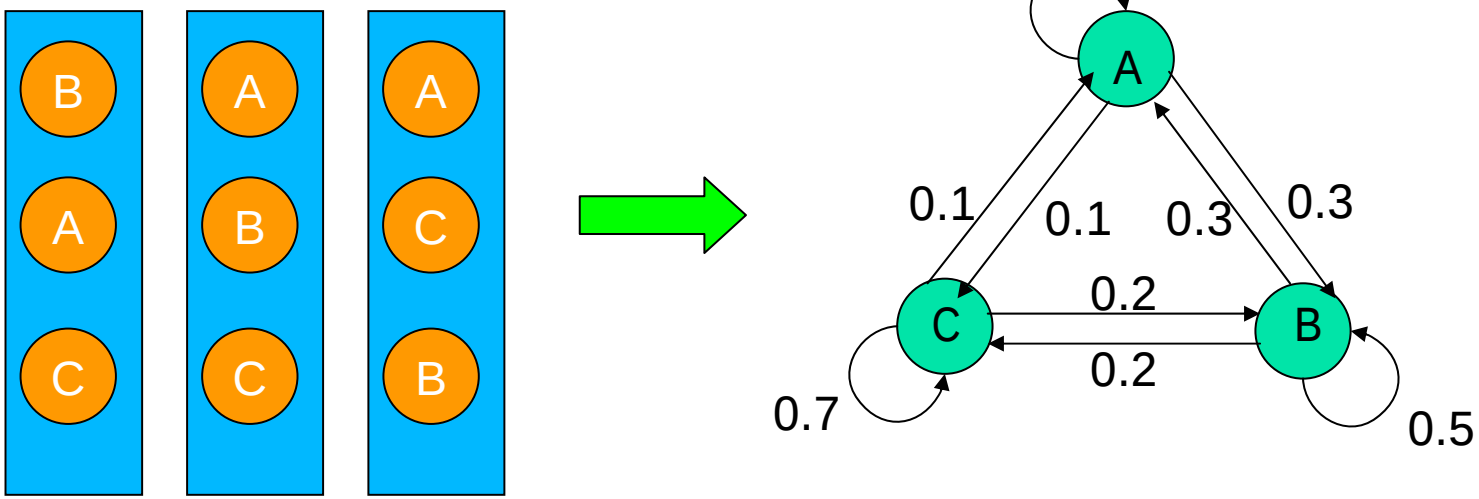
- **Under some nice condition, system eventually reaches a fixed point irrespective of the initial state x**

Second result: Markov chain methods



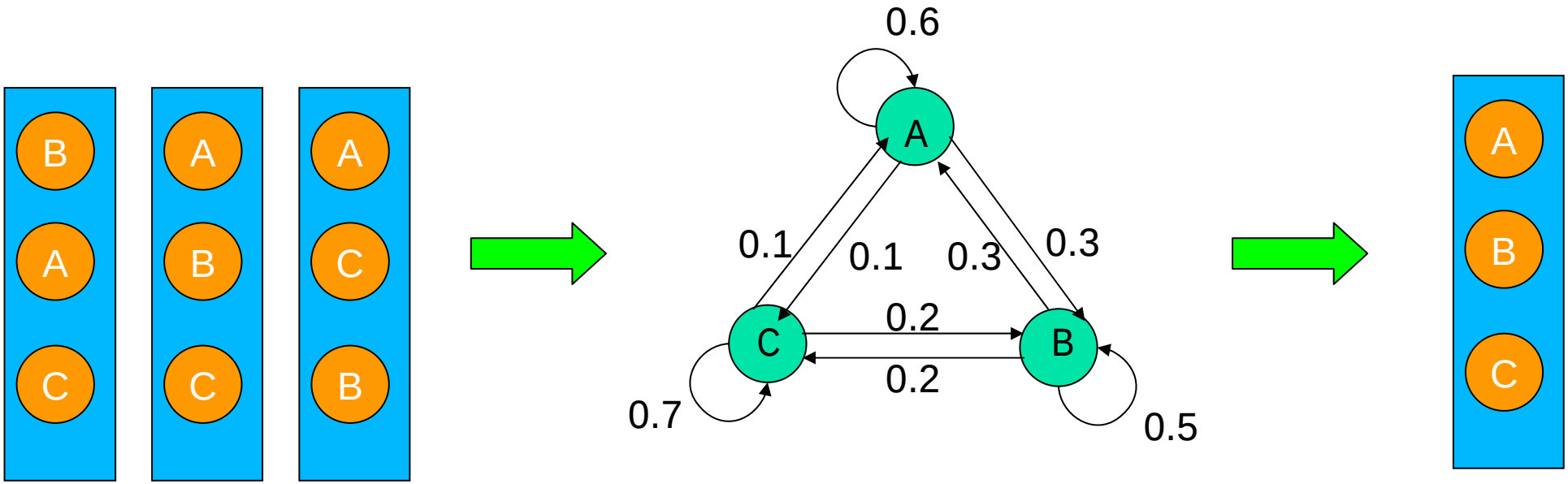
Original rankings

Second result: Markov chain methods



Original rankings

Second result: Markov chain methods



Original rankings

Aggregated ranking

Second result: Markov chain methods

- **Assume the current state is page P**
- **MC_1: The next state is chosen uniformly from the multiset of all pages that were ranked higher than or equal to P by some search engine that ranked P**
- **Please refer to the paper for the rest …**
  - **MC_2**
  - **MC_3**
  - **MC_4**

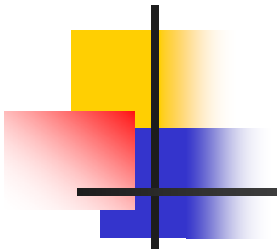
Second result: Markov chain methods

- **Assume the current state is page P**
- **MC_1: The next state is chosen uniformly from the multiset of all pages that were ranked higher than or equal to P by some search engine that ranked P**
- **Please refer to the paper for the rest …**
  - **MC_2**
  - **MC_3**
  - **MC_4**

Outline

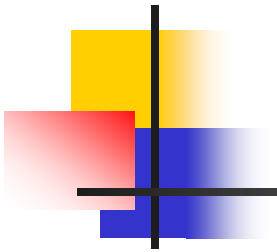
- What is rank aggregation problem
- Motivation
- Challenges
- Preliminaries
- First result: spam resistance in meta-search
- Second result: Markov chain methods
- Applications
- Experiments
- Conclusion

Applications



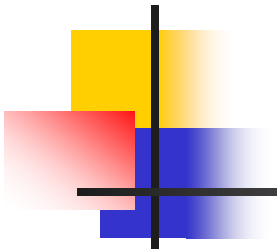
- **Meta-search**
- **Spam reduction**
- **Multi-criteria search**
- **Search engine comparison**

Applications



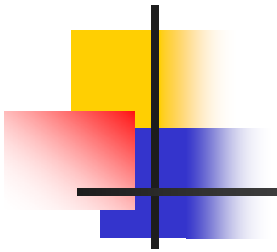
- **Meta-search**
- **Spam reduction**
- **Multi-criteria search**
- **Search engine comparison**

Applications



- **Meta-search**
- **Spam reduction**
- **Multi-criteria search**
- **Search engine comparison**

Applications

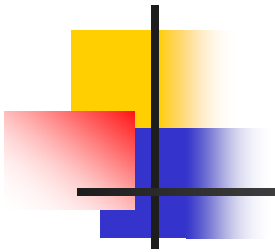


- **Meta-search**
- **Spam reduction**
- **Multi-criteria search**
- **Search engine comparison**

Outline

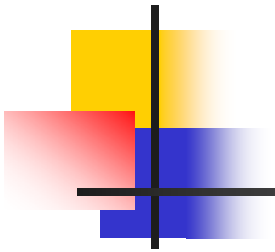
- What is rank aggregation problem
- Motivation
- Challenges
- Preliminaries
- First result: spam resistance in meta-search
- Second result: Markov chain methods
- Applications
- Experiments
- Conclusion

Experiments



- **Experiments were conducted by using the following search engines: Altavista(AV), Alltheweb(AW), Excite(EX), Google(GG), Hotbot(HB), Lycos(LY) and Northernlight(NL)**
- **Experiment on meta-search using several keywords: “affirmative action”, alcoholism, sushi, ...**

Experiments



- **Experiments were conducted by using the following search engines: Altavista(AV), Alltheweb(AW), Excite(EX), Google(GG), Hotbot(HB), Lycos(LY) and Northernlight(NL)**
- **Experiment on meta-search using several keywords: “affirmative action”, alcoholism, sushi, ...**

Experiments

	K		IF		SF	
	− LK	+ LK	− LK	+ LK	− LK	+ LK
Borda	0.221	0.214	0.353	0.345	0.440	0.438
SFO	0.112	0.111	0.168	0.167	0.137	0.137
MC_1	0.133	0.130	0.216	0.213	0.292	0.291
MC_2	0.131	0.128	0.213	0.210	0.287	0.286
MC_3	0.116	0.114	0.186	0.183	0.239	0.239
MC_4	0.105	0.104	0.151	0.149	0.181	0.181

Table 2: Performance of various rank aggregation methods for meta-search. "K" is Kendall distance, "IF" is induced footrule distance, and "SF" is scaled footrule distance. "− LK" and "+ LK", respectively, denote without and with Local Kemenization.

Experiments

	K		IF		SF	
	− LK	+ LK	− LK	+ LK	− LK	+ LK
Borda	0.221	0.214	0.353	0.345	0.440	0.438
SFO	0.112	0.111	0.168	0.167	0.137	0.137
MC_1	0.133	0.130	0.216	0.213	0.292	0.291
MC_2	0.131	0.128	0.213	0.210	0.287	0.286
MC_3	0.116	0.114	0.186	0.183	0.239	0.239
MC_4	0.105	0.104	0.151	0.149	0.181	0.181

SFO and MC_4 outperform the other 4 algorithms

MC_4 performs better than SFO most of the time

Experiments

- **Experiment on spam reduction using queries: Feng shui, organic vegetable, gardening**

Experiments

url	AV	AW	GG	HB	LY	NL	SFO	MC_4
www.lucky-bamboo.com	4	43			41		144	63
www.cambriumcrystals.com		9	51		5		31	59
www.luckycat.com	11	14	26		13		49	36
www.davesorganics.com	84	19	1		17		77	93
www.frozen.ch		9		63	11		49	121
www.eonseed.com		18		6	16		23	66
www.augusthome.com	26	16		27	12	16	57	54
www.taunton.com		25			21		78	67
www.egroups.com		34			29		108	101

Experiments

url	AV	AW	GG	HB	LY	NL	SFO	MC_4
www.lucky-bamboo.com	4	43			41		144	63
www.cambriumcrystals.com		9	51		5		31	59
www.luckycat.com	11	14	26		13		49	36
www.davesorganics.com	84	19	1		17		77	93
www.frozen.ch		9		63	11		49	121
www.eonseed.com		18		6	16		23	66
www.augusthome.com	26	16		27	12	16	57	54
www.taunton.com		25			21		78	67
www.egroups.com		34			29		108	101

Experiments

url	AV	AW	GG	HB	LY	NL	SFO	MC_4
www.lucky-bamboo.com	4	43			41		144	63
www.cambriumcrystals.com		9	51		5		31	59
www.luckycat.com	11	14	26		13		49	36
www.davesorganics.com	84	19	1		17		77	93
www.frozen.ch		9		63	11		49	121
www.eonseed.com		18		6	16		23	66
www.augusthome.com	26	16		27	12	16	57	54
www.taunton.com		25			21		78	67
www.egroups.com		34			29		108	101

Experiments

url	AV	AW	GG	HB	LY	NL	SFO	MC_4
www.lucky-bamboo.com	4	43			41		144	63
www.cambriumcrystals.com		9	51		5		31	59
www.luckycat.com	11	14	26		13		49	36
www.davesorganics.com	84	19	1		17		77	93
www.frozen.ch		9		63	11		49	121
www.eonseed.com		18		6	16		23	66
www.augusthome.com	26	16		27	12	16	57	54
www.taunton.com		25			21		78	67
www.egroups.com		34			29		108	101

Experiments

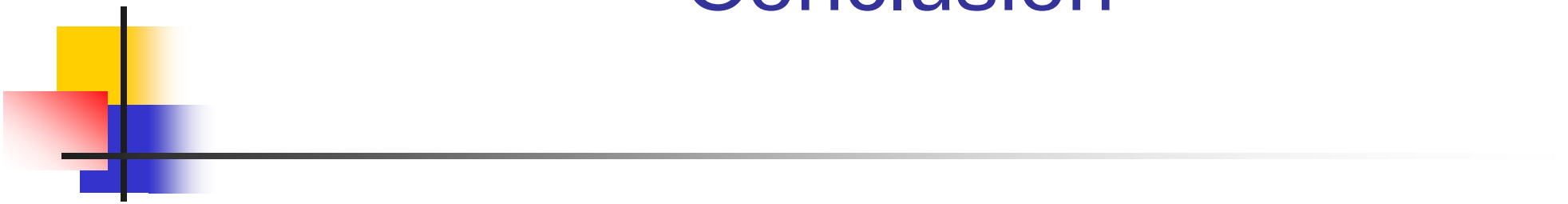
url	AV	AW	GG	HB	LY	NL	SFO	MC_4
www.lucky-bamboo.com	4	43			41		144	63
www.cambriumcrystals.com		9	51		5		31	59
www.luckycat.com	11	14	26		13		49	36
www.davesorganics.com	84	19	1		17		77	93
www.frozen.ch		9		63	11		49	121
www.eonseed.com		18		6	16		23	66
www.augusthome.com	26	16		27	12	16	57	54
www.taunton.com		25			21		78	67
www.egroups.com		34			29		108	101

Local Kemenization works!

Outline

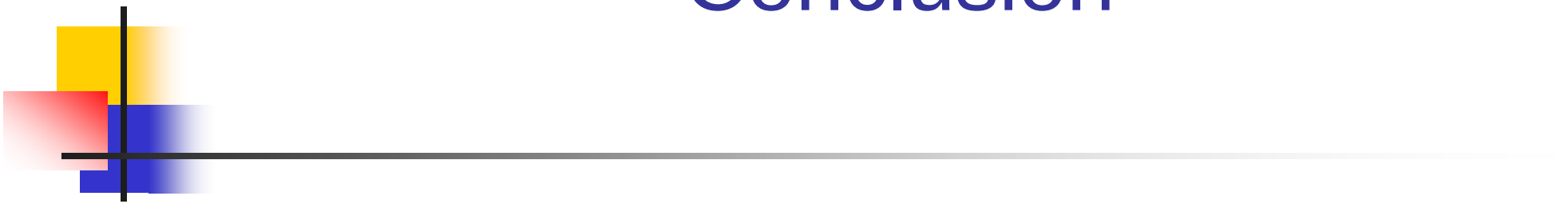
- What is rank aggregation problem
- Motivation
- Challenges
- Preliminaries
- First result: spam resistance in meta-search
- Second result: Markov chain methods
- Applications
- Experiments
- Conclusion

Conclusion



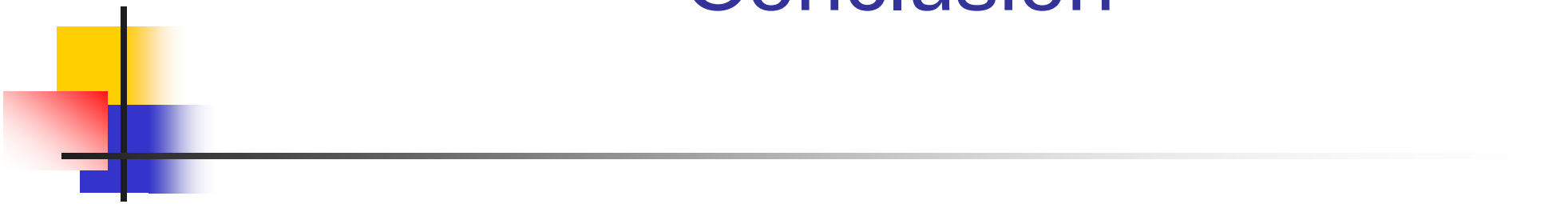
- **Proposed several rank aggregation techniques using Markov chain**
- **Established the value of Extended Condorcet Criterion (ECC)**
  - **Spam resistance**
- **Future work**
  - **Obtain a qualitative understanding of why Markov chain methods perform well**

Conclusion



- **Proposed several rank aggregation techniques using Markov chain**
- **Established the value of Extended Condorcet Criterion (ECC)**
  - **Spam resistance**
- **Future work**
  - **Obtain a qualitative understanding of why Markov chain methods perform well**

Conclusion



- **Proposed several rank aggregation techniques using Markov chain**
- **Established the value of Extended Condorcet Criterion (ECC)**
  - **Spam resistance**
- **Future work**
  - **Obtain a qualitative understanding of why Markov chain methods perform well**

Questions???