DBTech EXT

Virtual Laboratory Workshop
on
Knowledge Discovery from Databases
(KDD VLW)

Dimitris A. Dervos and Leonidas Karamitopoulos
Information Technology Department
Alexander Technology Educational Institute
Thessaloniki, Greece

A) The tutorial part

Video presentation and ppt slides on:

- Data vs. Information
- Shallow/Multidimensional/Hidden/Deep Information
- The KDD Process
- Data Mining Strategies
- Information Representation and Interpretation

B) Self-practicing part

Exercises with model answers on:

- Affinity Analysis (Apriori algorithm)
- Decision Trees (J48/C4.5)
- Clustering (K-Means, DBSCAN, Agglomerative)

C) Hands-On Self-Study and Practicing

- Association rules mining (WEKA, IBM I.M.)
- K-MEANS clustering (WEKA)
- Classification (J48/C4.5, WEKA)
- Using the K-MEANS output to improve the quality of the J48 output (WEKA)
- Recommender system type data processing with association rules mining, improved with demographic clustering*

  * **IBM DB2 DWE IM Tutorial: *Managing your business in retail***

     http://www.ibm.com/developerworks/edu/dm-dw-dm-retail_tutorial-i.html

D) Laboratory Report

Association Rules Mining:

- Software used: <u>WEKA</u> and <u>IBM Intelligent Miner</u>
- Datasets: (a) <u>HEAL-Link</u> <u>e-resources/journals</u> <u>dataset</u>
         (b) <u>Titanic dataset</u>

E) Quiz-Assessment

<u>MCQ type</u>
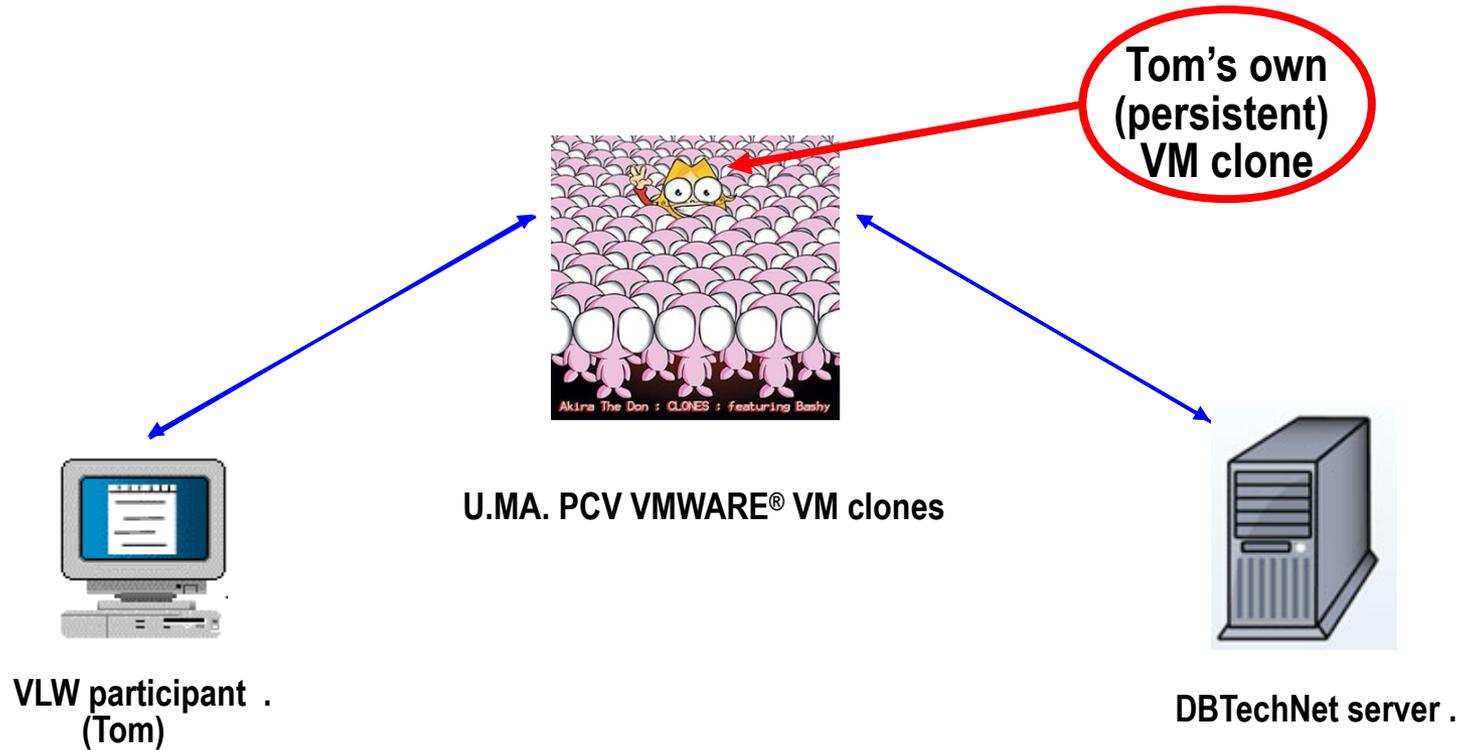
F) Supporting Material

'How To' instructions for installing the VLW software, user guides (with videos), trainer/trainee guides, bibliography, etc.

- <u>Three-level access authorization scheme</u>:
  public access, portal user access,
  KDD unit user access

- Single-point of entry to all VLW content
  (<u>KDD "RoadMap"</u>)

Tom's own (persistent) VM clone

U.MA. PCV VMWARE® VM clones

VLW participant .
(Tom)

DBTechNet server .

…resolves the commercial software licensing issue
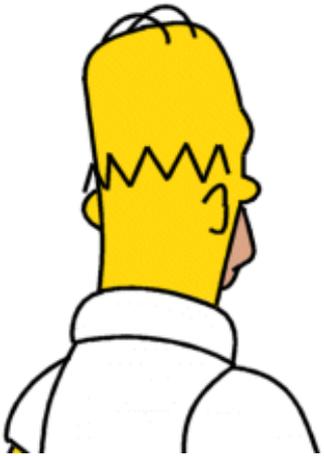
A) Asynchronous

- email
- Discussion fora
- Google Docs forms

B) Synchronous

- 1-1: Skype (with screen sharing, if needed)
- 1-M: Adobe Connect Pro (UMA, HAAGA-HELIA)

- DBTech EXT KDD VLW overview
- Content outline and organization
- Self-study, self-practicing (model answers, step-by-step instructions for the 'hands-on' part of the VLW
- Open source and commercial software used
- Virtualization technology: used both for the distribution of the software, as well as for overcoming the commercial software user licensing barrier

Thank You!