제 6회 소프트웨어무결점 연구센터의 하계 워크샵
경북대학교 데이터베이스 연구실
Milestones
2011년 6월 25일
2011년 연구 내용

- 그래프 데이터베이스에서 기존의 인덱싱 기법의 성능을 공정하게 분석할 수 있는 프레임워크를 개발 (VLDB 2010)하고, 여러 알고리즘의 성능을 효과적으로 분석 할 수 있는 툴 개발 (SIGMOD 2011)
- 시계열 데이터베이스에서 순위 지원 서브시퀀스 검색 방법에 대한 연구 (SIGMOD 2011)

연구 성과
- **SIGMOD 2011 2편 논문 발표 (데모 포함)**
- “iGraph in Action: Performance Analysis of Disk-Based Graph Indexing Techniques.”
- “A New Approach for Processing Ranked Subsequence Matching Based on Ranked Union.”
2010년도 연구 성과

- VLDB 2010, ICDE 2010에 총 2편 발표

- "iGraph: A Framework for Comparisons of Disk Based Graph Indexing Techniques."
  Wook-Shin Han, Jinsoo Lee, Minh-Duc Pham, and Jeffrey Xu Yu.
  *Proceedings of the Very Large Data Base Endowment (PVLDB).* 3 (1). 2010.

- "On Supporting Effective Web Extraction."
  Wook-Shin Han, Wooseong Kwak, and Hwanjo Yu.
  *26th IEEE International Conference on Data Engineering (ICDE’10).* 2010.
2009년도 연구 성과

- SIGMOD 2009에 논문 발표

- "Dependency–Aware Reordering for Parallelizing Query Optimization in Multi–Core CPUs."
  Wook–Shin Han and Jinsoo Lee.
"StreamTX: Extracting Tuples from Streaming XML Data."
Wook-Shin Han, Haifeng Jiang, Howard Ho, and Quanzhong Li.
*Proceedings of the Very Large Data Base Endowment (PVLDB).* 1 (1). 2008.

"Parallelizing Query Optimization."
Wook-Shin Han, Wooseong Kwak, Jinsoo Lee, Guy M. Lohman, and Volker Markl.
*Proceedings of the Very Large Data Base Endowment (PVLDB).* 1 (1). 2008.

"Ranked Subsequence Matching in Time-Series Databases."
Wook-Shin Han, Jinsoo Lee, Yang-Sae Moon, and Haifeng Jiang.
*Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB'07).* 2007.

"Progressive Optimization in a Shared-Nothing Parallel Database."
Wook-Shin Han, Jack Ng, Volker Markl, Holger Kache, and Mokhtar Kandil.
Similarity search in time-series databases
What are time-series data?
Red Howler Monkey

Mantled Howler Monkey

Orangutan (juvenile)

Borneo Orangutan

Mantled Howler Monkey
Hand at rest
Hand moving to shoulder level
Steady pointing
Hand at rest
Hand moving above holster
Hand moving to shoulder level
Steady pointing
Hand moving above holster
Hand at rest
Hand moving down to grasp gun
Hand at rest

(c) Eamonn Keogh, eamonn@cs.ucr.edu
Hey Jude don’t make it bad,

Take a sad song and make it better.
What is Similarity Search?
Euclidean Distance
Euclidean
$Q = q_1 \ldots q_n$

$C = c_1 \ldots c_n$

$$D(Q, C) \equiv \sqrt{\sum_{i=1}^{n} (q_i - c_i)^2}$$
R* trees
Dimensionality
With R* tree

Without R* tree

Elapsed Time

Dimension
Dimension Reduction!

M (256) → N (8)
Piecewise Aggregate Approximation
\[
\overline{C_i} = \frac{N}{n} \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} C_j
\]
Important Lemma
D(Q, C) < \varepsilon \Rightarrow D(PAA(Q), PAA(C)) < \varepsilon
Pruning!
We can prune any C such that
\[ D(\text{PAA}(Q), \text{PAA}(C)) > \text{cur\_top\_kth\_similar} \]