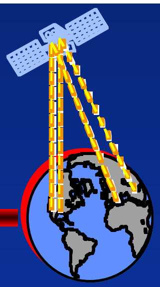
Nearest Neighbor Queries

Nick Roussopoulos
Stephen Kelley
Frederic Vincent

University of Maryland
May 1995

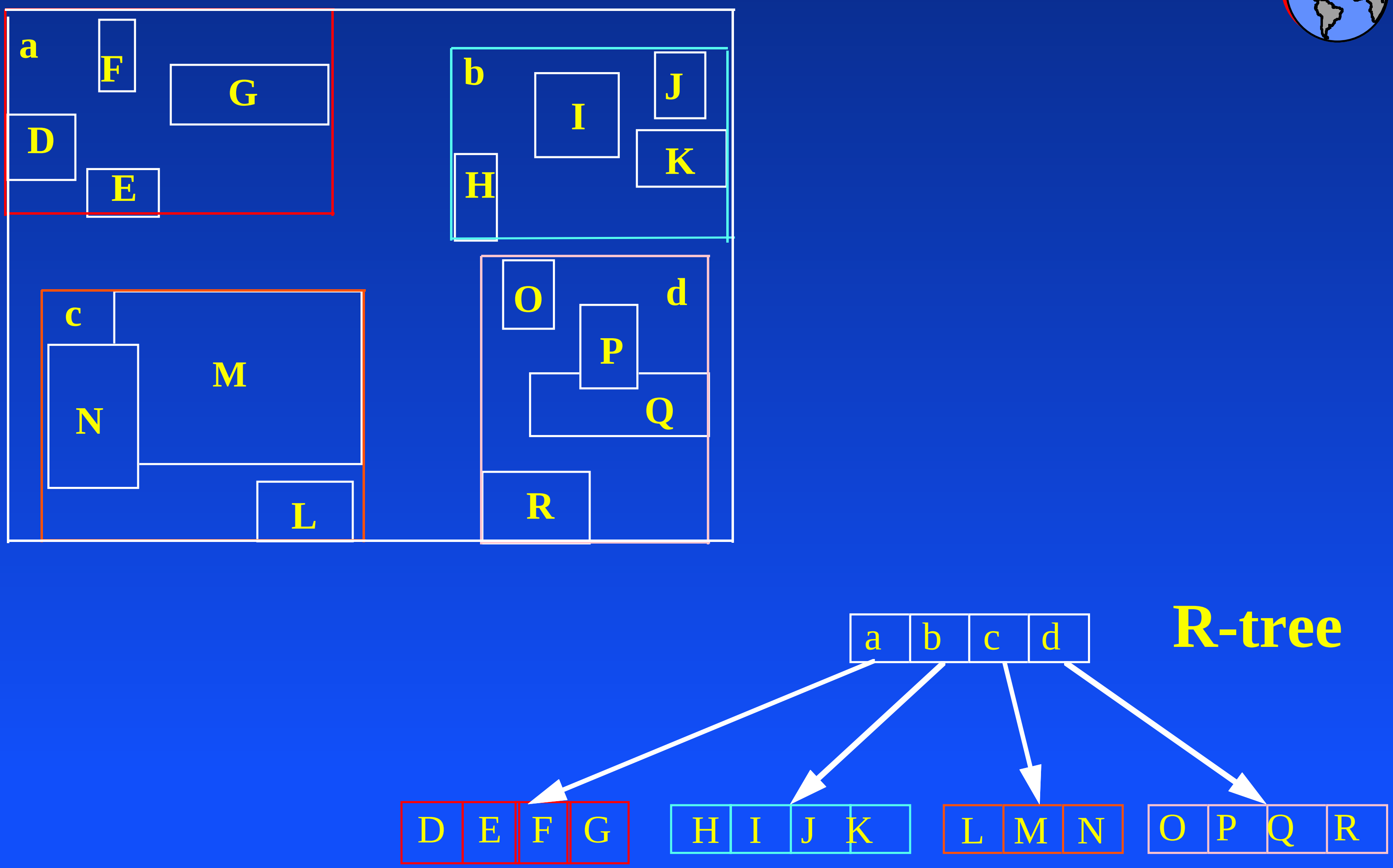
Problem / Motivation



Given a point in space, find the k NN

- **classic NN queries** ***(find the nearest 5* French restaurants)***
- **bounded queries** ***(gas stations between 10 and 15 mile range)***
- **spatial joins combined with NN** ***(3 NN restaurants to each movie theater)***
- **useful to formulate queries when exact location specification is hard** ***(astrophysics)***
- **combined with other geographic relationships and spatial predicates** ***(5 NN cities west of Mississippi)*** ***(NN houses on lots larger than 2 acres)***
- **Furthest Neighbors and other distance ordering functions**

R-trees

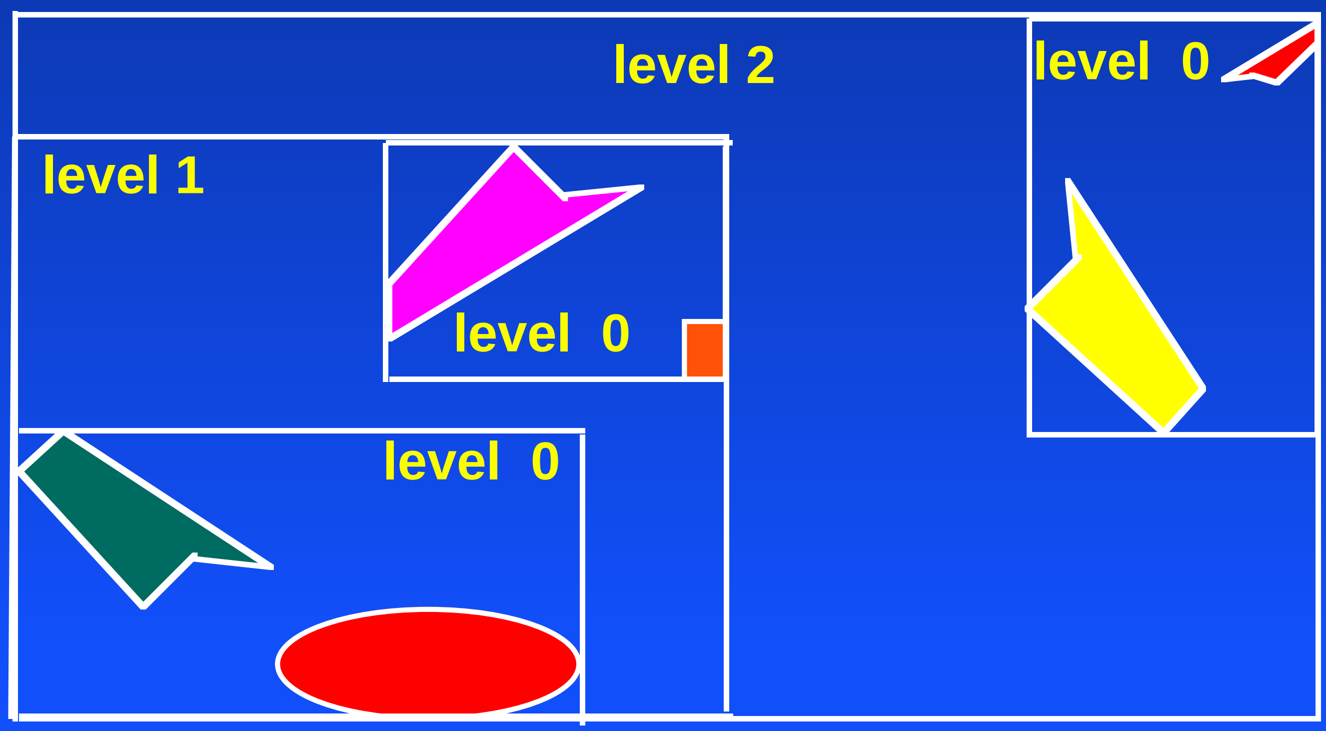


Need for Formalization of NN Queries

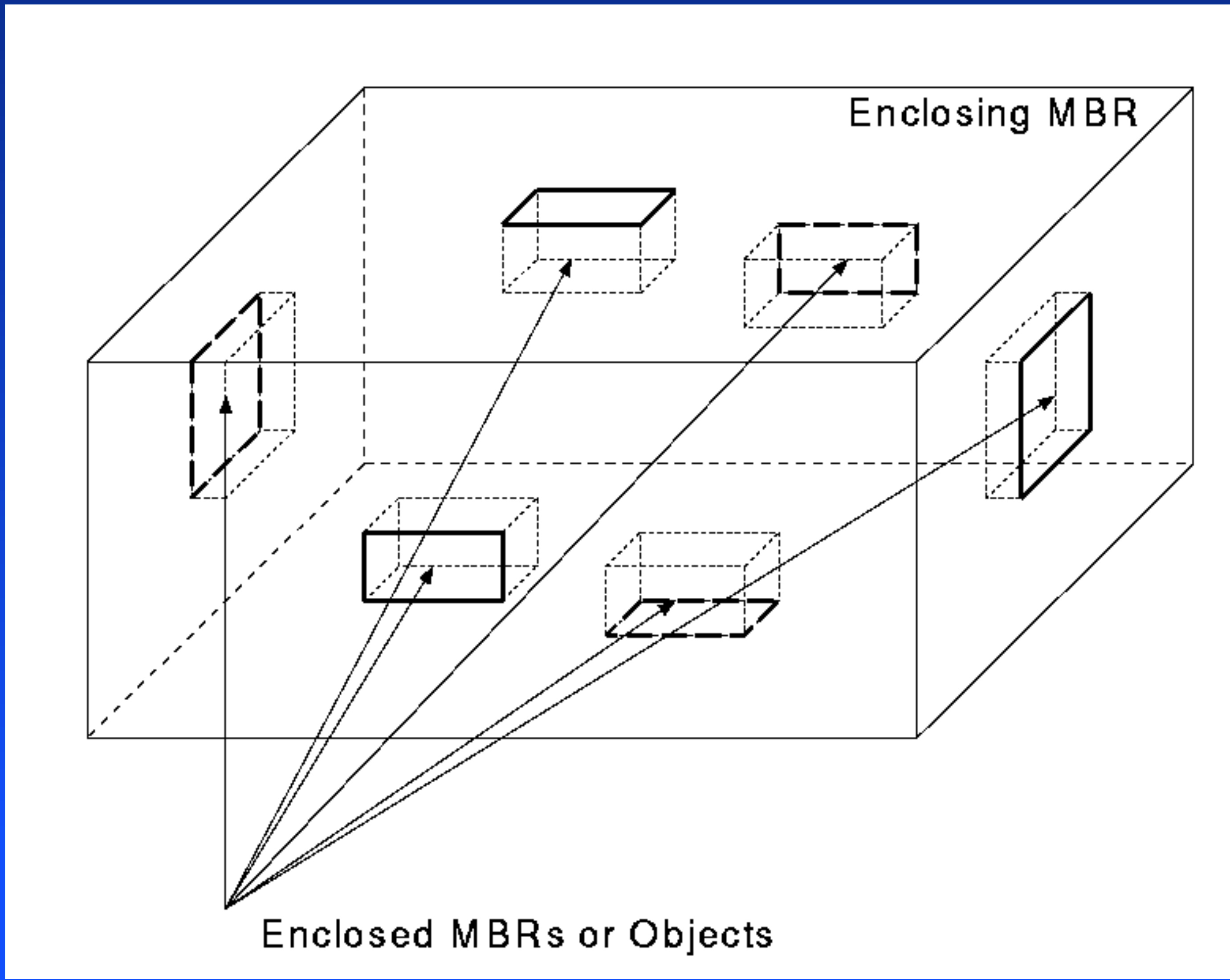
- **No formalism for NN search**
- **No metrics for ordering and pruning the search**
- **R-trees only for overlap/containment queries**
- **R-tree based spatial joins used only overlap/containement predicates**

The MBR Face Property

Lemma: **Every face of any MBR contains at least one point of an actual spatial object**



The MBR Face Property in 3-d



NN Metrics

- **MINDIST(P,R): the shortest distance from P to R**

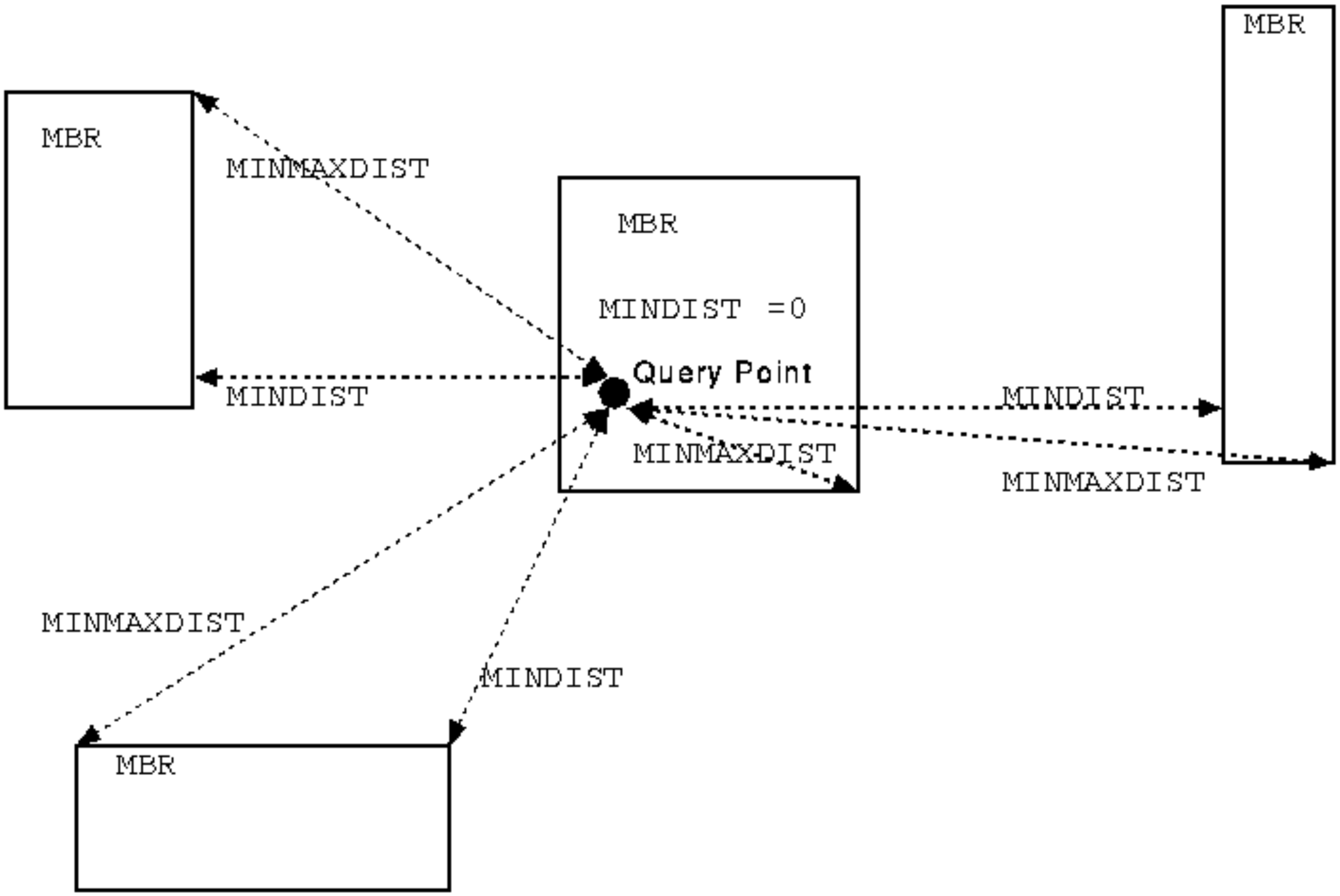
Theorem: **Any object O in R has distance from P that is at least as large as MINDIST**

- **MINMAXDIST(P,R) the minimum over all dimensions distance from P to the furthest point of the closest face of the R**

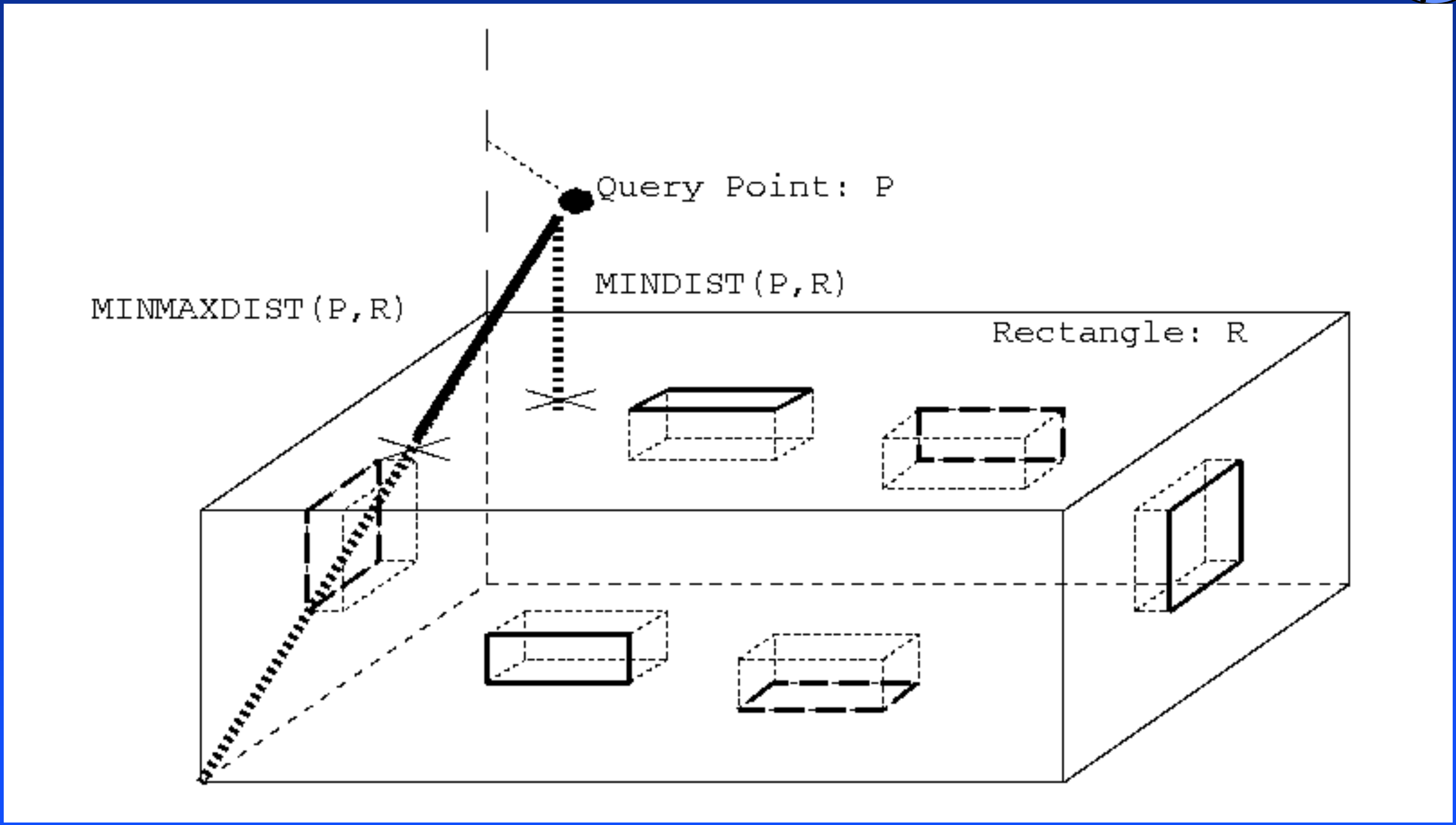
Theorem: **There exists at least one object within R with distance <= to MINMAXDIST**

MINDIST(P,R) <= NN(P) <= MINMAXDIST(P,R)

MINDIST & MINMAXDIST



MINDIST & MINMAXDIST in 3-d



NN Branch-and-Bound Algorithm

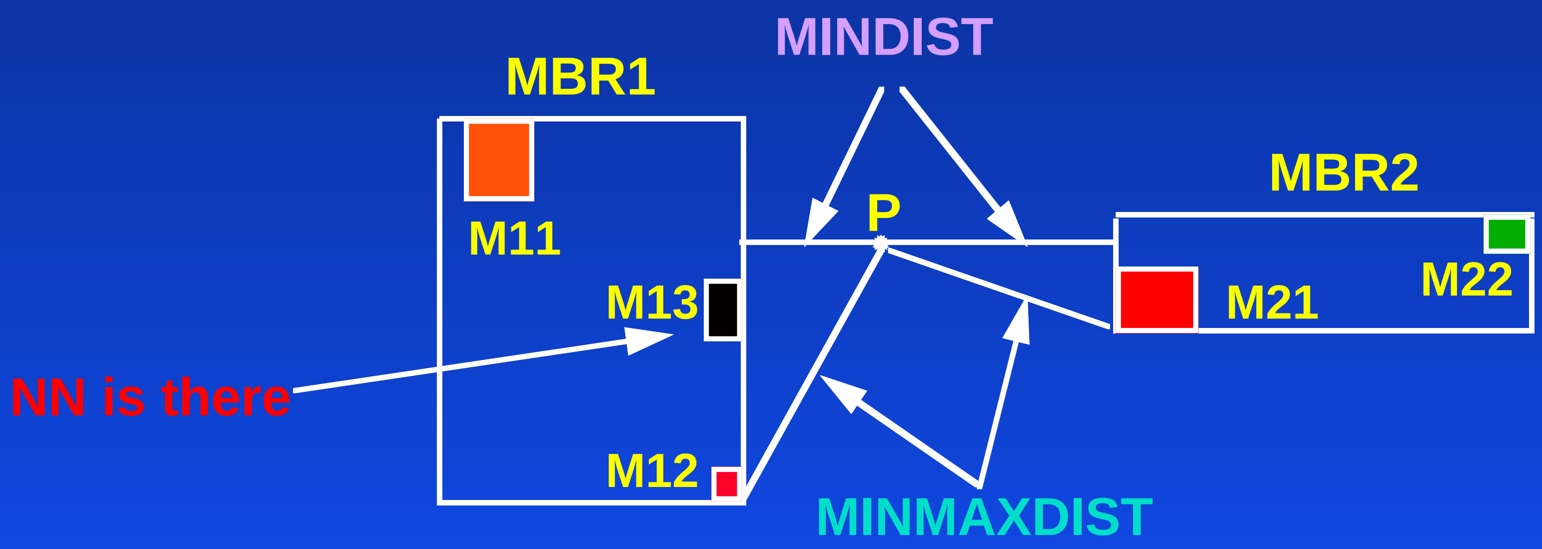
Ordering search alternatives

- **MINDIST is the most optimistic**
- **MINMAXDIST is the most pessimistic ever needed be considered**

Pruning search alternatives

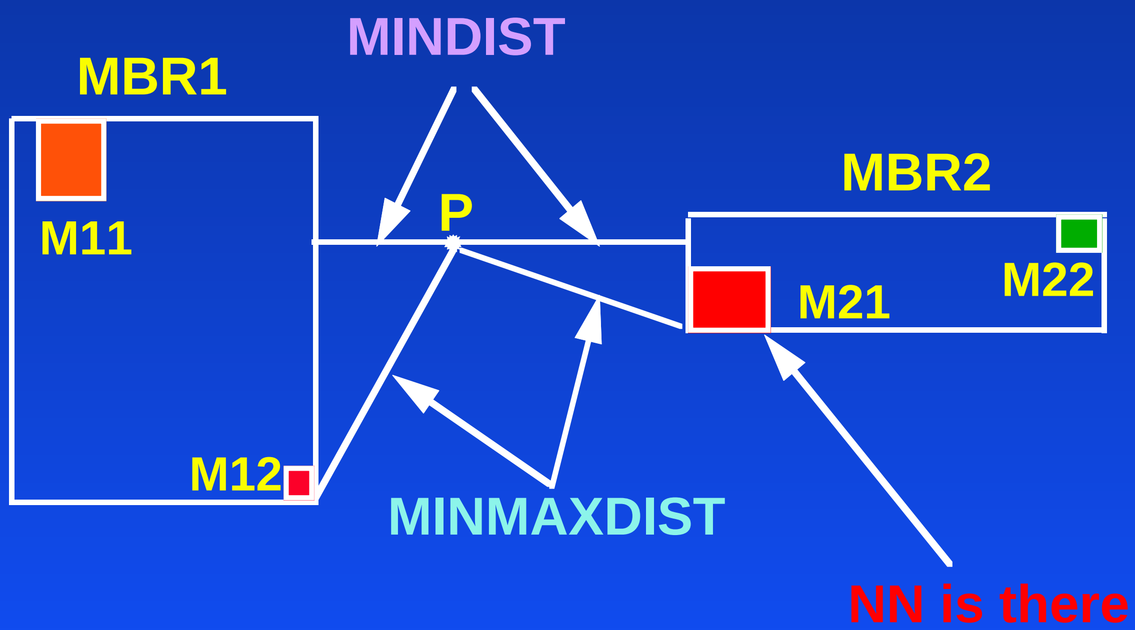
- **downward pruning: an MBR R is discarded if there exists an R' s.t.**
  MINDIST(P,R) > MINMAXDIST(P,R')
- **downward pruning: an object O is discarded if there exists an R s.t.**
  ACTUAL-DIST(P,O) > MINMAXDIST(P,R)
- **upward pruning: an MBR R is discarded if an object O is found s.t.**
  MINDIST(P,R) > ACTUAL-DIST(P,O)

MINDIST vs MINMAXDIST Ordering



MINDIST leads the search to MBR1 which finds the NN in M13 and prunes MBR2 before visiting it

MINDIST vs MINMAXDIST Ordering



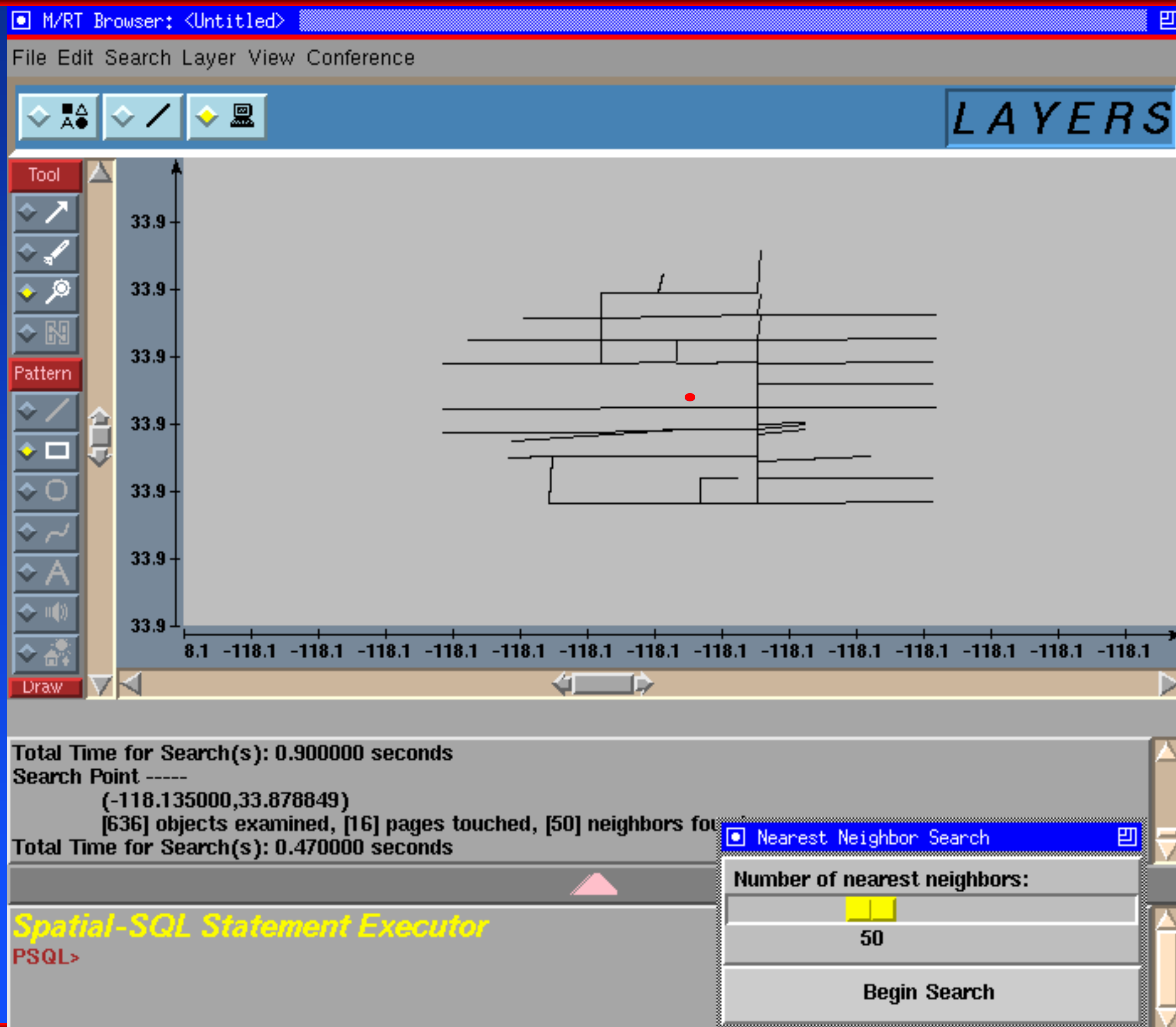
Generalization to k NN

- **Keep a sorted buffer of at most k NN**
- **Pruning is applied according to the furthest NNs in the buffer**
- **Extra pruning based on additional predicates**

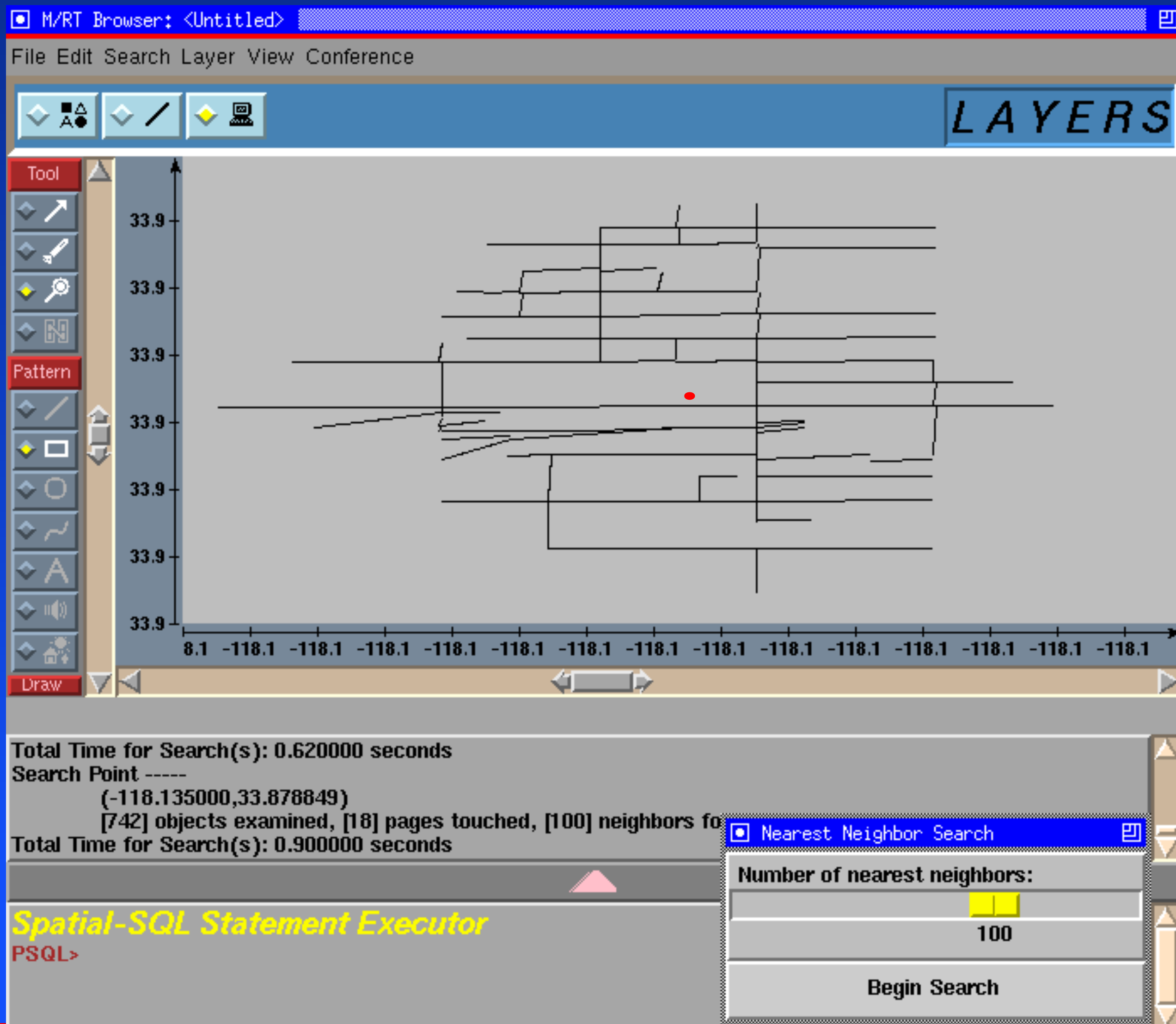
IMPLEMENTATION

- Extended PSQL syntax
- Uses our version of high performance R-trees
- Also runs on top of ORACLE / INGRES

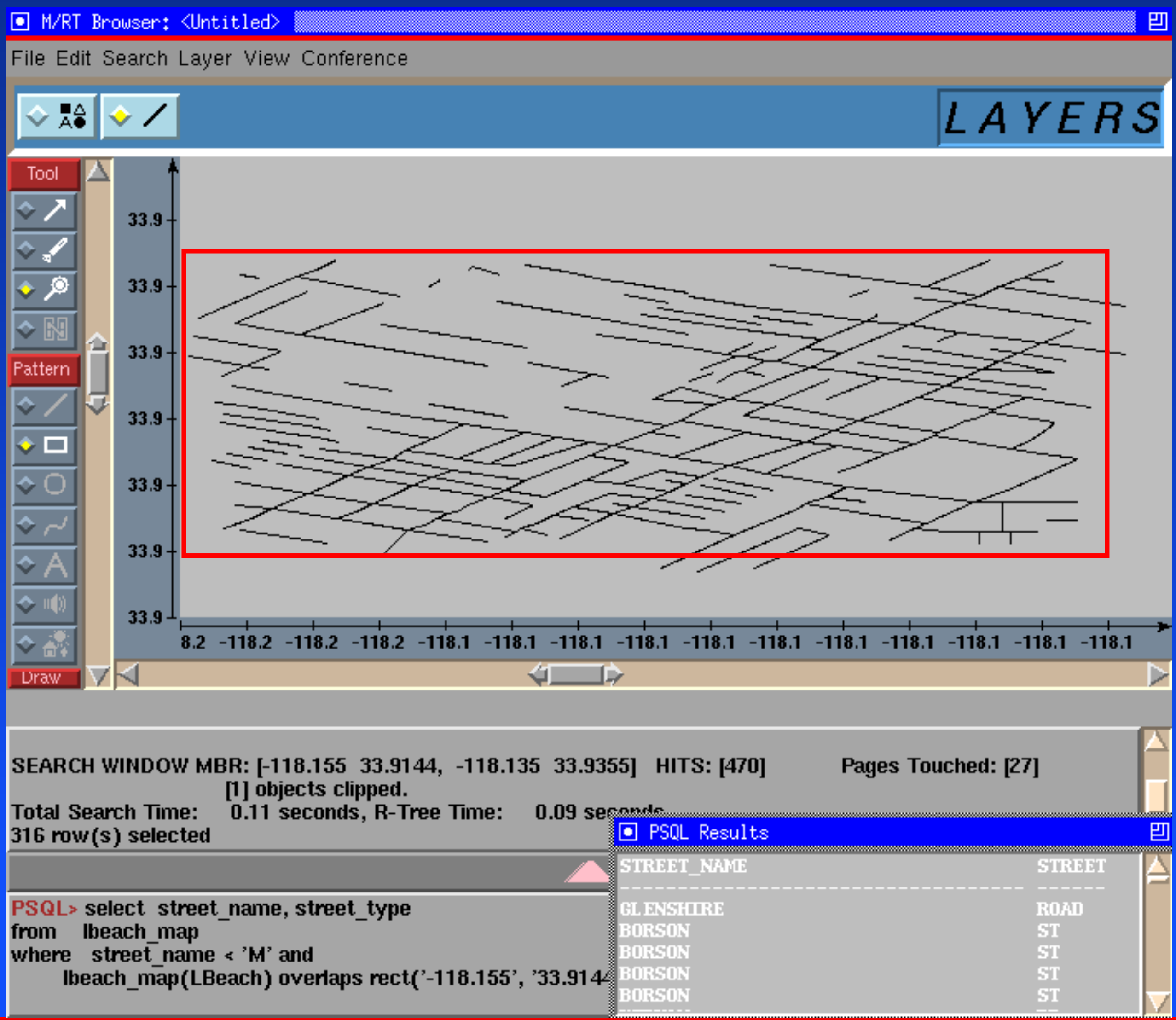
50 NN in LBeach



100 NN in LBeach

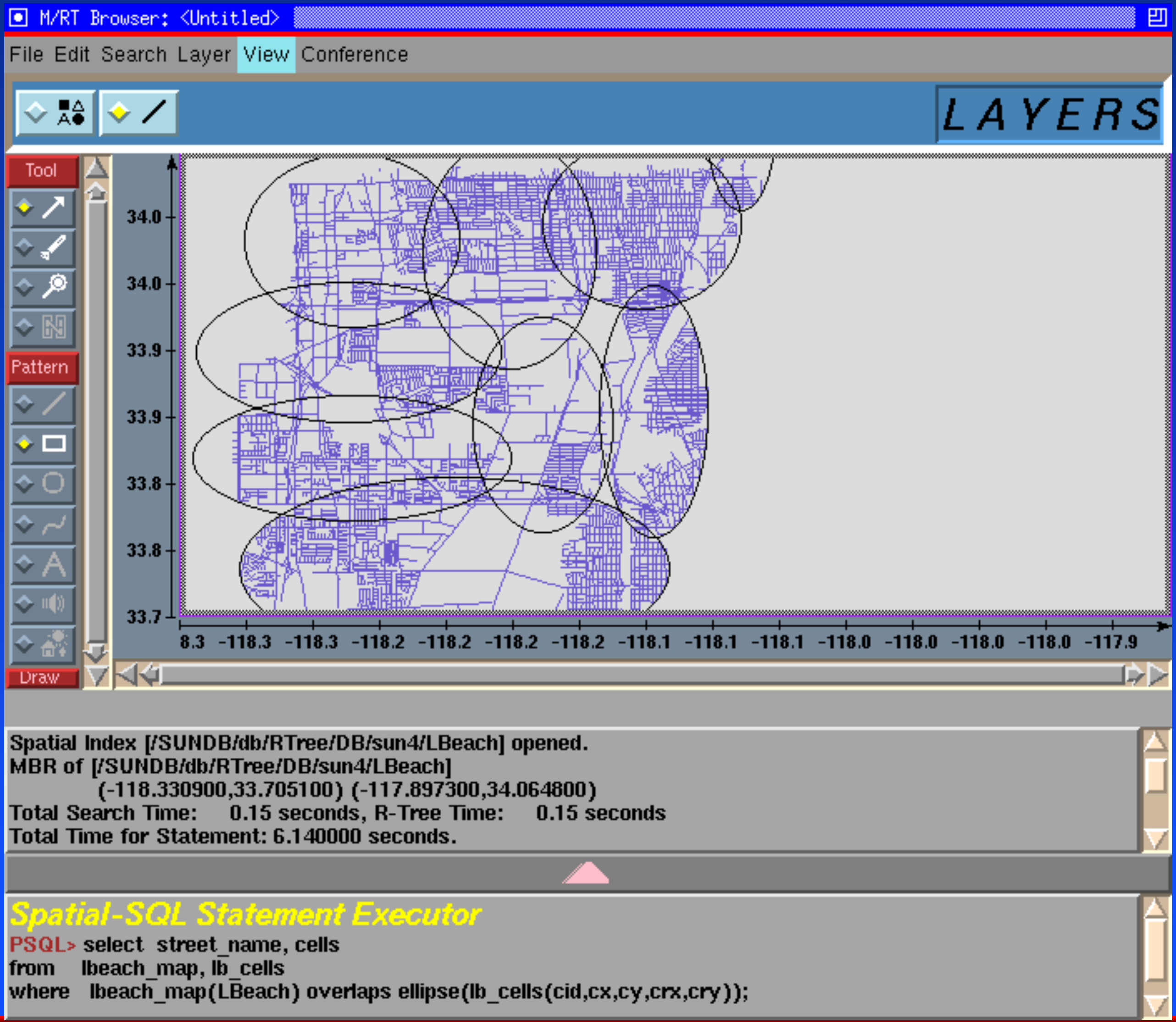


Range Query



M/RT Browser: <Untitled>
File Edit Search Layer View Conference
LAYERS
Tool
Pattern
Draw
33.9
8.2 -118.2 -118.2 -118.2 -118.1 -118.1 -118.1 -118.1 -118.1 -118.1 -118.1 -118.1 -118.1 -118.1 -118.1
SEARCH WINDOW MBR: [-118.155 33.9144, -118.135 33.9355] HITS: [470] Pages Touched: [27]
[1] objects clipped.
Total Search Time: 0.11 seconds, R-Tree Time: 0.09 seconds
316 row(s) selected
PSQL> select street_name, street_type
from lbeach_map
where street_name < 'M' and
lbeach_map(LBeach) overlaps rect('-118.155', '33.914
PSQL Results
STREET_NAME STREET
GLENSHIRE ROAD
BORSON ST
BORSON ST
BORSON ST
BORSON ST

Spatial Join with Elliptical Cells



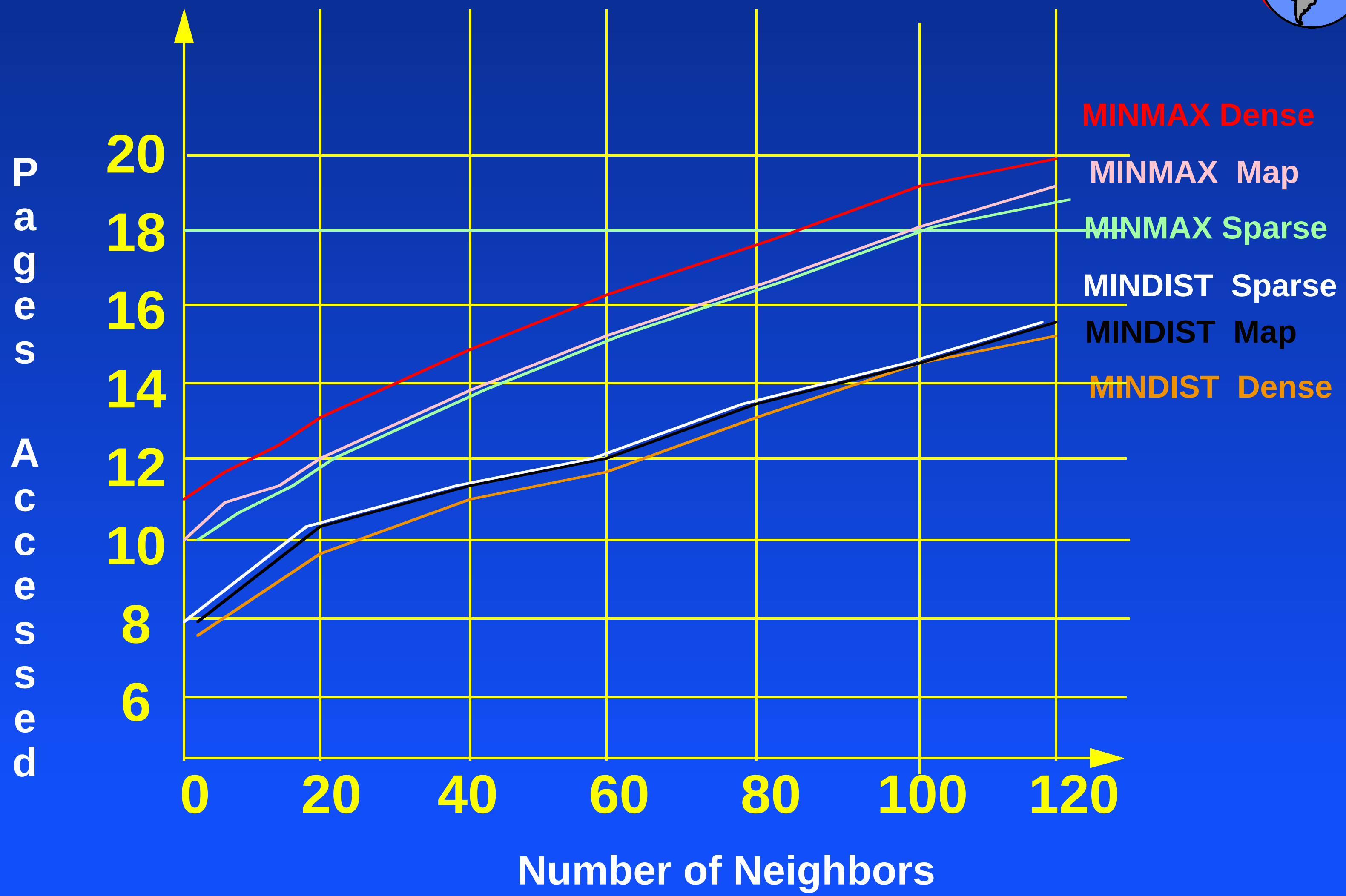
EXPERIMENTS

- TIGER data sets of Long Beach, CA and Montgomery County, MD
- International Ultraviolet Explorer (IUE)
- Synthetic data sets 1K-256K points
- Data sets were sorted according to their Hilbert order and R-trees were packed using [Rous+Leif1985],[Kame+Falo1993]
- Achieved the “maximum” performance for R-tree overlap, containment and NN searches

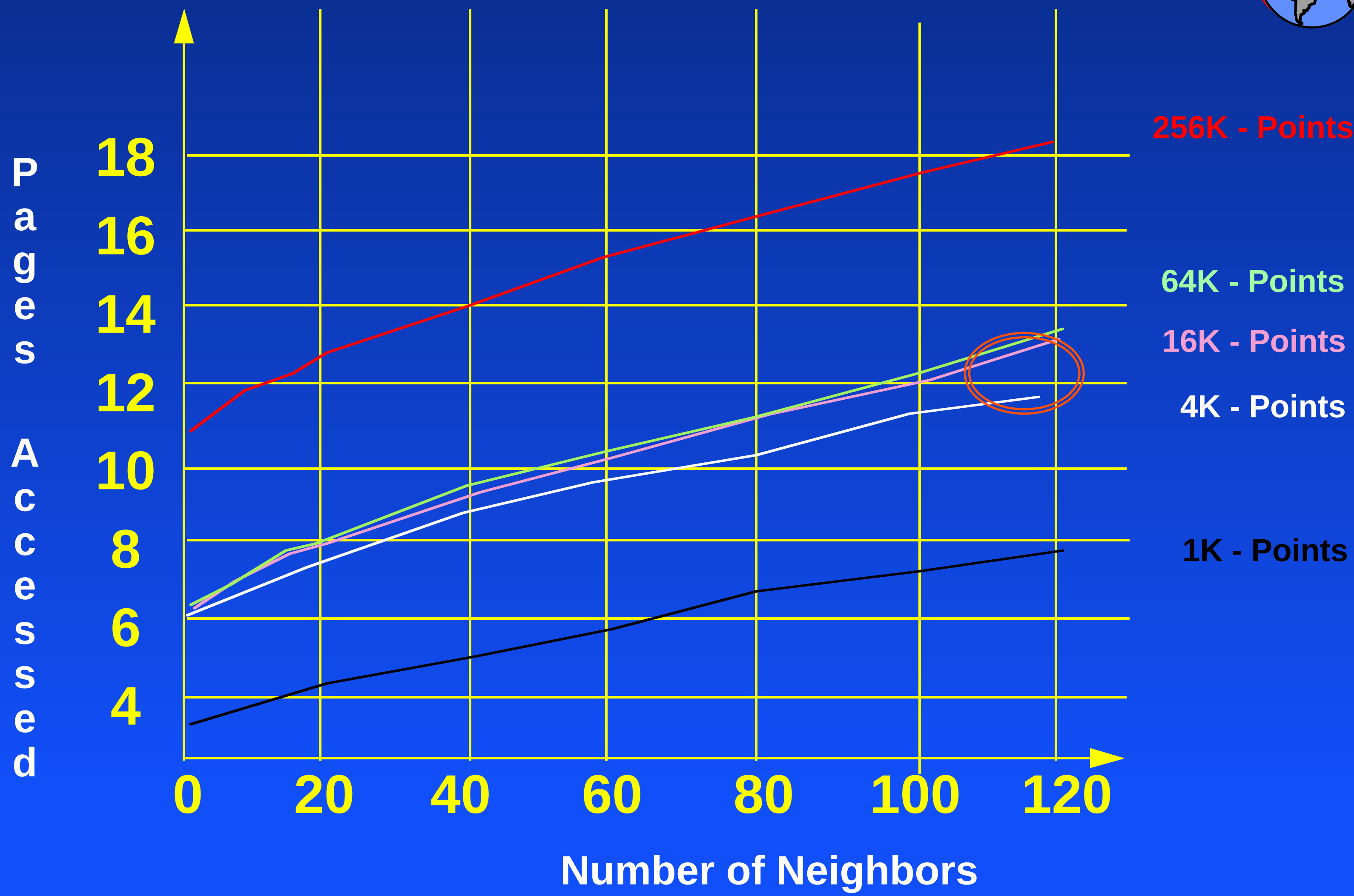
Comparisons

- Optimistic vs pessimistic ordering
- Scalabilty in the number of NN
- Scalability in the size of the index

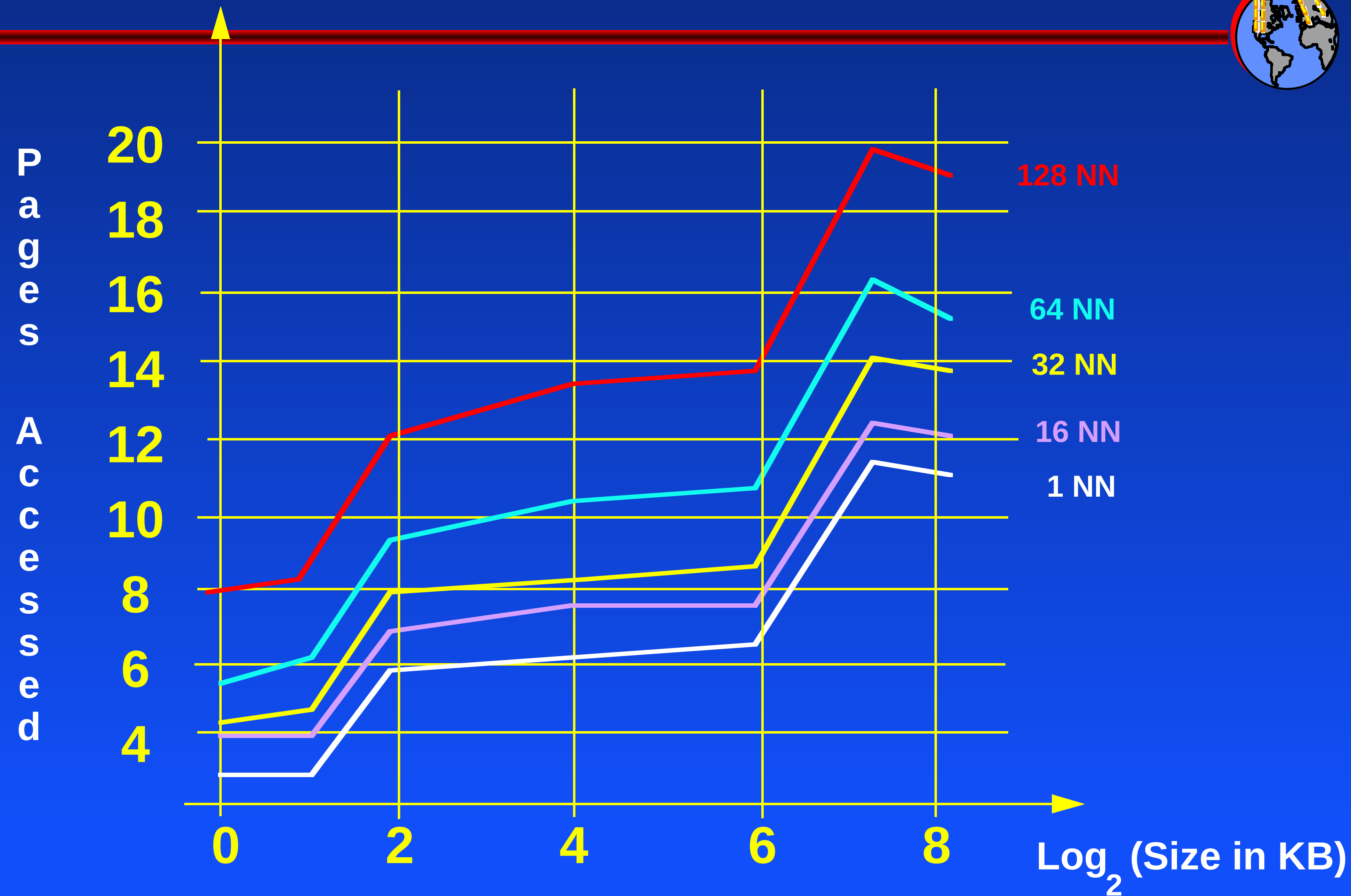
Optimistic vs. Pessimistic Ordering & Scalability in NN - Long Beach Map



Size Scalability - Synthetic Data



Size Scalability - Synthetic Data



CONCLUSIONS

- **Branch-and-bound NN search algorithm**
- **Formalized and provided metrics for directing NN search**
- **The optimistic MINDIST metric performed better in our experiments but this is related to the construction of the R-trees**
- **Generalized and displayed the versatility of k NN search**