Hybrid Quantized Resource Descriptions for Geospatial Source Selection

Stefan Kufer and Andreas Henrich
stefan.kufer@uni-bamberg.de
University of Bamberg
Media Informatics Group

Shanghai, 10.11.2014

1. Motivation and Introduction

- ease of creation → huge amount of media items, maintained
  - in the world wide web
  - on private devices
  - …
- proliferation of location-aware devices: growing availability of geospatial metadata
- media items distributed across various resources (e.g. websites, peers in a P2P network, …), need for a large scale search mechanism

→ resource selection

→ additionally: constrained search, criteria: text, content features, geospatial data (focus of this work)

Distributed Indexing and Searching

- distributed indexing and search scenario:
  - Global indexing
    - single global distributed index
  - Local indexing
    - resource maintains local index and summaries (global representation of the local data)
    - query routing based on summaries

approach in this paper!

⚠ the terms "peer" and "resource" are used as synonyms!

Scenario: Resource Selection for GeoIR

- given: distributed database O
  distance measure d
  query object q
- wanted: similar data points w.r.t. query object [criterion: $d(q,o)$ with $o \in O$] (*exact* search, not *approximate* search!)
- example:

query: [Lat=31.24, Lon=121.49]

result:

1. [Lat=31.241 Lon=121.494]

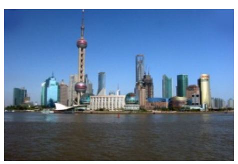
2. [Lat=31.235, Lon=121.501]

…

Scenario: Resource Selection for GeoIR

resource A


resource B




resource C

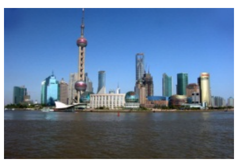

Scenario: Resource Selection for GeoIR

resource A

[Lat=48.2188,
Lon =11.6247]

[Lat=-33.858,
Lon=151.215]

resource B



[Lat=31.2304,
Lon=121.473]

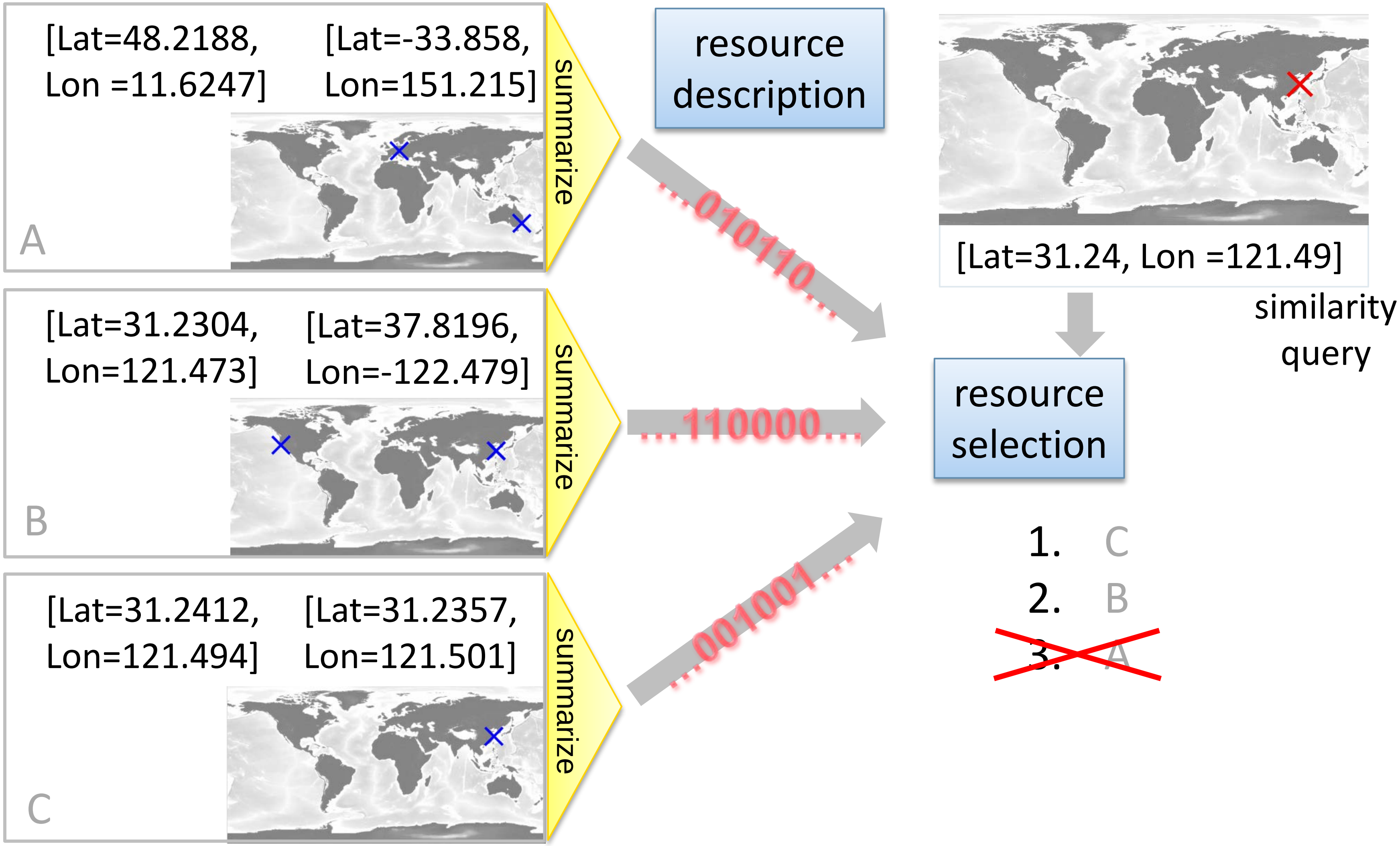
[Lat=37.8196,
Lon=-122.479]

resource C

[Lat=31.2412,
Lon=121.494]

[Lat=31.2357,
Lon=121.501]

Scenario: Resource Selection for GeoIR

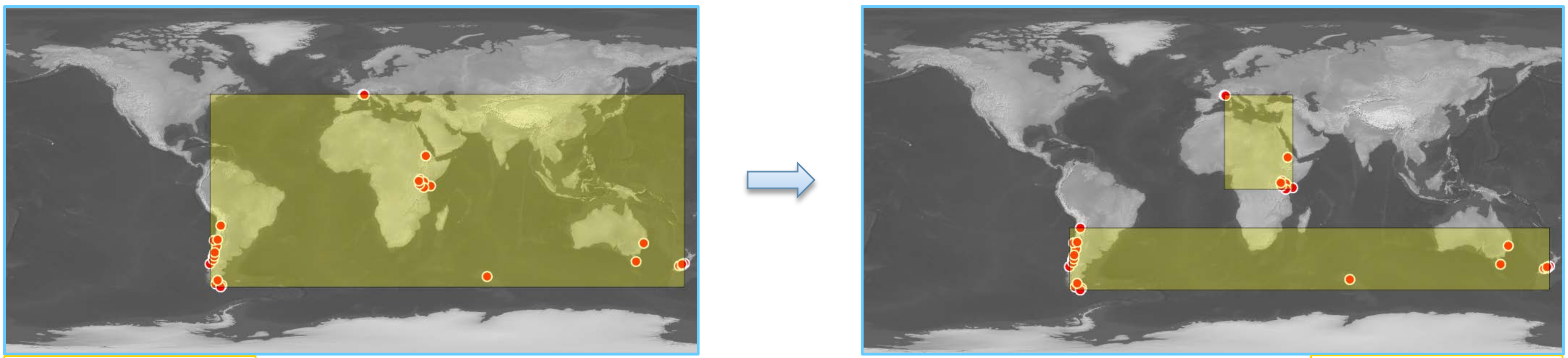


2. Categories of Resource Description Techniques

- general requirements for summaries: selectivity & space efficiency
- three categories of summary techniques for geographic data:
  - Geometric approaches
  - Space partitioning approaches
  - Hybrid approaches (combining properties of the other two)

Geometric approaches

- describe resource data with one (or several) geometric shapes
- examples:
  - Minimum Bounding Rectangle (MBR)
  - RecMAR$_k$ (**Rec**ursive calculation of up to *k* **M**inimum **A**rea **R**ectangles [MARs])



MBR

RecMAR$_2$

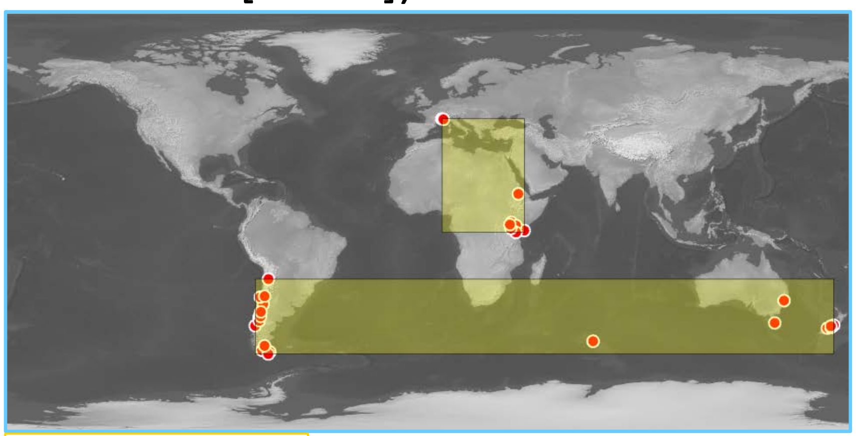
● = resource data point

MAR computation: [Becker et. al. 1991]

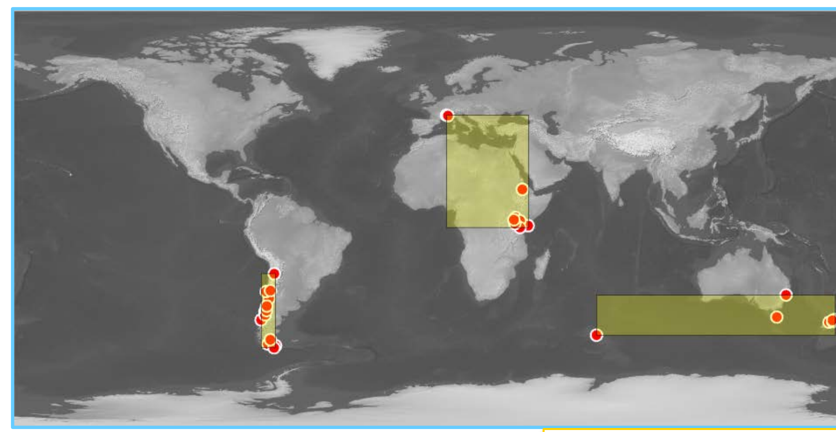
2. Categories of Resource Description Techniques

Geometric approaches

- describe resource data with one (or several) geometric shapes
- examples:
  - Minimum Bounding Rectangle (MBR)
  - RecMAR_k (**Rec**ursive calculation of up to *k* **M**inimum **A**rea **R**ectangles [MARs])



RecMAR_2



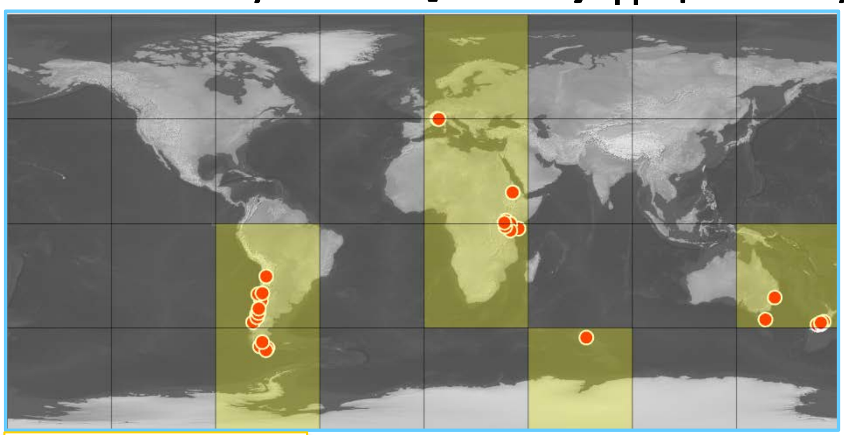
RecMAR_3

● = **resource data point**

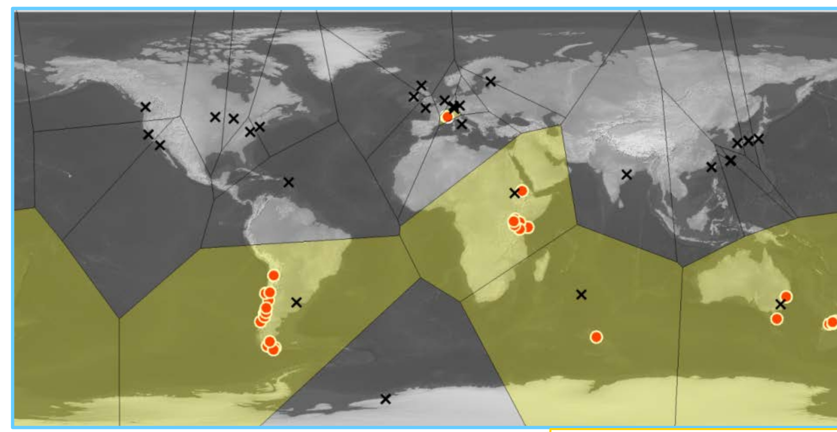
MAR computation: [Becker et. al. 1991]

Space partitioning approaches

- segment data space into *n* subspaces
- examples:
  - $Grid_r$ (r = # number of rows, number of columns = $2 \cdot r \rightarrow n = 2 \cdot r^2$)
  - UFS_n (*n = # of subspaces*)
- summary data: [**binary** || quantity] information



$Grid_4$



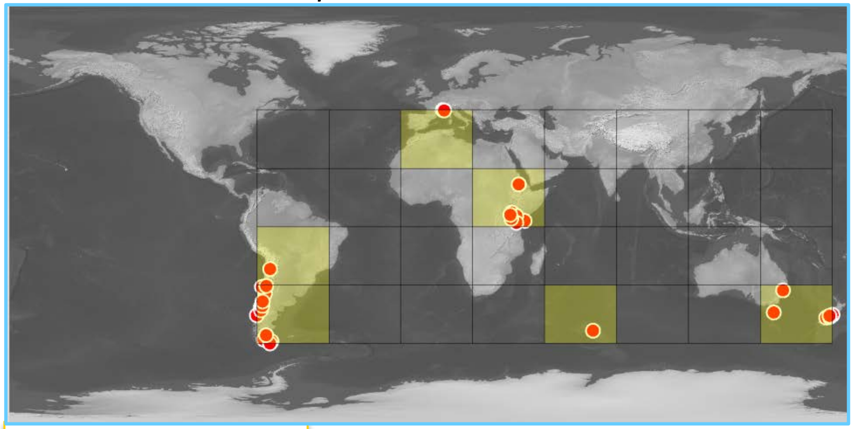
UFS_{32}

× = reference point

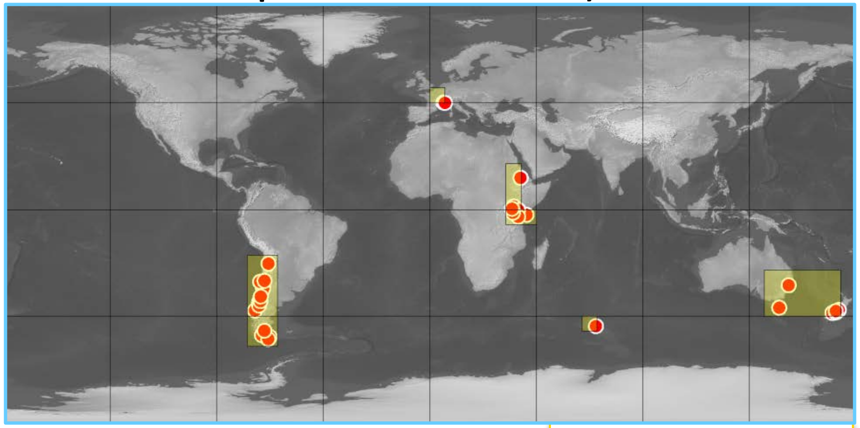
● = resource data point

Hybrid approaches

- combining properties of the other two aforementioned approaches, two subcategories:
  - shape 1st, partitioning 2nd || partitioning 1st, shape 2nd
- examples:
  - MBRGrid_r (shape 1st)
  - GridMBR_r^b (partitioning 1st ; b = # bits used for 1 quantized value)



MBRGrid_4



GridMBR_4^3

● = resource data point

2. Categories of Resource Description Techniques

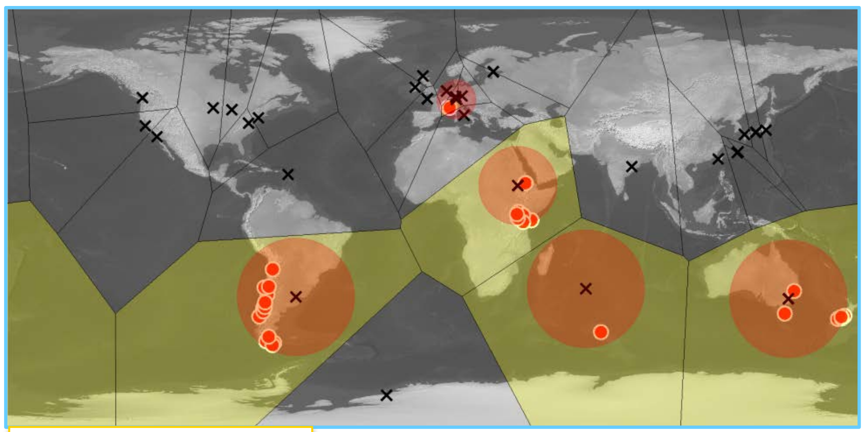
- in general: best hybrid approaches outperform best geometric approaches as well as best space partitioning approaches!

  → see: [3]: Using Hybrid Techniques for Resource Description and Selection in the Context of Distributed Geographic Information Retrieval.

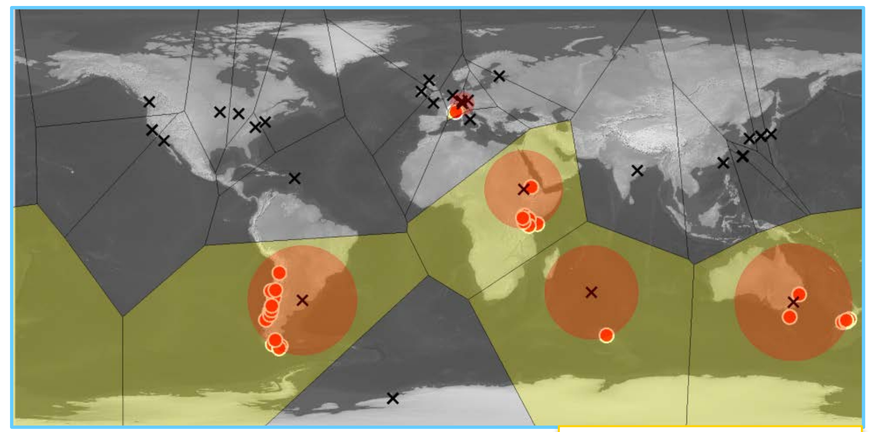
- contribution of this work: four novel hybrid approaches, all using quantization in order to address the space efficiency criterion (trade-off with selectivity!)

❶ DFS_n^b (Distance fine-grained summaries)

- combines Voronoi-like space partitioning with distance information
- parameters n: # reference points; b: # bits used for 1 distance value
  see slide 20
- summary: bit vector
  - first 32 bits: peer's threshold value (base distance for quantization)
  - binary information about cell occupancy (0 = empty, 1 = occupied)
  - after '1': *b* bits with quantized distance value (max. distance)



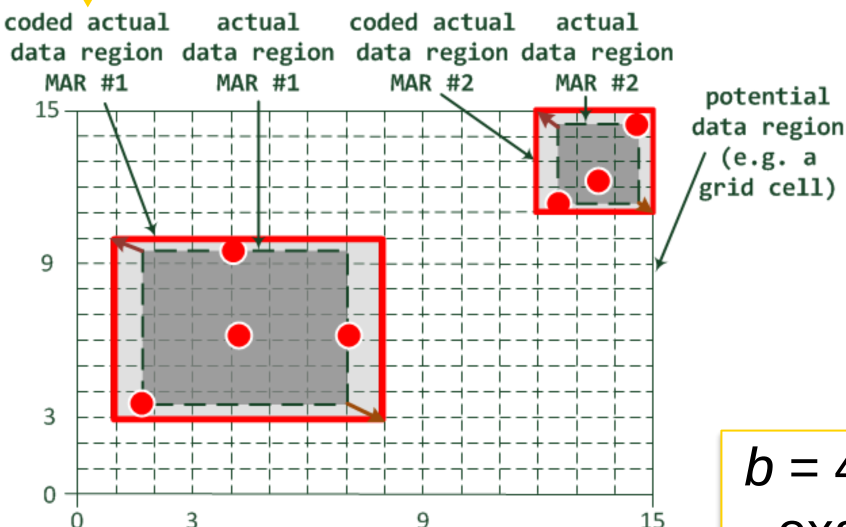
DFS_{32}^{2}



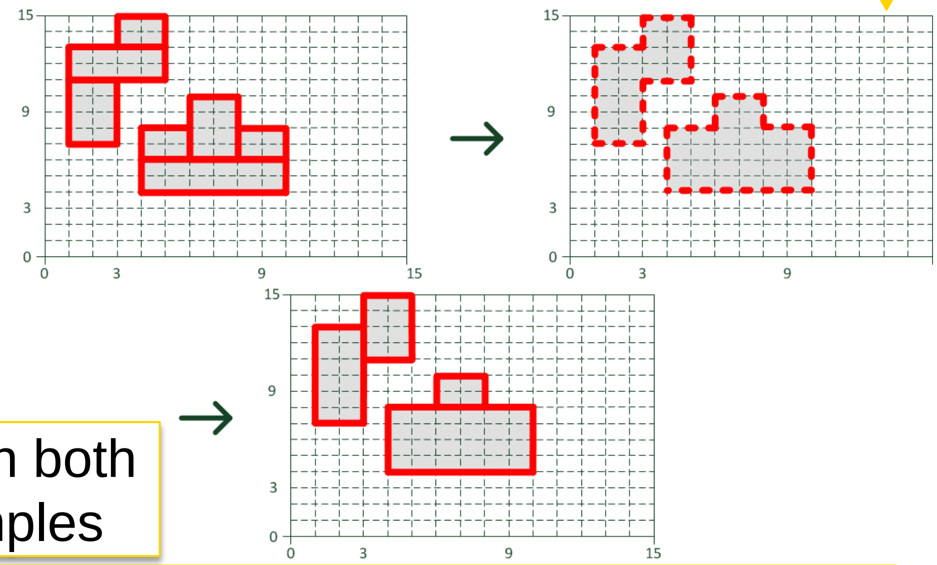
DFS_{32}^{4}

❷ GridMAR $_r^{b,k}$

- encodes up to *k* interior MARs for each occupied cell of a uniform grid
- reducing storage space:
  - encoding quantized MARs ($4 \cdot b$ bits for each rectangle)
  - minimizing the number of MARs → polygon decomposition problem

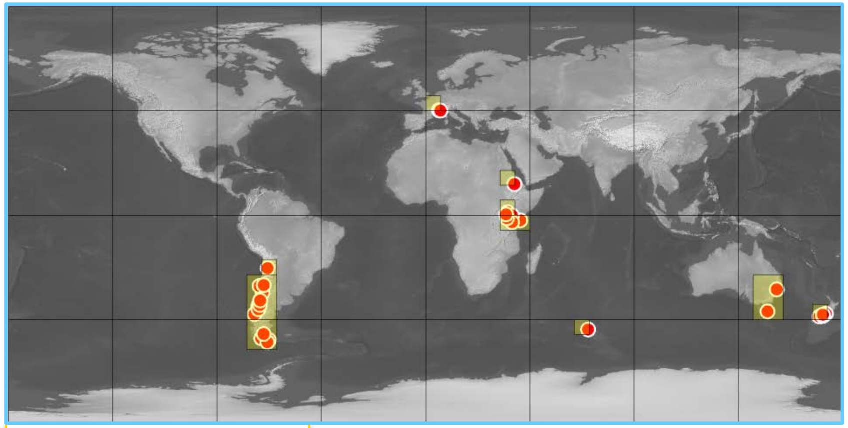


b = 4 in both examples

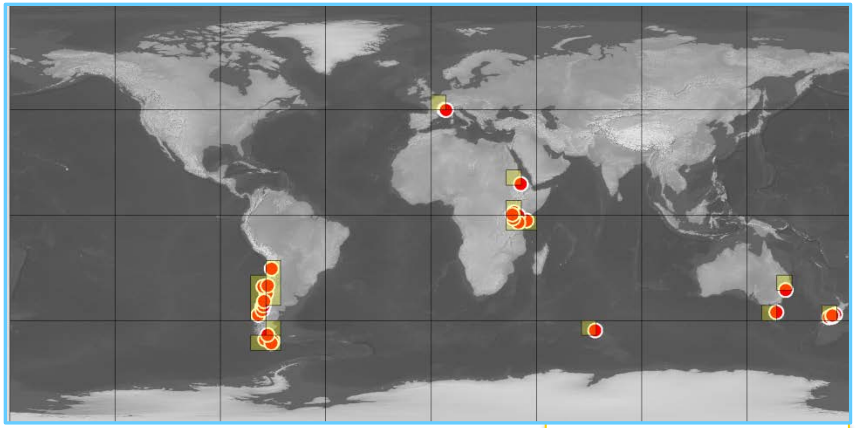


❷ GridMAR $_r^{b,k}$

- parameters: r → resolution of global grid (r rows, $2 \cdot r$ columns)
  b → # bits for encoding one (of four) MAR value(s)
  k → max. # of rectangles per cell
- summary: bit vector
  - each cell represented by 1 bit indicating cell occupancy
  - after ‘1’: # of MARs ($\lceil \log_2 k \rceil$ bits), $4 \cdot b$ bits for each MAR



GridMAR$_4^{3,2}$



GridMAR$_4^{3,3}$

❸ K-D-MAR$_n^{b,k}$

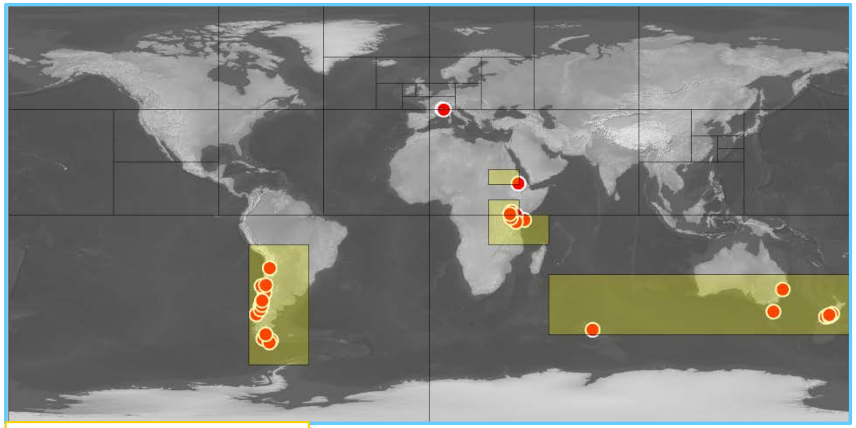
- k-d-tree like space partitioning (learned from training data), up to k cell interior MARs

see slide 20

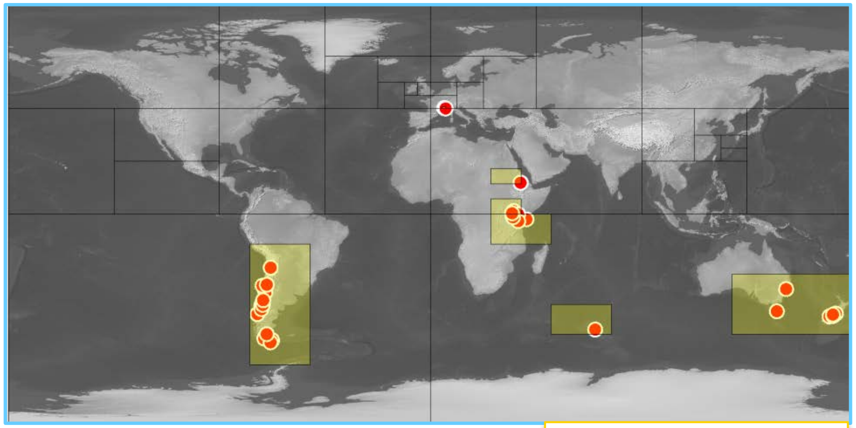
- parameters: n → # of subspaces/cells
  b → # of bits for encoding one (of four) MBR value(s)
  k → max. # of rectangles per cell;

MAR minimization is conducted!

- summary: same as GridMAR$_r^{b,k}$



K-D-MAR$_4^{3,2}$



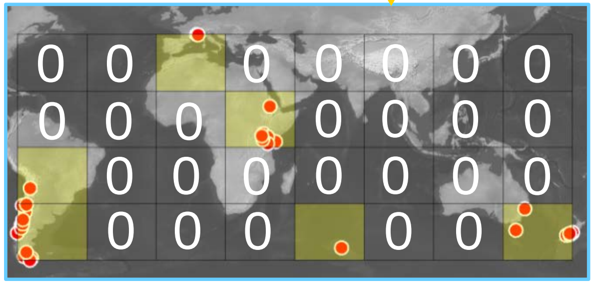
K-D-MAR$_4^{3,3}$

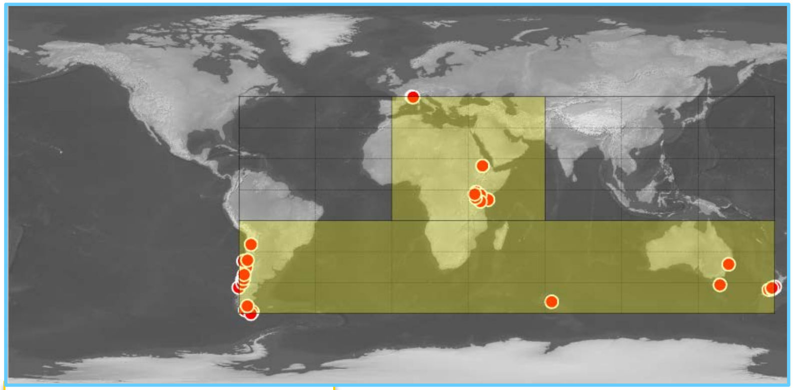
3. Novel Hybrid Resource Description Techniques

❹ MBR-MAR^{b,k}

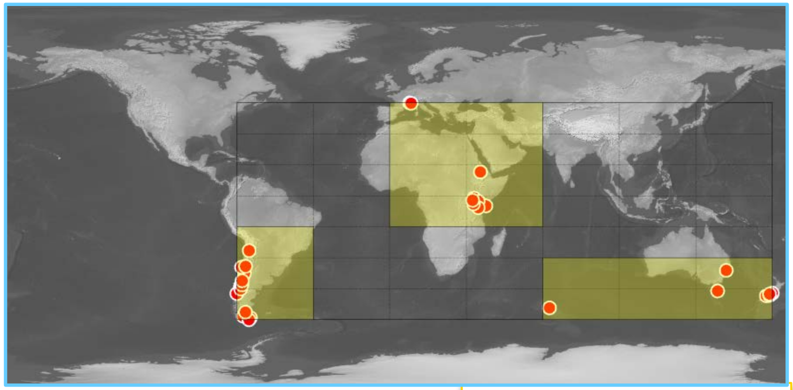
MAR minimization is conducted!

- takes MBR as base, refinement with MBR-interior MARs
- advancement over aforementioned MBRGrid$_r$
- parameters: b → # of bits for encoding one (of four) MBR value(s)
  k → max. # of rectangles per cell
- summary: bit vector
  - first 4 · 32 bits: MBR extents (float precision)
  - 4 · b bits for each MAR; # of MARs not required



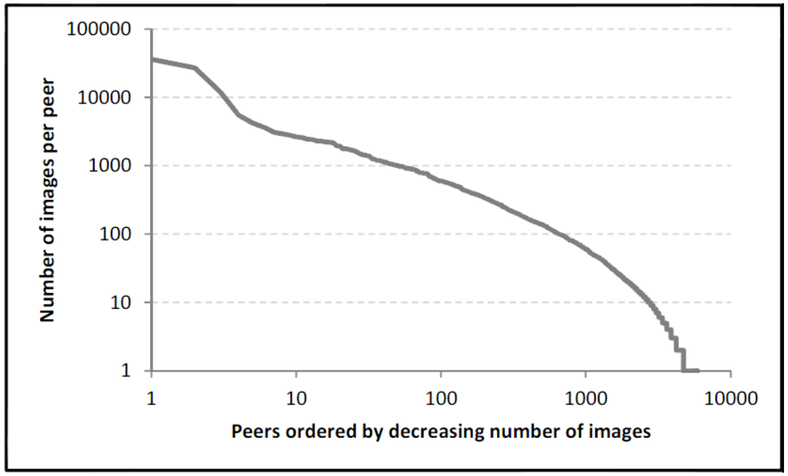


MBR-MAR^{3,2}



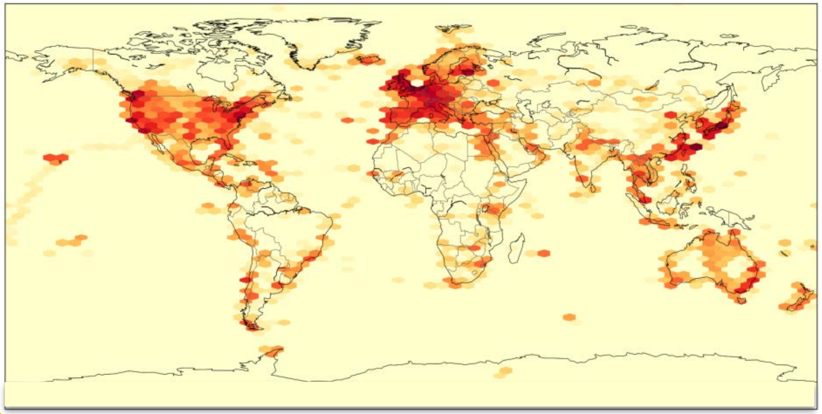
MBR-MAR^{3,3}

4. Experimental Setting and Evaluation



- geo-tagged images from flickr
  - 406.450 different geo-tagged images
  - 5.951 different users (unique user IDs)
  - 1 Flickr user corresponds to 1 peer

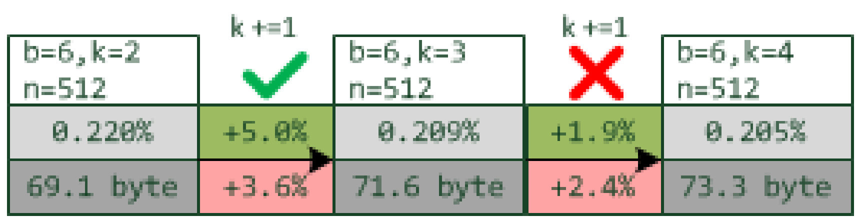
  → 5.951 peers for evaluation
- k-nearest-neighbor queries: 200 point queries from data collection (random data points chosen from random peers); $k = 50$
- main performance measures:
  - fraction of contacted peers [in %]
  - avg. summary sizes [in byte] (gzip compression if beneficial)



4. Experimental Setting and Evaluation

- data sources for reference points/training data:
  - randomly chosen from data collection ("internal data")
  - Geonames gazetteer in compliance with GDP per country statistics from WorldMapper → approximated data distribution ("external data")

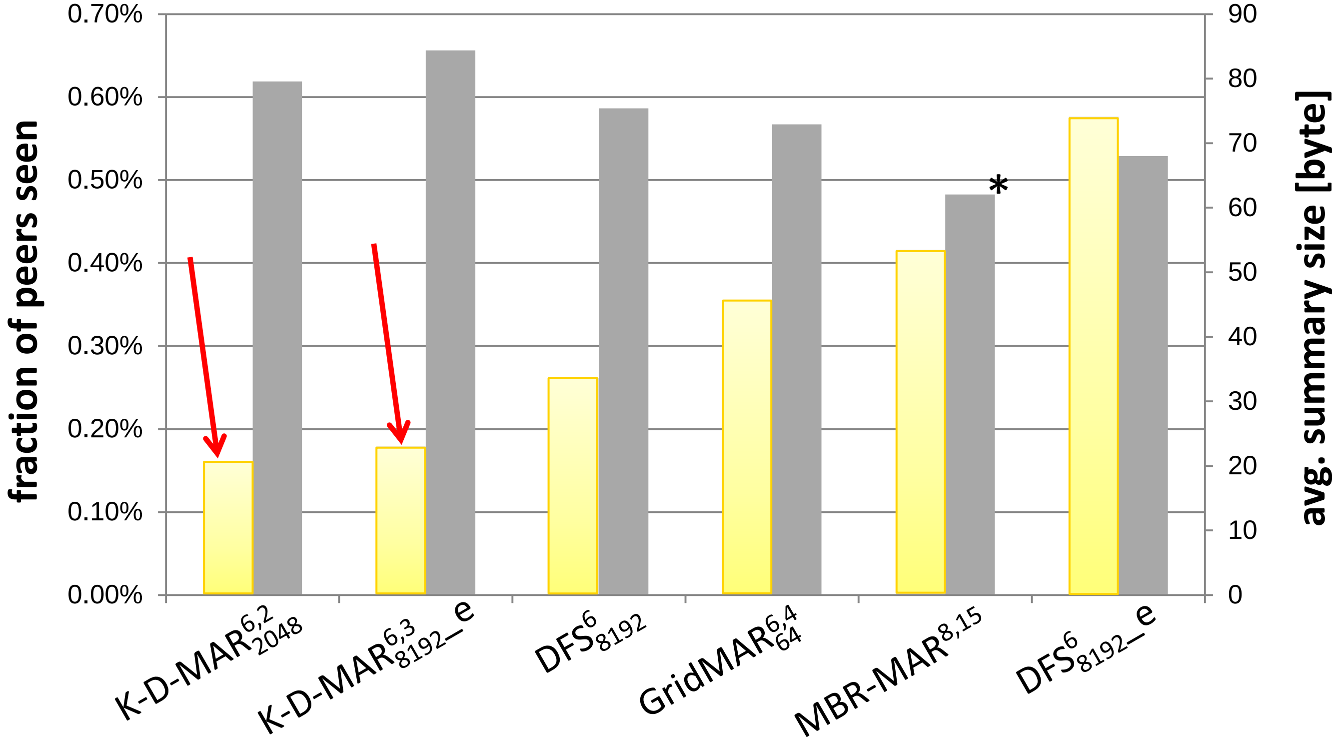
    denoted with an attached "_e", e.g. "DFS_n^b_e"
  - http://www.geonames.org http://www.worldmapper.org
- parameters have been varied in a certain range for each approach → details in the paper
- comparison of respective "best" parameterizations in the following
  - as long as percentage selectivity gains are higher than percentage summary size growth → beneficial to raise parameters
  - example:



Overview

Direct transmission of data points:
0.138% (fraction of peers)
265.8 byte (summ. size)

*
no gzip compression

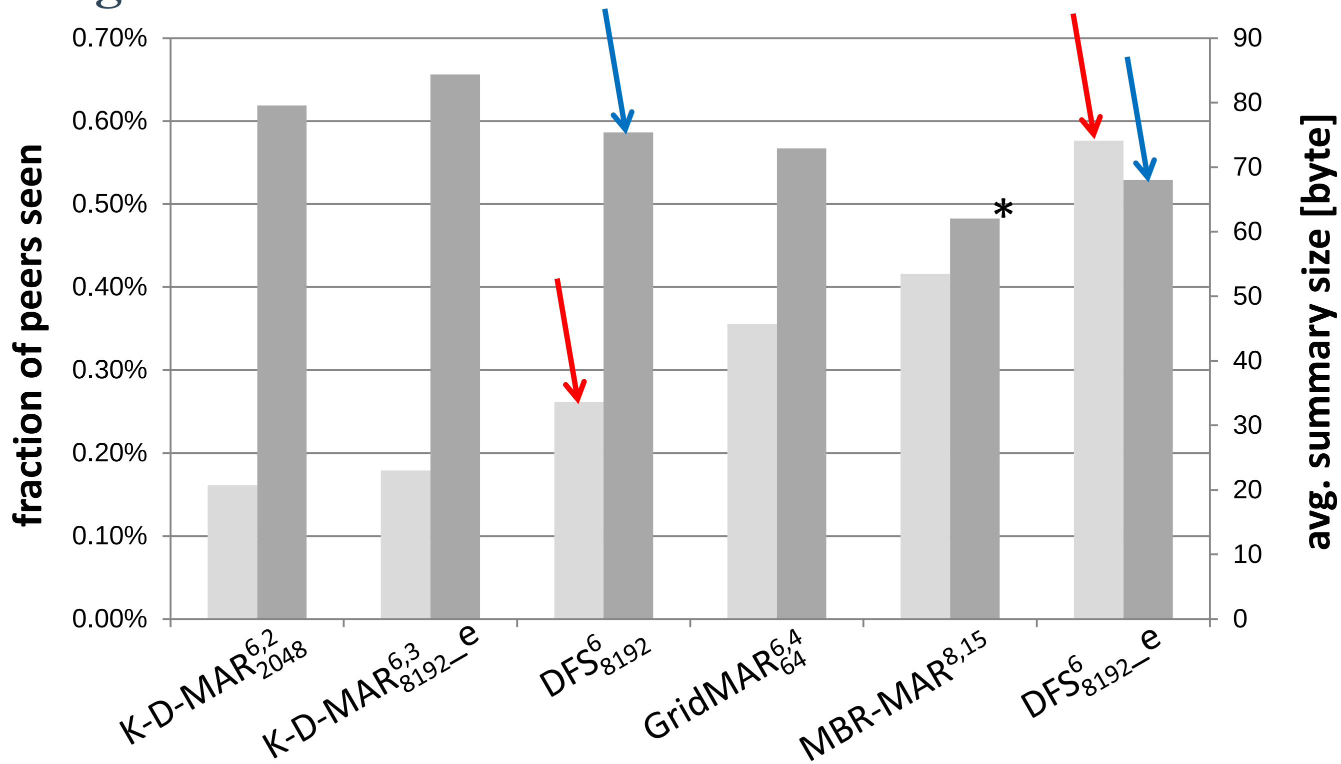


- K-D-MAR$^{b,k}_{n}$ offers best selectivity with still moderate avg. summary sizes (both with internal and external data); more selective and smaller summaries with internal data (different parameterization)
- in general: same parameterization → internal data: more selectivity, bigger summaries (see DFS$^{6}_{8192}$ vs DFS$^{6}_{8192}$ _e)

Overview

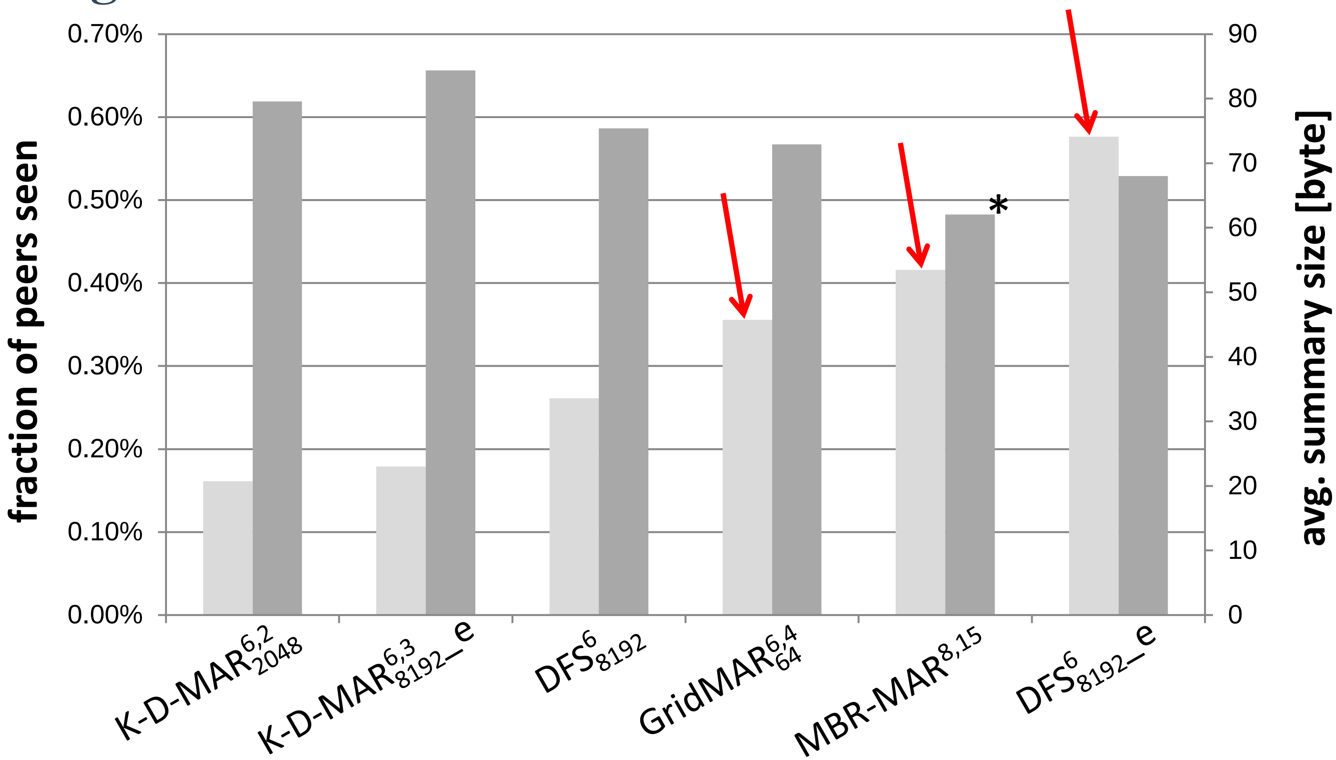
Direct transmission of data points:
0.138% (fraction of peers)
265.8 byte (summ. size)

*
no gzip compression



- K-D-MAR$_n^{b,k}$ offers best selectivity with still moderate avg. summary sizes (both with internal and external data); more selective and smaller summaries with internal data (different parameterization)
- in general: same parameterization → internal data: more selectivity, bigger summaries (see DFS$^6_{8192}$ vs DFS$^6_{8192}$_e)

Overview



Direct transmission of data points:
0.138% (fraction of peers)
265.8 byte (summ. size)

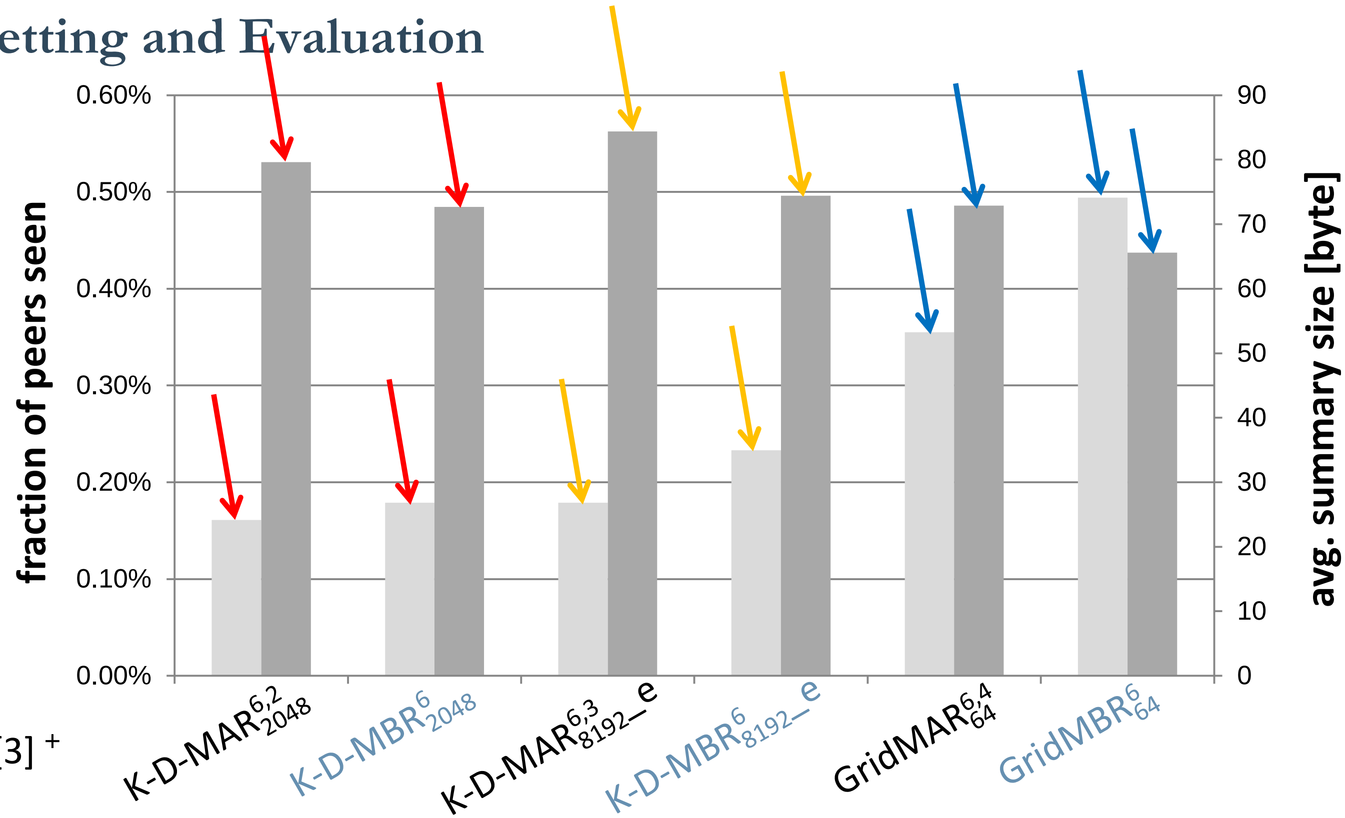
*
no gzip compression

- GridMAR$^{6,4}_{64}$ and MBR-MAR^{8,15} (no specific space partitioning adjustments): better selectivity than DFS$^{6}_{8192}$ _e

  → quantized MARs pay off
- both come close to DFS$^{6}_{8192}$ (internal data) → overhead justifiable?

Overview

[3]: Using Hybrid Techniques for Resource Description and Selection in the Context of Distributed Geographic Information Retrieval.



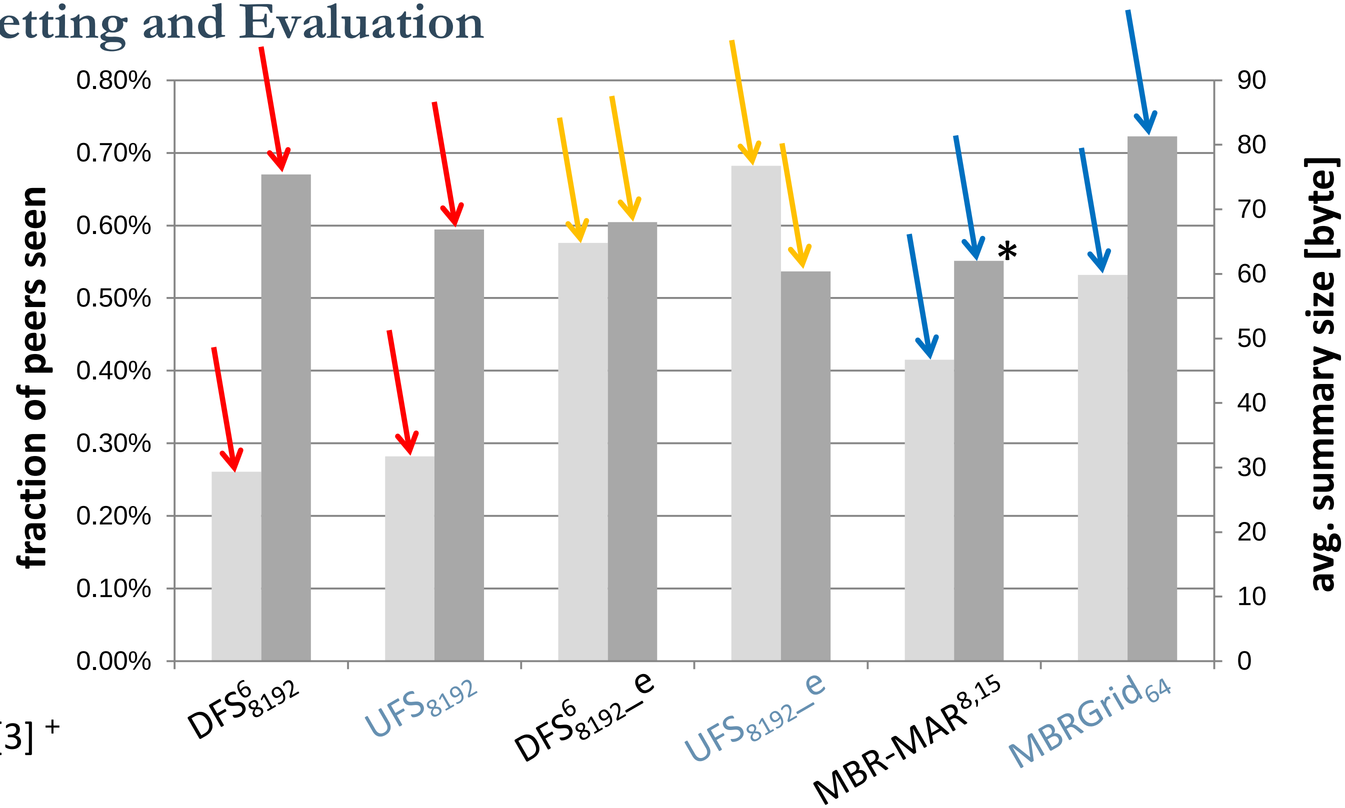
*
no gzip compression

ABC$^{y}_{x}$
old technique, evaluated in [3] +

- comparison to old techniques evaluated in [3]+:
  - relying on only 1 quantized, cell-interior rectangle (K-D-MBR$_n^b$, GridMBR$_r^b$)
  - the less accurate the basic approach, the more beneficial the computation of several rectangles

→ K-D-MAR$_n^{b,k}$ slight improvement, GridMAR$_r^{b,k}$ notable improvement

Overview



+ [3]: Using Hybrid Techniques for Resource Description and Selection in the Context of Distributed Geographic Information Retrieval.

* no gzip compression

ABC^{y}_{x} old technique, evaluated in [3] +

- comparison to old techniques evaluated in [3]+:
  - DFS$^{6}_{8192}$: not worthwhile compared to just encoding binary information (UFS$_n$)
  - MBR-MAR^{8,15} : notable improvement in selectivity & avg summary sizes

Quick Summary

- K-D-MAR$_{n}^{b,k}$: adjusted space partitioning possible, max. selectivity desired
- GridMAR$_{r}^{b,k}$ MBR-MAR^{b,k} : no adjusted space partitioning possible, good selectivity desired
- DFS$_{n}^{b}$: better alternatives existent

5. Future Work

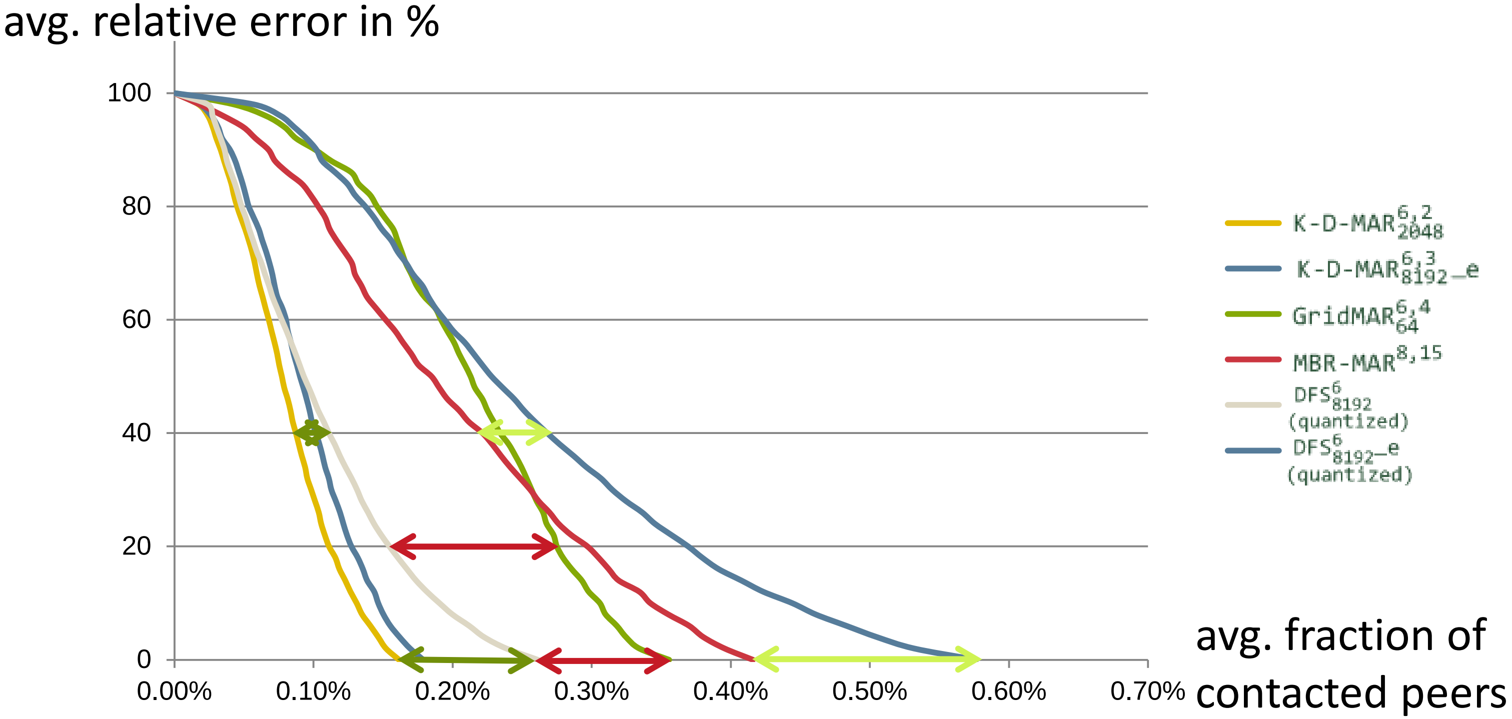
- new data collection*:* 70,790,823 geo-tagged tweets from 4,780,464 Twitter users
- utilization of quadtree encodings (e.g. cell-interior quadtrees, i.e. *GridQuadTree* or *K-D-QuadTree*)
- evaluation of techniques in different application fields (centralized index structures)

Thank you very much for your attention!

Appendix

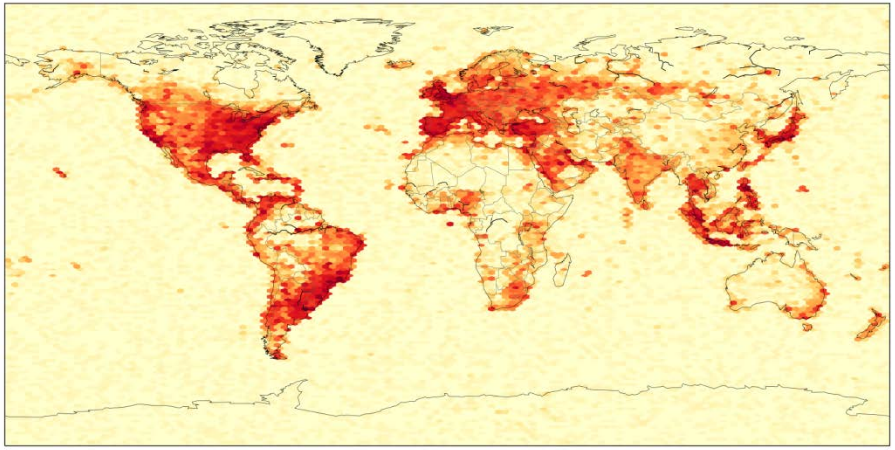
Relative error (approximate search)

- in general: selectivity resembles exact/precise search performance
- DFS_n^b/_e falls off for the last 20%-40%
- exception: $\mathrm{GridMAR}_r^{b,k}$



New data collection – Twitter

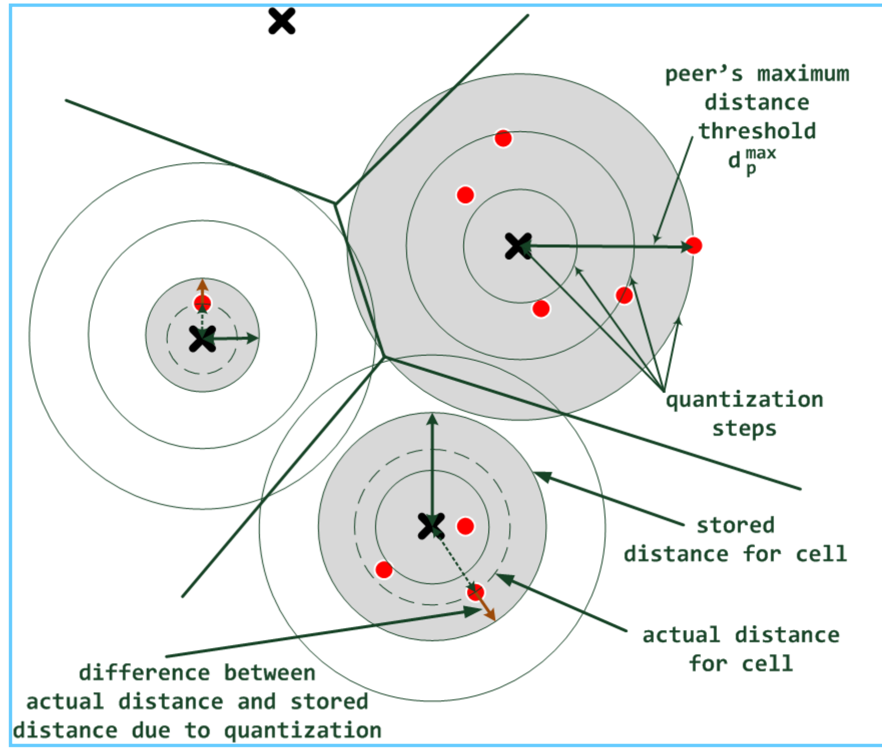
- 70,790,823 geo-tagged tweets from 4,780,464 users
- gathered from 25.08.2014 to 01.09.2014
- 3354 three-minute-intervals, 67 languages, 235 countries
- 91,818,675 different terms
- Global distribution:



- Open questions: Scenario (i.e. how to assign data points to peers)? Include all data || filter by e.g. language?

Quantization for DFS_n^b

- global threshold for quantization not reasonable
- for each cell: maximum distance between the cell's reference point and any of its associated data points is calculated
- biggest value over all these distances: individual threshold d_p^{max}
  → is encoded into summary (first 32 bits)



Quantization for DFS_n^b, resulting numbers

- quantized DFS_n^b vs. non-quantized DFS_n vs. old UFS_n
  - DFS_n^b encodes distance information for each occupied cell
  - UFS_n just encodes binary occupation information

			contacted peers change	summary size change
DFS_{8192} (non-quant.)	0.261%	75.4	-	-
DFS_{8192}^{6} (quant.)	0.256%	93.4	-1.92%	+23.87%
UFS_{8192}	0.282%	66.9	+8.05%	-11.27%
DFS_{8192}_e (non-quant.)	0.576%	68	-	-
DFS_{8192}^{6}_e (quant.)	0.556%	77.3	-3.47%	+13.68%
UFS_{8192}_e	0.682%	60.4	+10.60%	-11.18%

New techniques vs. old techniques, numbers

1/2

- K-D-MAR$_n^{b,k}$ vs. K-D-MBR$_n^b$
  - K-D-MAR$_n^{b,k}$: up to *k* cell-interior rectangles
  - K-D-MBR$_n^b$: exactly one cell-interior rectangle

  for each occupied cell

			contacted peers change	summary size change
K-D-MAR$_{2048}^{6,2}$	0.161%	79.6	-	-
K-D-MBR$_{2048}^{6}$ (old)	0.179%	72.7	+11.18%	-8.67%
K-D-MAR$_{8192}^{6,3}$ _e	0.179%	84.4	-	-
K-D-MBR$_{8192}^{6}$_e(old)	0.233%	74.4	+30.17%	-11.85%

New techniques vs. old techniques, numbers 2/2

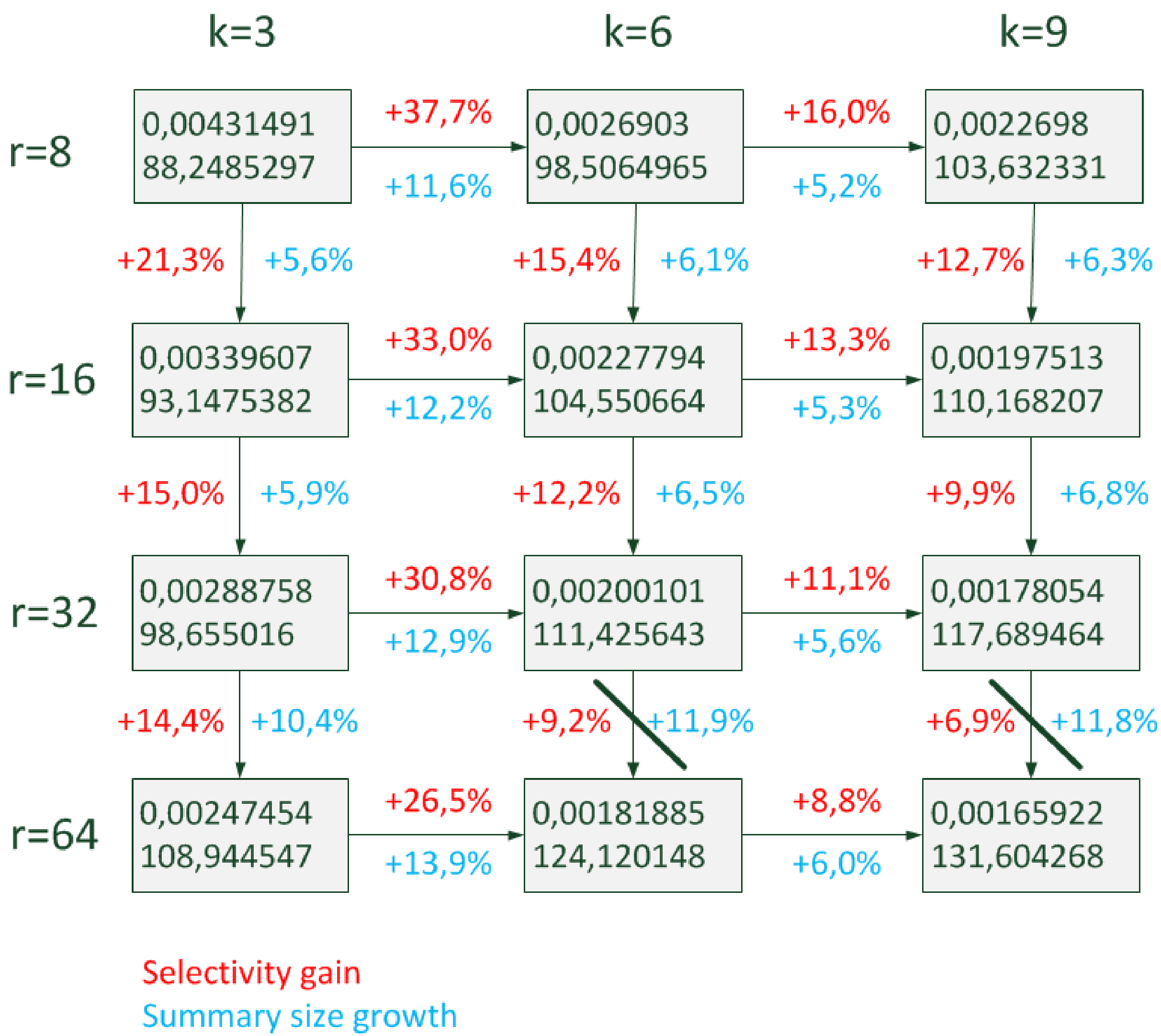
- GridMAR$_r^{b,k}$ vs. GridMBR$_r^b$
  - GridMAR$_n^{b,k}$: up to *k* cell-interior rectangles
  - GridMBR$_n^b$: exactly one cell-interior rectangle for each occupied cell
- MBR-MAR^{b,k} vs. MBRGrid$_r$
  - MBR-MAR^{b,k}: MBR-interior set of up to *k* rectangles
  - MBRGrid$_r$: MBR-interior grid with resolution *r*

			contacted peers change	summary size change
GridMAR$_{64}^{6,4}$	0.355%	72.9	-	-
GridMBR$_{64}^{6}$ (old)	0.494%	65.6	+39.15%	-10.01%
MBR-MAR^{8,15}	0.415%	62.0	-	-
MBRGrid$_{64}$ (old)	0.532%	81.3	+21.99%	+31.12%

Parameter variation

- DFS_{n}^{b}
  - ***n***: 512, 2048, 8192; ***b*** = 2, 4, 6
- $\text{GridMAR}_{r}^{b,k}$
  - ***r***: 16, 32, 64; ***b*** = 2, 4, 6; ***k*** = 2, 3, 4, 6
- $\text{K-D-MAR}_{n}^{b,k}$
  - ***n*** = 512, 2048, 8192; ***b*** = 2, 4, 6; ***k*** = 2, 3, 4, 6
- $\text{MBR-MAR}^{b,k}$
  - ***b*** = 2, 4, 6, 8; ***k*** = 3, 6, 9, 12, 15

Finding best parameterization



Ranking for rectangle-based descriptions

- approach describe several rectangular areas containing data points!
- 1. ∀ peer:
  - 1.1 ∀ rectangle:
    - calculate minimum distance to query and area covered
    - store information in R-Entry and insert R-Entry into a queue
  - 1.2 Sort queue by a) minimum distance and b) minimum area
- 2. Ranking between Peer A and Peer B
  - 2.1 i=1;
  - 2.2 Choose i-th R-Entry for Peer A ($RE_a i$) and Peer B ($RE_b i$)
    - if [min. dist. $RE_a i$ < min. dist. $RE_b i$]
      - → $P_a \succ P_b$ rank Peer A higher than Peer B
    - else if [area $RE_a i$ < area $RE_b i$]
      - → $P_a \succ P_b$ rank Peer A higher than Peer B
    - else
      - i++; GOTO 2.1;

min. dist. is equal!

Ranking for DFS_n^b

- sort reference points c_j ($j \in \{0;n-1\}$) in ascending order w.r.t. the query location

  → sorted list L; 1st element of L = query cluster
- ranking between Peer A and Peer B
  - 1. i=1;
  - 2. choose i-th element of L
    - if peer P_a administers documents in this cluster while P_b does not
      - → $P_a \succ P_b$ rank Peer P_a higher than Peer P_b
    - else
      - i++; GOTO 2.;

⚠ distance information used for pruning only since binary ranker is superior to ranker utilizing max. balls (comparable to ranker for rectangle-based descriptions; using balls instead of rectangles)!

kNN algorithm to determine top-50 data points

- Series of range queries with dynamic query range

1. Initially rank all peers (Set S) based on their resource description and the query point

2. Determine the 50 nearest images to the query point from the 10 best ranked peers; remove these peers from S

3. While S is not empty, do

- Set query range to the distance of the currently 50th nearest neighbor
- Remove all peers that can not have relevant data points from S
- Analyze images from the 10 best-ranked peers from remainder of set S
- Remove these peers from S

→ when S is empty: top 50 have been determined

MAR # minimization (1)

$\text{K-D-MAR}_n^{b,k}$_e

512	maxMARs	2	3	4	6
bits	2	1.315966	1.558697	1.722596	1.884406
	4	1.14865	1.353516	1.560239	1.922503
	6	1.12677	1.288717	1.472187	1.790822

2048	maxMARs	2	3	4	6
bits	2	1.450705	1.827869	2.059217	2.289281
	4	1.235556	1.488667	1.757412	2.20744
	6	1.150625	1.398605	1.634789	2.036511

8192	maxMARs	2	3	4	6
bits	2	1.641861	2.082329	2.345985	2.614248
	4	1.341598	1.703566	2.077178	2.689399
	6	1.256509	1.55632	1.827026	2.330333

$\text{K-D-MAR}_n^{b,k}$

512	maxMARs	2	3	4	6
bits	2	1.441413	1.799591	2.010840724	2.25946
	4	1.181172	1.445213	1.700238	2.190078
	6	1.110882	1.296807	1.491037	1.827071

2048	maxMARs	2	3	4	6
bits	2	1.737922	2.301289	2.687026	3.096139
	4	1.317148	1.707319	2.083832	2.730999
	6	1.159325	1.384177	1.636404	2.083129

8192	maxMARs	2	3	4	6
bits	2	2.219819	3.127232	3.701462	4.237606
	4	1.479281	2.032496	2.528858	3.414974
	6	1.245918	1.547946	1.863541	2.417923

MAR # minimization (2)

MBR-MAR^{b,k}

	maxMARs	3	6	9	12	15
bits	2	1.582734	1.429084	1.582734	1.675515	1.721776
	4	1	1.179163	1.43907	1.701452	1.935171
	6	1	1.146597	1.363556	1.574673	1.807345
	8	1	1.147239	1.347958	1.515963	1.698765

16	maxMARs	2	3	4	6
bits	2	1.16091954	1.27756654	1.35897436	1.4312
	4	1.09259259	1.20716783	1.33957034	1.58362573
	6	1.04365079	1.13211382	1.24471959	1.47576602
32	maxMARs	**2**	**3**	**4**	**6**
bits	2	1.23659674	1.40556492	1.49205148	1.6172927
	4	1.08726753	1.23832528	1.40669241	1.70366379
	6	1.0918197	1.18834081	1.33779264	1.58469388
64	maxMARs	**2**	**3**	**4**	**6**
bits	2	1.33634538	1.57446809	1.73214286	1.90194805
	4	1.14321608	1.31821454	1.50517241	1.87122338
	6	1.12816456	1.26149915	1.4344473	1.7388724

GridMAR$_r^{b,k}$

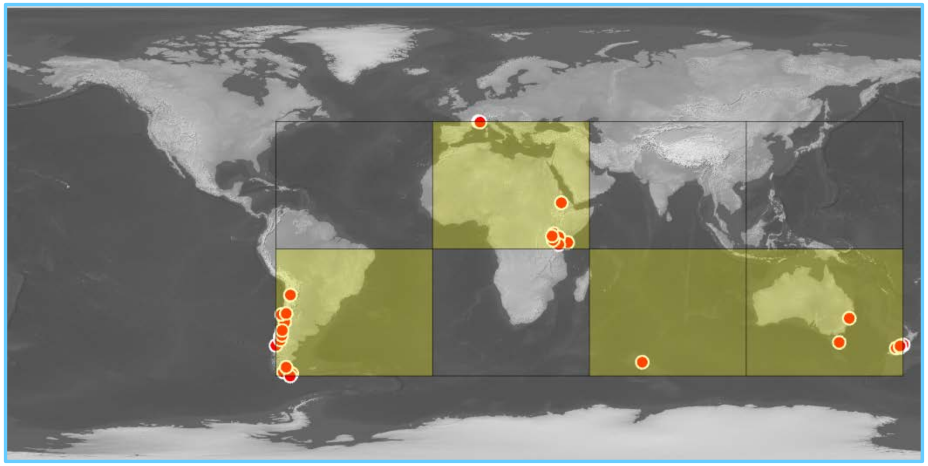
Comparison of techniques

technique	selectivity	avg. summary sizes
MBR	4.150%	43.0*
ECH	2.070%	61.4*
kMeans++_9	0.440%	88.7*
RecMAR_9	0.380%	69.3*
Grid_{8192}	2.320%	58.1
K-D bin. $_{8192}$	0.630%	68.4
K-D bin. $_{8192}$_e	0.790%	63.4
UFS_{8192}	0.280%	66.9
UFS_{8192}_e	0.682%	60.4
MBRGrid_{64}	0.532%	81.3
KMARGrid_9^{32}	0.178%	117.7
GridMBR_{64}^6	0.494%	65.6
K-D-MBR_{2048}^6	0.179%	72.7
K-D-MBR_{8192}^6_e	0.233%	74.4
DFS_{8192}^6	0.261%	75.4
DFS_{8192}^6_e	0.576%	68.0
DFS_{8192} no quant	0.256%	93.4
DFS_{8192} no quant_e	0.556%	77.3
$\text{GridMAR}_{64}^{6,4}$	0.355%	72.9
$\text{K-D-MAR}_{2048}^{6,2}$	0.161%	79.6
$\text{K-D-MAR}_{8192}^{6,3}$_e	0.179%	84.4
$\text{MBR-MAR}^{8,15}$	0.415%	62.0*

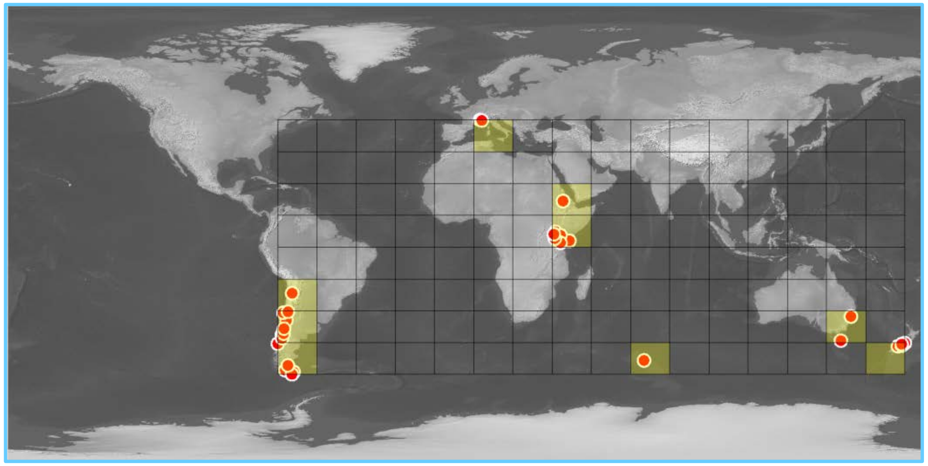
Old Hybrid Resource Description Techniques (geom. 1st, space part. 2nd)

❶ MBRGrid $_r$

- combines a MBR with an internal grid
- parameter r: resolution of the internal grid (r rows, $2r$ columns)
- summary: bit vector
  - first $4 \cdot 32$ bit values: MBR boundaries
  - remainder: $r \cdot 2r$ bits indicating cell occupancy (1 bit for 1 cell)



$r = 2$

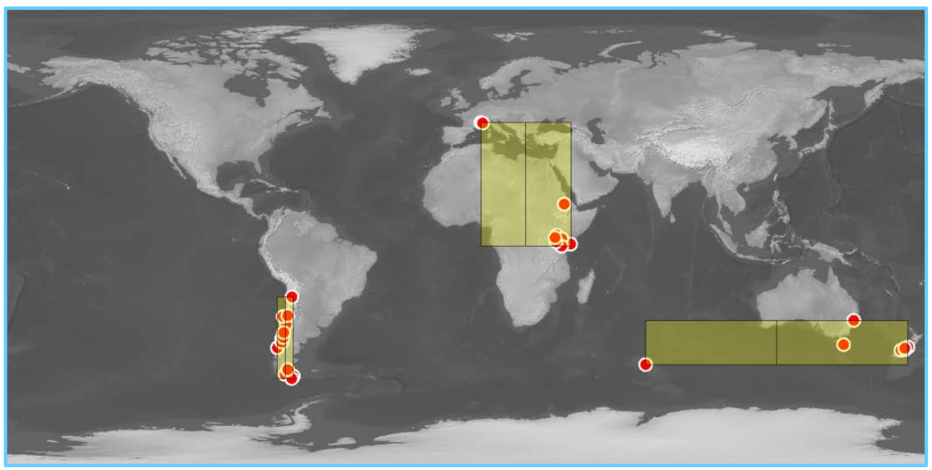


$r = 8$

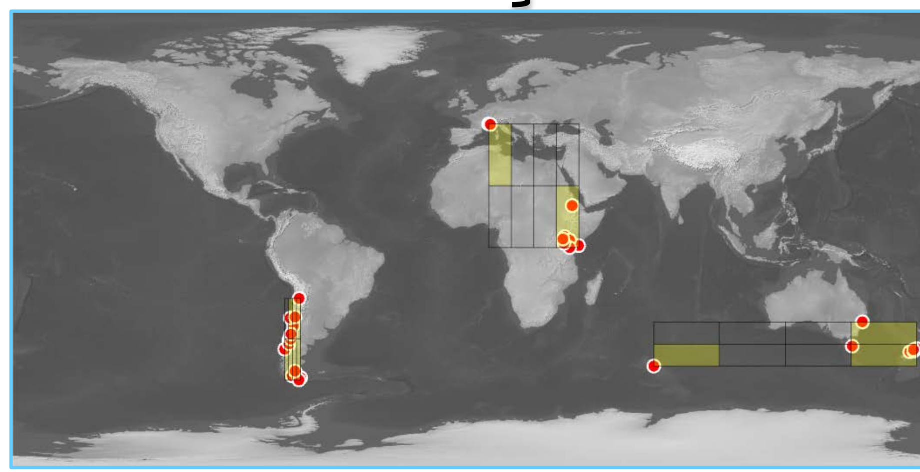
❷ KMARGrid_k^r

- combines RecMAR_k with internal grids for each MAR
- parameters: r → resolution of MAR internal grids (r rows, $2r$ columns)
  k → max. number of MARs
- summary: bit vector
  - first 4 · 32 bit values: MAR boundaries
  - remainder: $r \cdot 2r$ bits indicating cell occupancy (1 bit for 1 cell)

up to k times



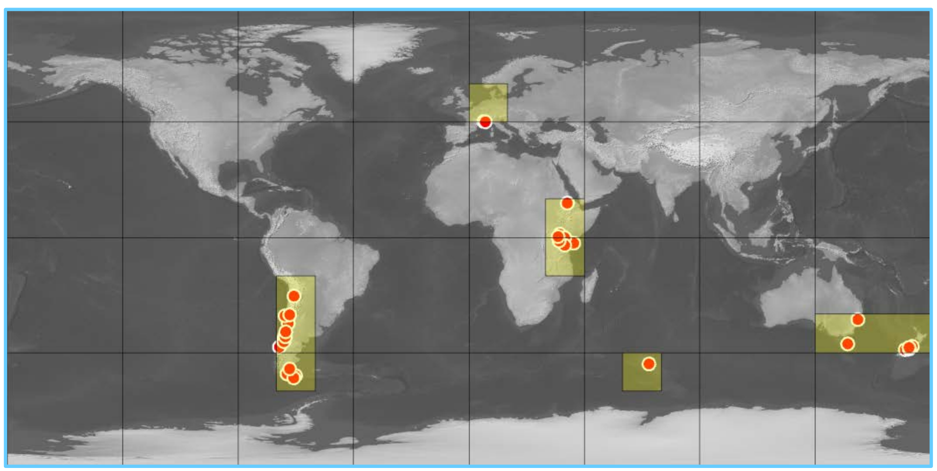
$k = 3, r = 1$



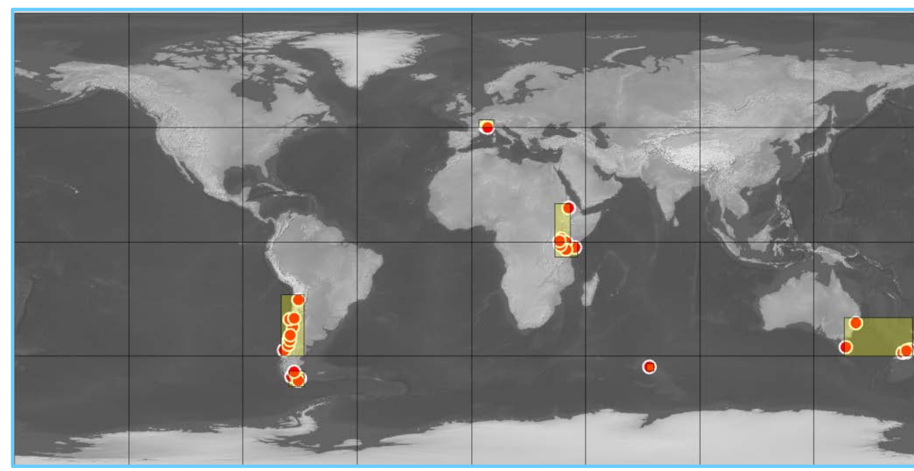
$k = 3, r = 2$

❸ GridMBR $_r^b$

- parameters: r → resolution of global grid (r rows, $2r$ columns)
  b → number of bits for encoding one (of four) MBR value(s)
- summary: bit vector
  - each cell represented by 1 bit indicating cell occupancy
  - after an '1': $4 \cdot b$ bits for cell interior MBR



$r = 4, b = 2$

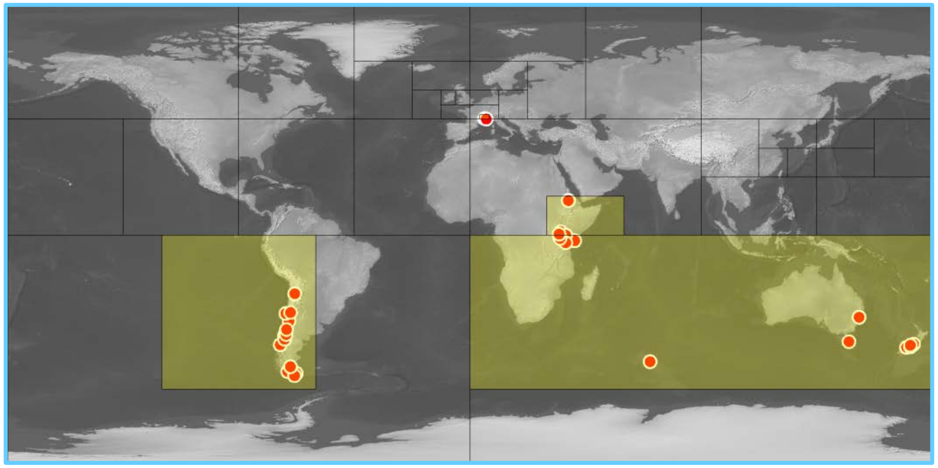


$r = 4, b = 4$

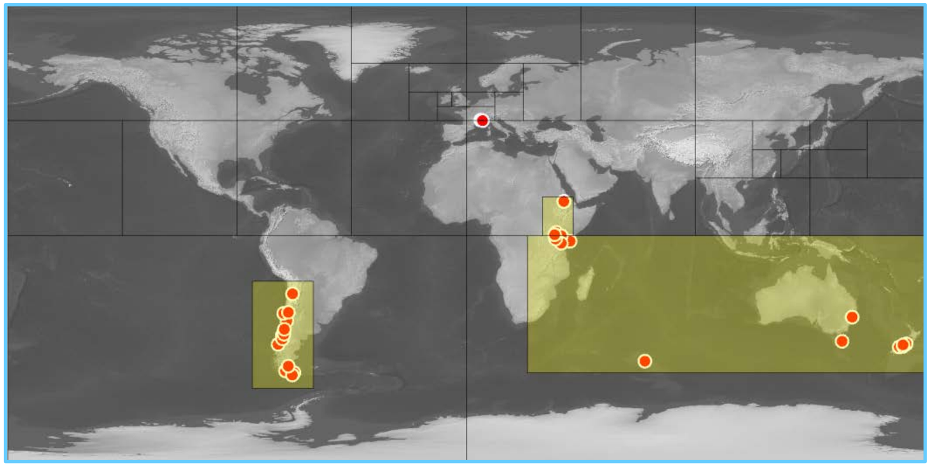
Old Hybrid Resource Description Techniques (space part. 1st, geom. 2nd)

❹ K-D-MBR$_n^b$

- k-d-tree like space partitioning (learned from training data), cell interior MBRs
- training data: from data collection [→ Sec.4] or from external source
- parameters: r → resolution of MAR internal grids (r rows, $2r$ columns)
  b → number of bits for encoding one (of four) MBR value(s)
- summary: same as GridMBR$_r^b$



$n = 32, b = 2$



$n = 32, b = 3$

Literature 1/3

[1] F. Cuenca-Acuna, C. Peery, R. P. Martin, T. D. Nguyen. PlanetP: Using gossiping to build content addressable peer-to-peer information sharing communities. IEEE Intl. Symp. on High Performance Distributed Computing 236-246, 2003

[2] S. Kufer, D. Blank, A. Henrich. Techniken der Ressourcenbeschreibung und -auswahl für das geographische Information Retrieval. Proceedings of the IR Workshop at LWA 2012 1-8, September 2012.

[3] S. Kufer, D. Blank, A. Henrich. Using Hybrid Techniques for Resource Description and Selection in the Context of Distributed Geographic Information Retrieval. Advances in Spatial and Temporal Databases. Lecture Notes in Computer Science 8098:330-347, August 2013.

[4] D. Blank, A. Henrich. Describing and Selecting Collections of Georeferenced Media Items in Peer-to-Peer Information Retrieval Systems. Diaz, Laura ; Granell, Carlos ; Huerta, Joaquin (eds.): Discovery of Geospatial Resources: Methodologies,Technologies, and Emergent Applications 1-20, 2012.

[5] B. Becker, P. G. Franciosa, S. Gschwind, T. Ohler, G. Thiemt, P. Widmayer. An Optimal Algorithm for Approximating a Set of Rectangles by Two Minimum Area Rectangles. Intl. Workshop on Computational Geometry. LNCS 553:22-29, 1991.

[6] D. Blank, A. Henrich. Resource Description and Selection for Range Query Processing in General Metric Spaces. BTW 2013: 15. GI-Fachtagung Datenbanksysteme für Business, Technologie und Web. GI-Fachbereich "Datenbanken und Informationssysteme" 93-112, 2013.

Literature 2/3

[7] A. Henrich. The LSDh-tree: An Access Structure for Feature Vectors. Proc. of the 14th intl. conf. on Data Engineering 262-369, 1998.

[8] B. Seeger, H.-P. Kriegel. The buddy-tree: an efficient and robust access method for spatial data base systems. Proc. of th 13th intl. conf. on VLDB 590{601, 1990.

[9] J. M. Keil Polygon Decomposition. J.-R. Sack, J. Urrutia (Eds.), Handbook of Computational Geometry 491-518, 1999.

[10] D. S. Franzblau, D. J. Kleitman. An algorithm for covering polygons with rectangles. Inf. Control 63(3):164-189, 1986

[11] V. Soltan, A. Gorpinevich. Minimum dissection of rectilinear polygon with arbitrary holes into rectangles. SCG '92 Proceedings of the eighth annual symposium on Computational geometry 296-302, 1992.

[12] J. L. Bentley. Multidimensional Binary Search Trees Used for Associative Searching. Commun. ACM 18(9):509-517, 1975.

[13] V. Murdock. Your Mileage May Vary: On the Limits of Social Media. SIGSPATIAL Special 3(2):62-66, July 2011.

[14] H. Samet. Foundations of Multidimensional and Metric Data Structures. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA 2005

Literature 3/3

[15] D. Ahlers, S. Boll. Adaptive geospatially focused crawling. CIKM '09 Proceedings of the 18th ACM conference on Information and knowledge management 445-454, 2009