The Performance Gap

Agma J. M. Traina

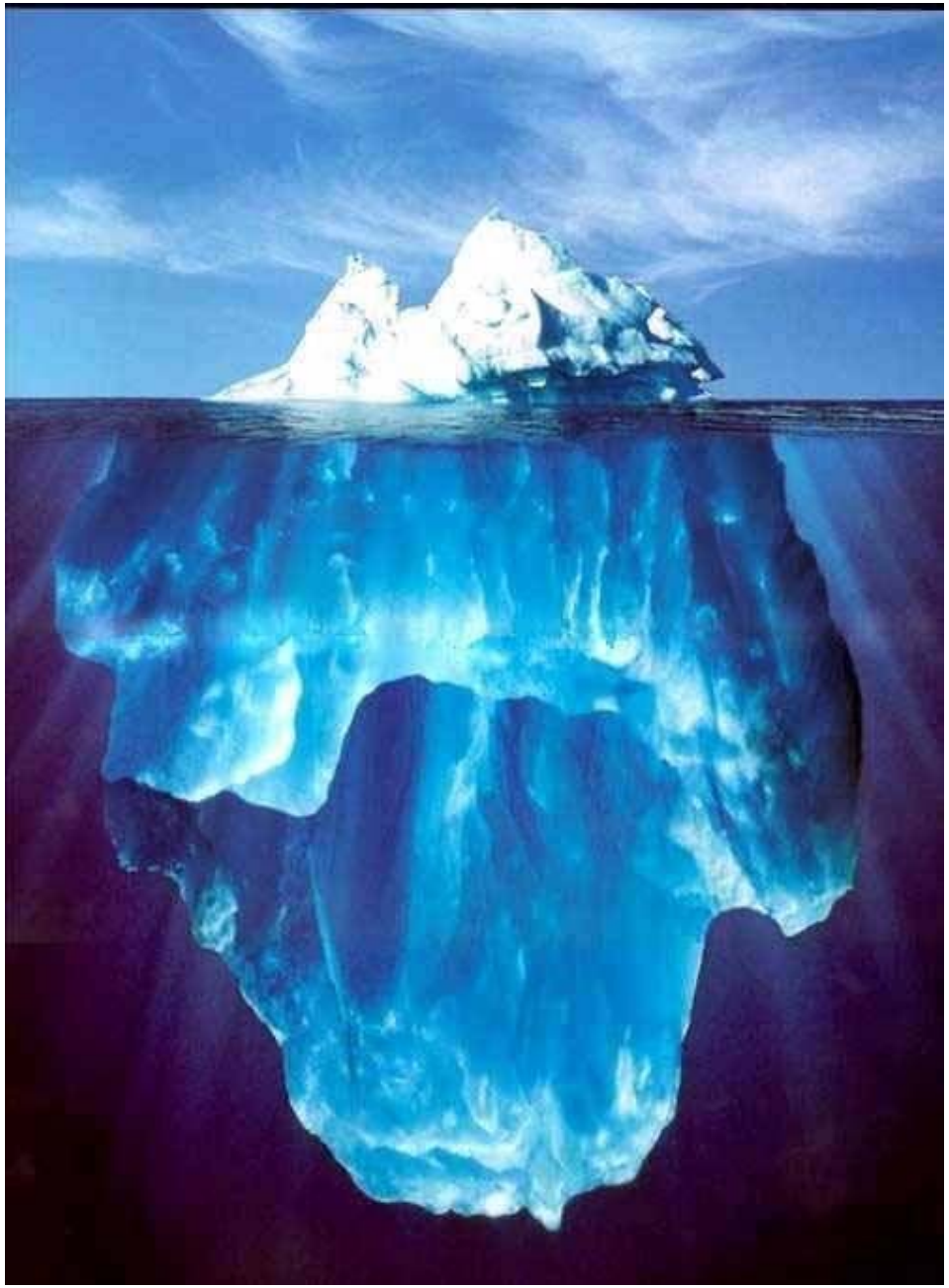
GBDI – Database and Image Group
Department of Computer Science
University of São Paulo at São Carlos
SP- Brazil

Motivation

What impairs the acceptance of CBIR Systems ?

Motivation

The problem is deeper than one first imagine!



The Problem – Performance Gap

- A CBIR system should bring the **meaningful** answer in a **timely manner**, being **integrated** to the clinical workflow
- Employs the necessary means to do so:
  - Feature selection
  - Indexing
  - Query optimization
  - Evaluation

Feature Selection

- A large number of features extracted from the images
  - **Dimensionality Curse![1]**
- Many features are correlated
  - **Blurs the images discrimination/separability[2,3]**
- Feature Selection => aims at keeping the features that best differentiate the images
- Approaches: decision trees, association rules, fractal theory, …

- Goal: to get **few** and **highly** distinguishing features.
- Challenge: How to **select** them… and fast

Indexing

- Access Methods to accelerate the query processing
- Similarity Search
  - **k-Nearest Neighbor Query:** *"Find the 5 images most similar to the John Doe's XRay-Chest"*
  - **Range Query:** *"Find the images that differ from the John Doe's XRay-Chest* up to 10 units"
- Metric Access Methods – only the features and distances are necessary (e.g.: M-tree[4], Slim-tree[5,6], Omni Family[7], …)

- **Goal:** to keep up with the voluminous and ever increasing number of images and complex data
- **Challenge**: speed and scalability

Query optimization

- Develop optimized query algorithms inside de Database Management Systems[8],
- Make the most efficient query plan (pipeline) to perform complex queries[9],
- Prepare beforehand the most efficient plans to execute the most frequently asked queries,
- “New rules to a new world”, dealing with images and complex data.

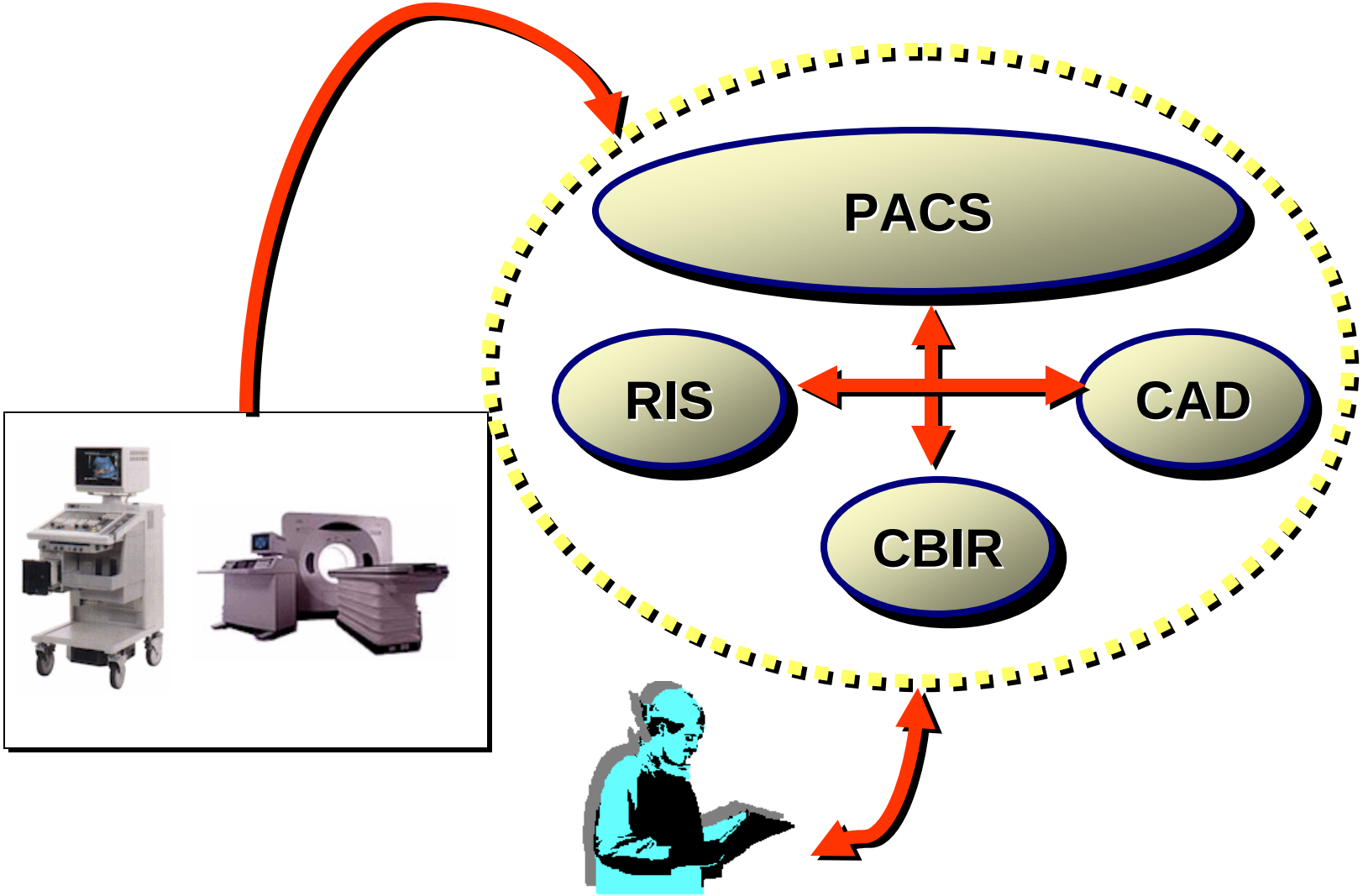
- **Goal:** to have the DBMS extended to query images by content
- **Challenge**: image data type defined and integrated to DBMS

Results evaluation

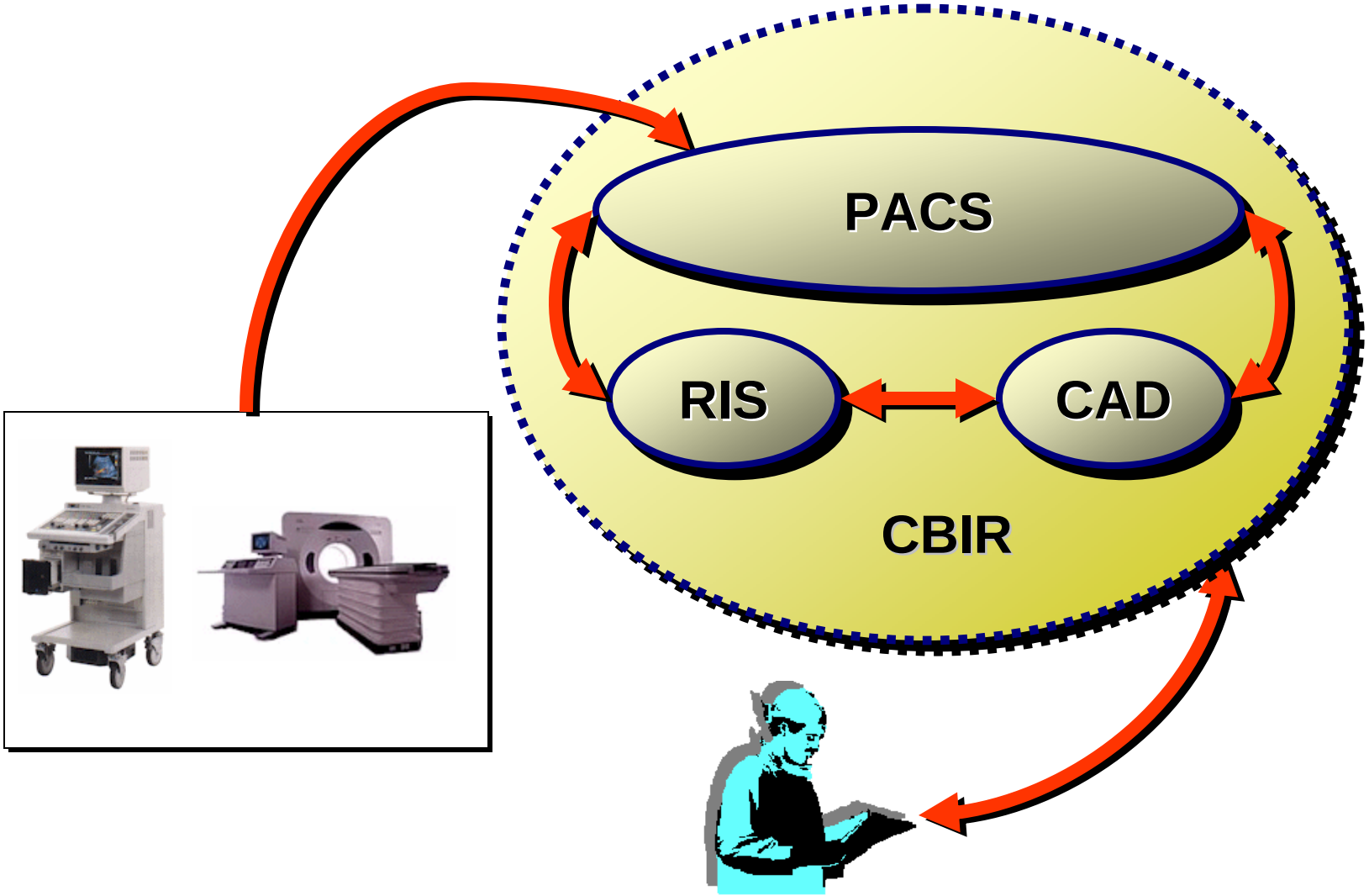
- Subjective: depends on the user interest at the moment,
- Include the user in the refinement step (Relevance Feedback)[10]
- Create user profiles.

- **Goal:** to allow real comparison among different methods/ algorithms
- **Challenge**: Ground truth, human perception

Integration to other Clinical systems



Integration to other Clinical systems



Pointers to further readings

- [1] F. Korn, B.-U. Pagel, and C. Faloutsos, "On the 'Dimensionality Curse' and the 'Self-Similarity Blessing'," IEEE Transactions on Knowledge and Data Engineering (TKDE), Vol. 13, No. 1, January/February 2001, pp. 96-111.
- [2] K. Beyer, J. Godstein, R. Ramakrishnan, and U. Shaft, "When is "Nearest Neighbor" Meaningful?," Proc. International Conference on Database Theory (ICDT), Jerusalem, Israel, January 10-12, 1999, pp. 217-235.
- [3] C. Domeniconi, D. Gunopulos, S. Ma, B. Yan, M. Al-Razgan, and D. Papadopoulos, "Locally adaptive metrics for clustering high dimensional data," Data Mining and Knowledge Discovery, Vol. 14, No. 1, February, 2007, pp. 63-97.
- [4] P. Ciaccia, M. Patella, and P. Zezula, "M-tree: An efficient access method for similarity search in metric spaces," Proc. International Conference on Very Large Databases (VLDB), Athens, Greece, September 1997, pp. 426-435.
- [5] C. Traina Jr., A. J. M. Traina, C. Faloutsos, and B. Seeger, "Fast Indexing and Visualization of Metric Datasets Using Slim-trees," IEEE Transactions on Knowledge and Data Engineering (TKDE), Vol. 14, No. 2, March/April 2002, pp. 244-260.
- [6] C. Traina Jr., A. J. M. Traina, C. Faloutsos, and B. Seeger, "Fast Indexing and Visualization of Metric Datasets Using Slim-trees," IEEE Transactions on Knowledge and Data Engineering (TKDE), Vol. 14, No. 2, March/April 2002, pp. 244-260.
- [7] R. F. Santos Filho, A. J. M. Traina, C. Traina Jr., and C. Faloutsos, "Similarity Search without Tears: The OMNI Family of All-purpose Access Methods," Proc. IEEE International Conference on Data Engineering (ICDE), Heidelberg, Germany, April 2-6, 2001, pp. 623-630.
- [8] C. Traina, Jr., R. F. Santos Filho, A. J. M. Traina, M. R. Vieira, and C. Faloutsos, "The OMNI-Family of All-Purpose Access Methods: A Simple and Effective Way to Make Similarity Search More Efficient," The International Journal on Very Large Databases, Vol. 16, No. 4, October 2007, pp. 483-505.
- [9] M. R. Vieira, C. Traina, Jr., A. J. M. Traina, A. S. Arantes, and C. Faloutsos, "Estimating Suitable Query Radii to Boost kNearest Neighbor Queries," Proc. 19th International Conference on Scientific and Statistical Database Management (SSDBM 2007), Banff, Canada, July 9-11, 2007, pp. 10.
- [10] A. J. M. Traina, J. Marques, and C. Traina, Jr., "Fighting the Semantic Gap on CBIR Systems through Relevance Feedback Techniques," Proc. 19th IEEE Intl Symposium on Computer-Based Medical Systems (CBMS 2006), Salt Lake City, Utah, USA., 22-23 June 2006, pp. 881-886.

The Performance Gap

Thank you!

Agma J. M. Traina

agma@icmc.usp.br

GBDI – Database and Image Group
Department of Computer Science
University of São Paulo at São Carlos
SP- Brazil

Questions ?

CAPES FAPESP CNPq