Web Feature Selection using Ant Colony Optimization

Presented by:

Group members:

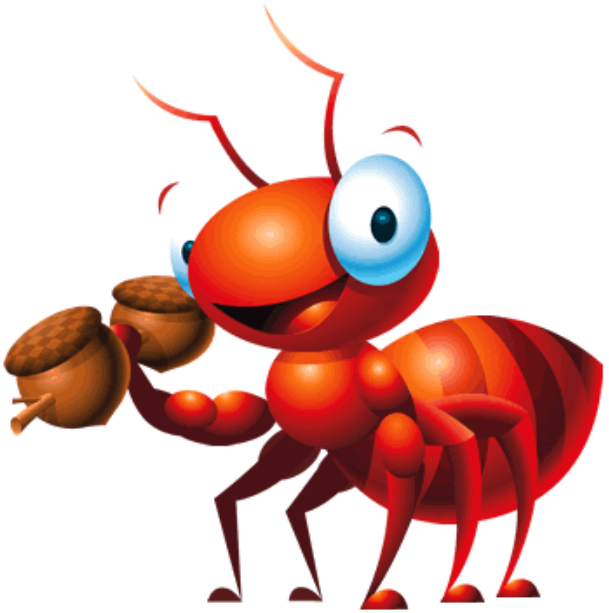
Group no 2

Antony Jino A T

Jithin Pradeep

Midhula Vijayan

Najeera P M



Plan

- Introduction to Feature Selection.
- ACO Software agents- ANT.
- Overview of Proposed Algorithm.
- System design.
- Related docs and implementation details.
- Conclusion & Future Works.
- References.
- Questions Hour.

Introduction to Feature Selection

- Feature selection (FS) is a commonly used step in machine learning which simplify a dataset by reducing its dimensionality and identify relevant underlying features without sacrificing predictive accuracy.
- It selects subset of relevant features
- In real world problems FS is a must due to the abundance of noisy, irrelevant or misleading features. Feature selection is useful in many fields, including text categorization, data mining, pattern recognition and signal processing

ACO & Software Agents -ANT

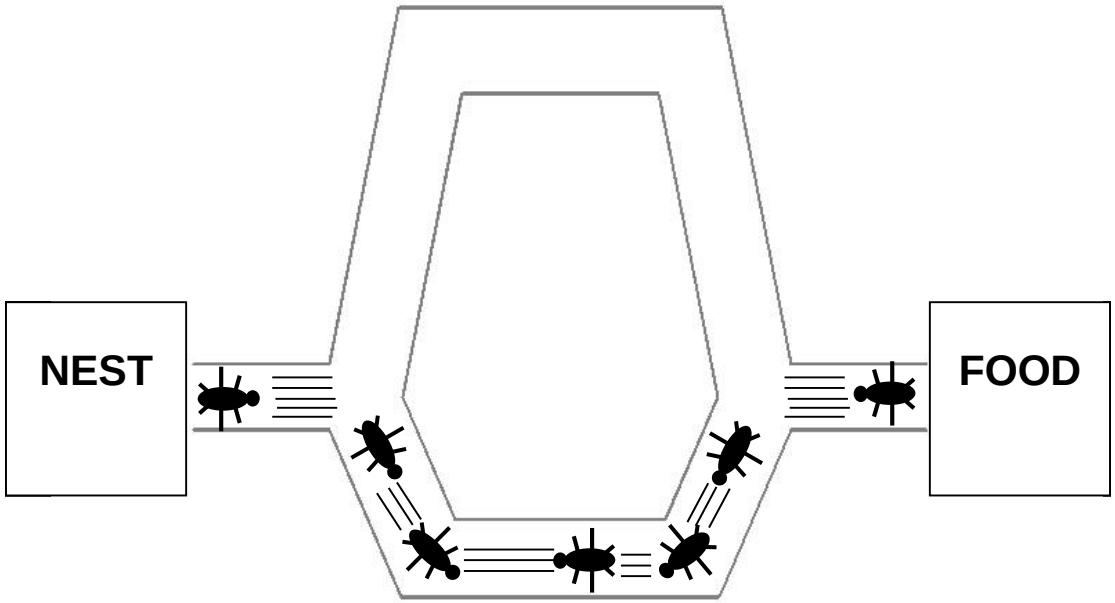
- We put forward a new ACO-based technique to the problem of feature selection in web page feature selection.

ACO- Ant colony optimization???

- ACO is an iterative algorithm. Belonging to class of algorithm in the branch of study called Swarms Intelligence

- Inspired by the natural behavior of ants. Works on pheromone model of computing.

Study Of Ants :Double bridge exp.

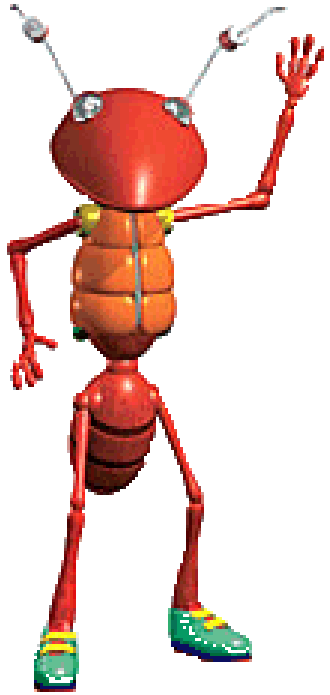


Real vs. Artificial ANTS

REAL ANT

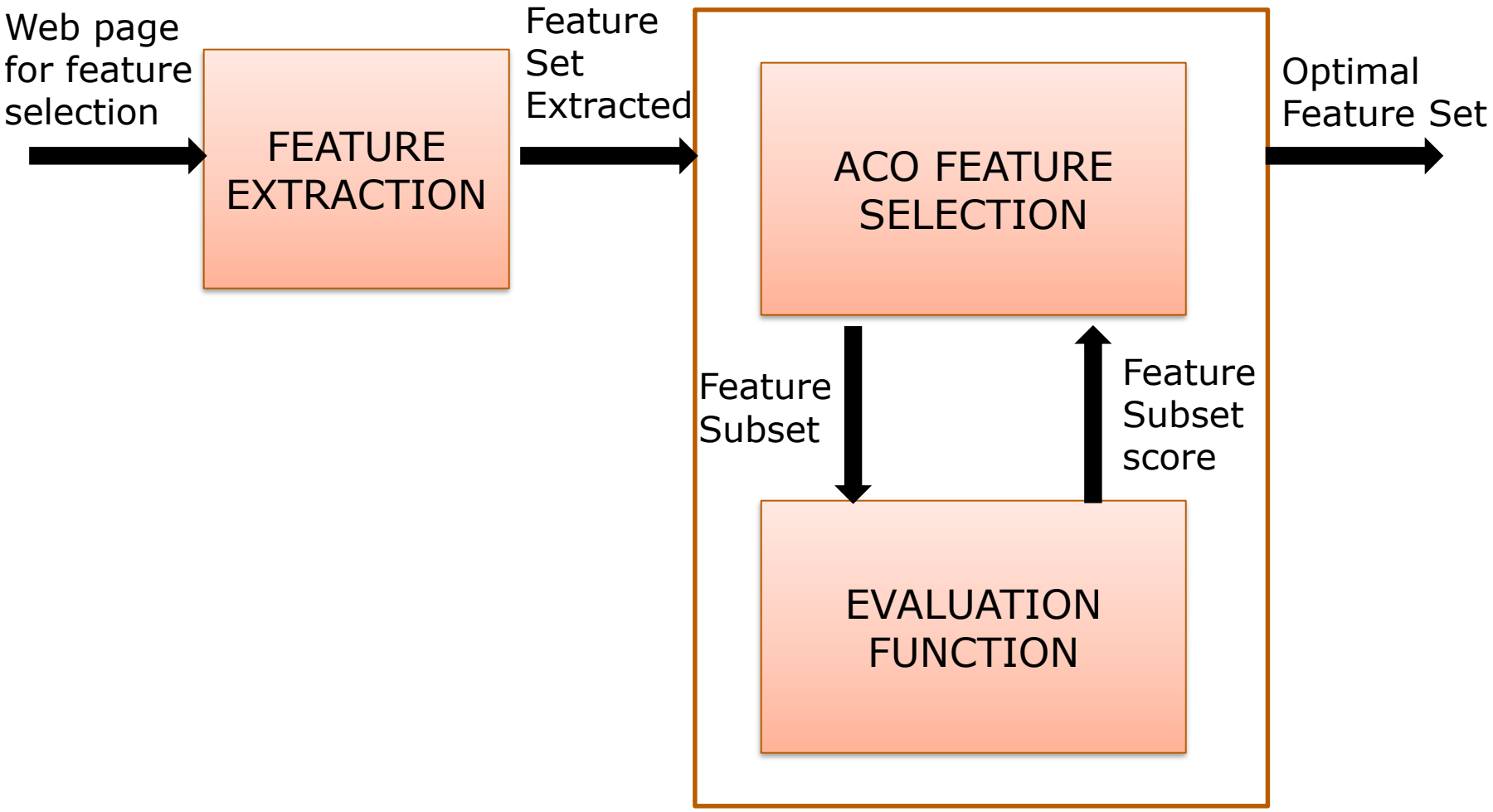
+

- Discrete time steps
- Memory Allocation
- Quality of Solution
- Time of Pheromone deposition
- Distance Estimation



SOFTWARE AGENT ANT

Overview of the Proposed System



Algorithm for Feature Extraction

Feature Extraction :

Input: The algorithm takes web document as input.

Output: Extracted feature set along with their normalized weight and occurrence.

Algorithm: Feature Extraction

Step1: Preprocess the web document.

Step2: Stop word removal, and store the features into feature set

Step 3: Perform Stemming on feature set.

Step 4: Assign a weight to each feature in feature set

Step 4.1: Compute the Occurrence of feature

Step 4.2: Compute the weight of feature based on Occurrence and there place of occurrence in the html document.

Feature Extraction algorithm contd

Step4.3: Add the weight of feature to Total weight.

Step 5: Compute normalized weight for each feature in feature set

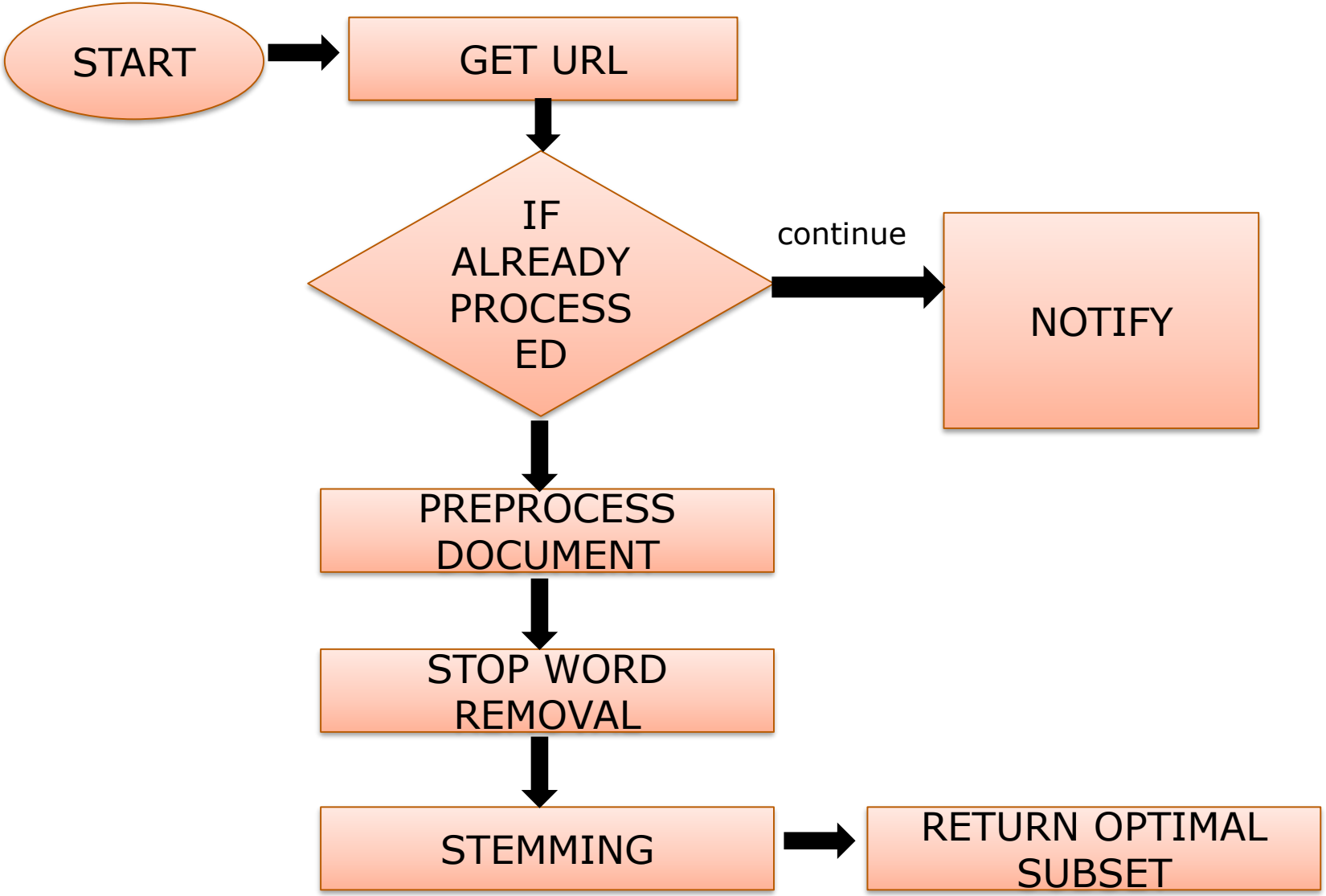
Step 5.1: Nr. weight is computed as

Nr. weight= (Weight of the feature/Total weight)*100

Step 5.2: Store Nr. weight and Occur of feature in the Extracted feature set.

Step 6: Return Extracted feature set.

Flow Chart for Feature Extraction



Algorithm for Feature Selection

Feature Selection :

Input: The algorithm takes Extracted feature set.

Output: Optimal feature set.

Algorithm: Feature Selection

Step1: Initialization.

Step1.1: Determine the population of ants.

Step1.2: Initialize the global list.

Step1.3: Set the intensity of pheromone trial associated with any feature.

Step1.4: Determine the maximum of allowed iterations.

Step2: Solution generation

Step2.1:Assign each ant randomly to one feature and initialize the local list of each ant

Feature Selection algorithm contd…

Step2.2: Each ant select a feature node which is not yet selected by the ant and based on probability value and update the local list of each ant

$$p_{ij}^{k}(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_{k \in allowed_k}[\tau_{ik}(t)]^{\alpha}[\eta_{ik}]^{\beta}} & \text{if } j \in \text{allowed}_k \\ 0 & \text{otherwise} \end{cases}$$

Step 3: Evaluation of the selected subsets.

Step 3.1: local list of each ant are compared with global list if found better global list is updated.

Feature Selection algorithm contd…

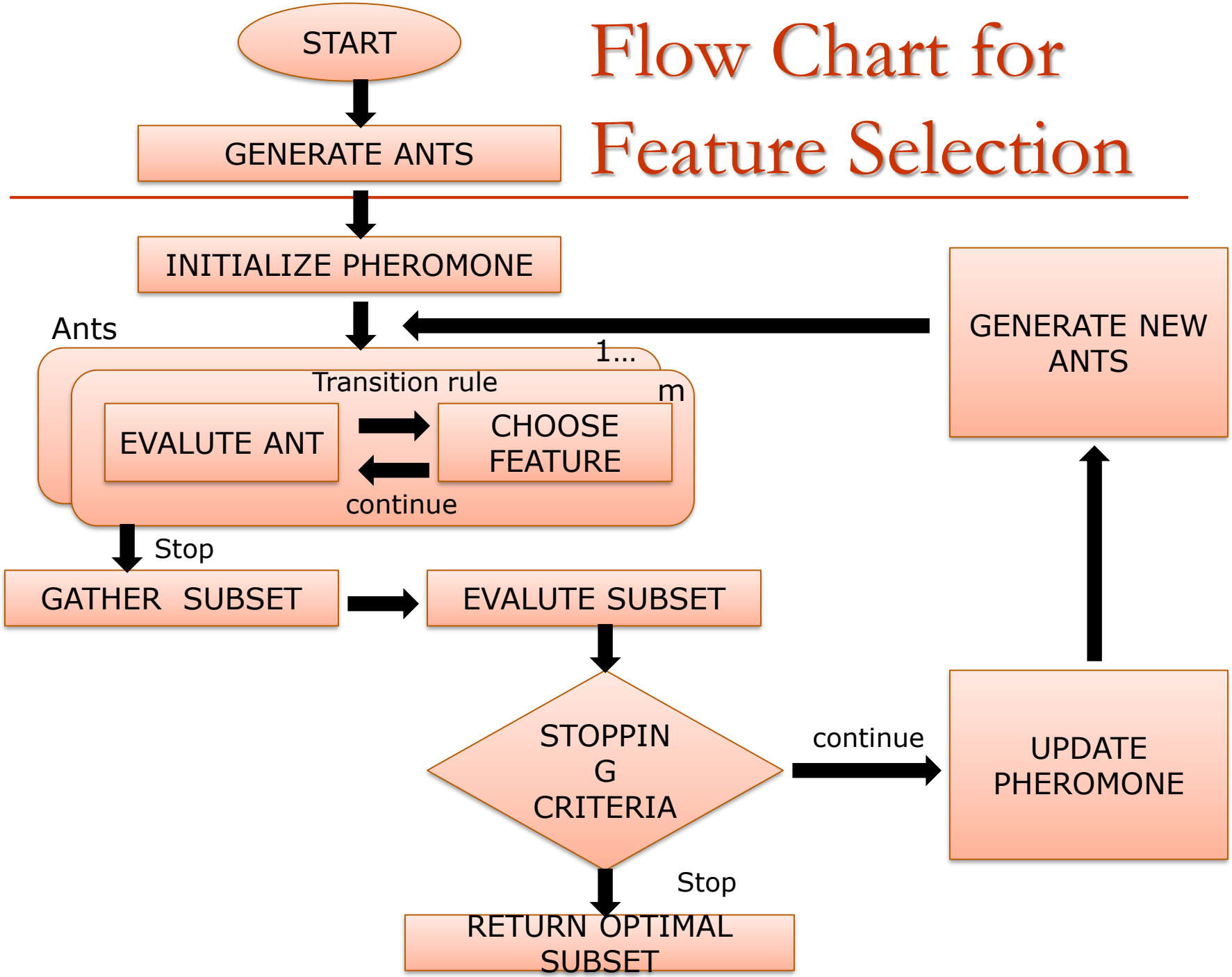
Step 4: Check the stopping criterion. If meet exist and move on step 6 otherwise continue with step 2.

Step 5: Pheromone updating

Step 5.1: Decrease pheromone concentrations of nodes then, all ants deposit the quantity of pheromone on graph. Finally, allow the best ant to deposit additional pheromone on nodes.

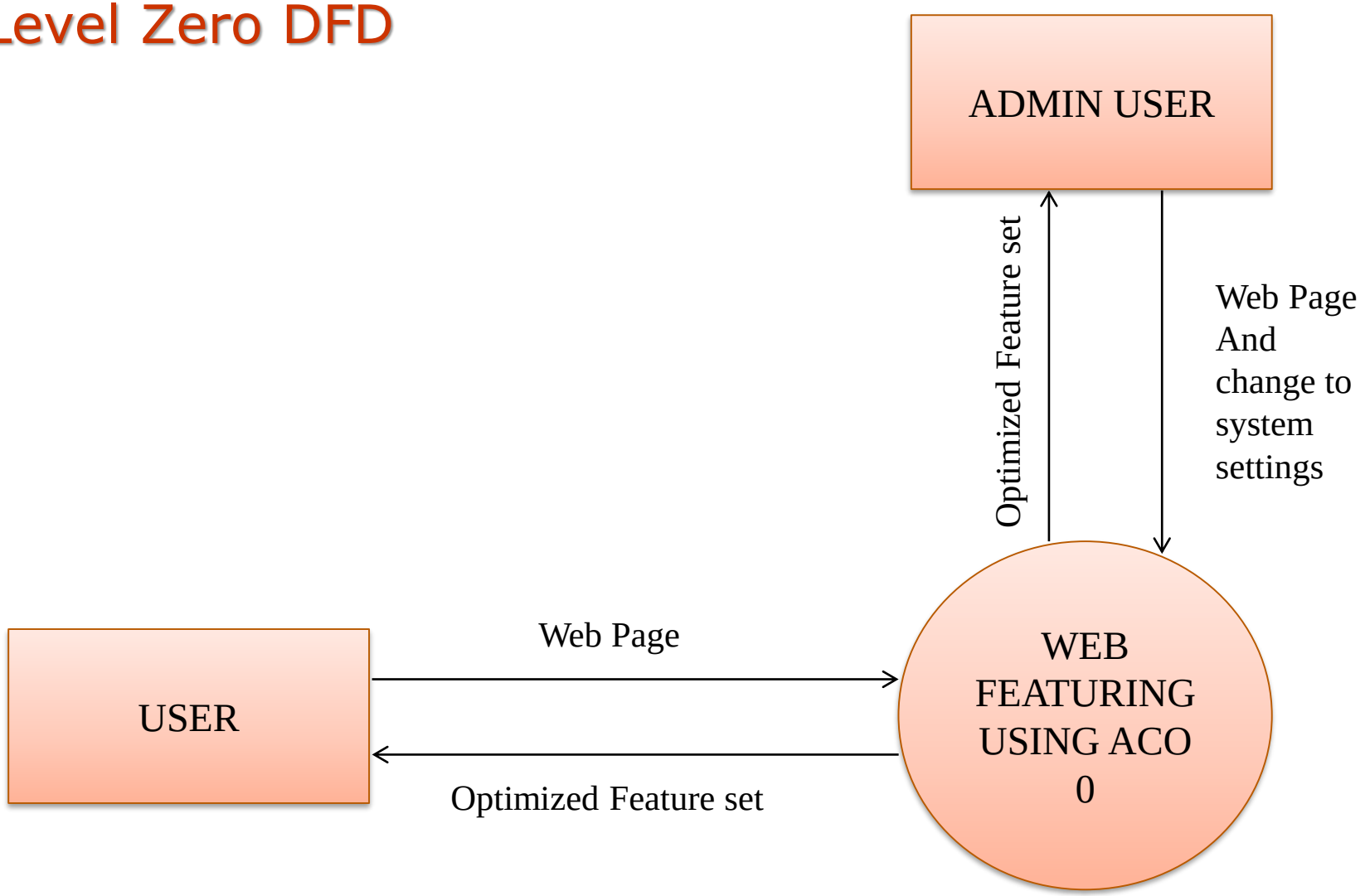
Step 6: Return the optimal feature set stored in the global list

Flow Chart for Feature Selection



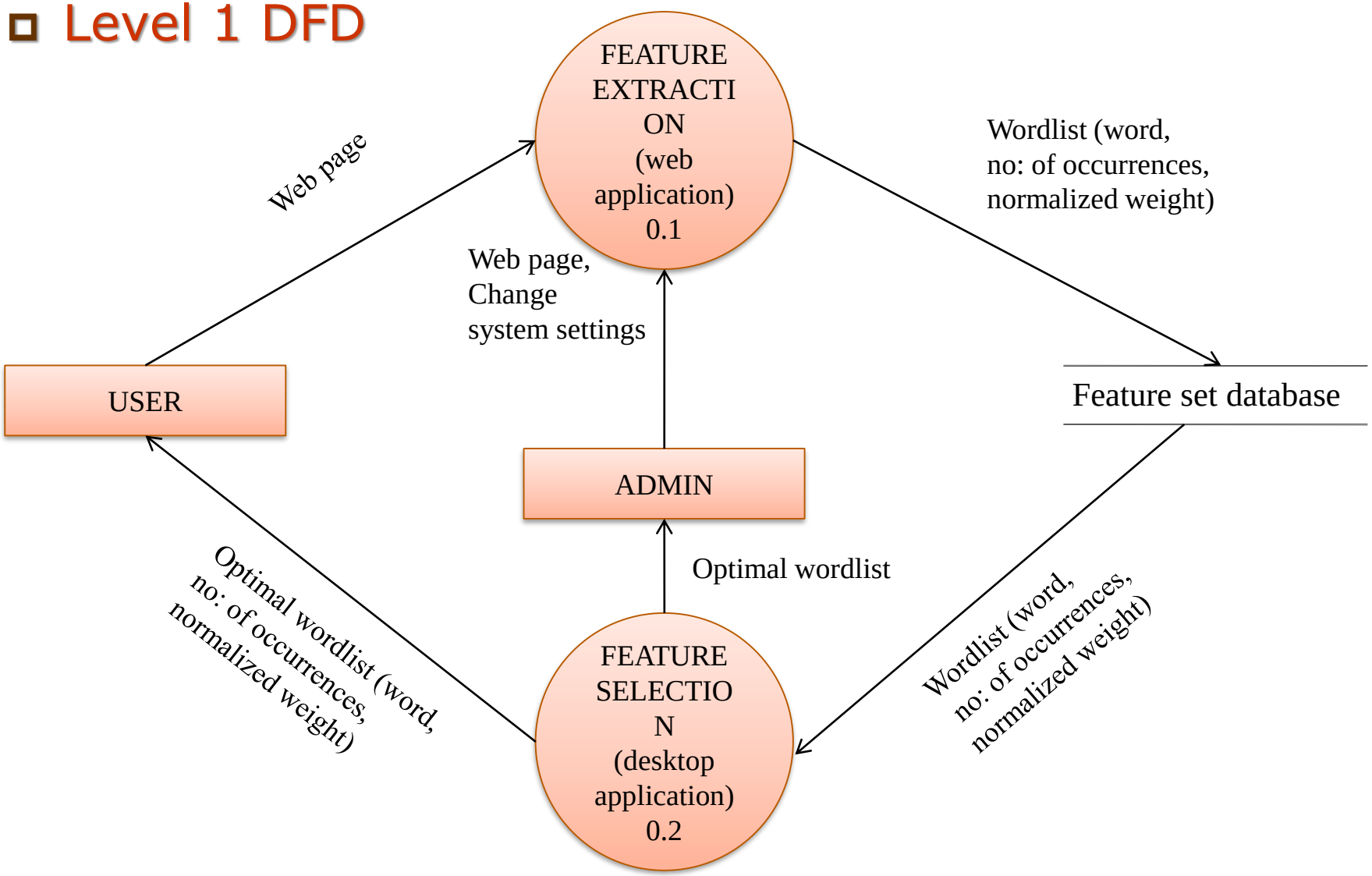
System design : Data Flow Diagram

- Level Zero DFD



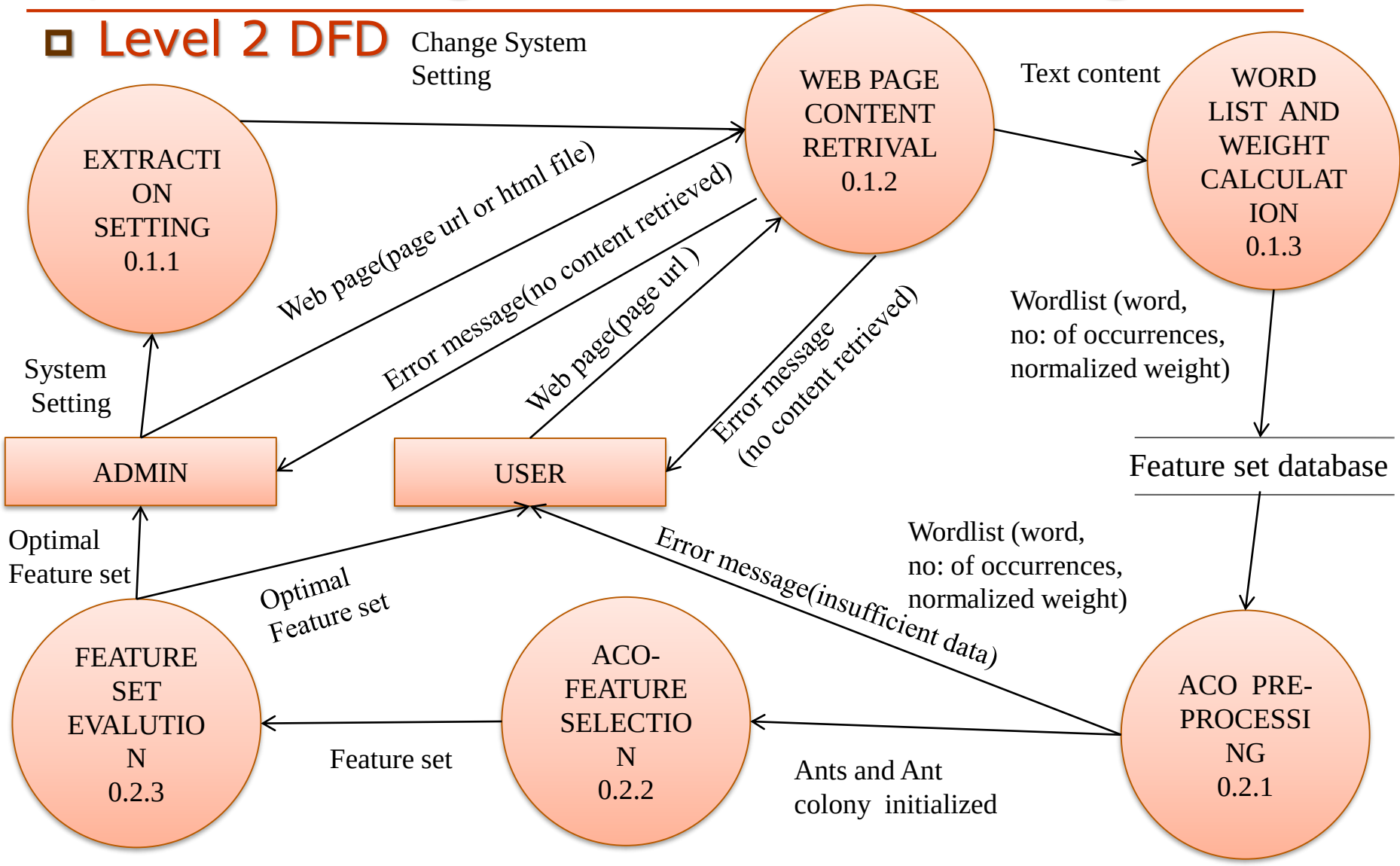
System design : Data Flow Diagram

- Level 1 DFD

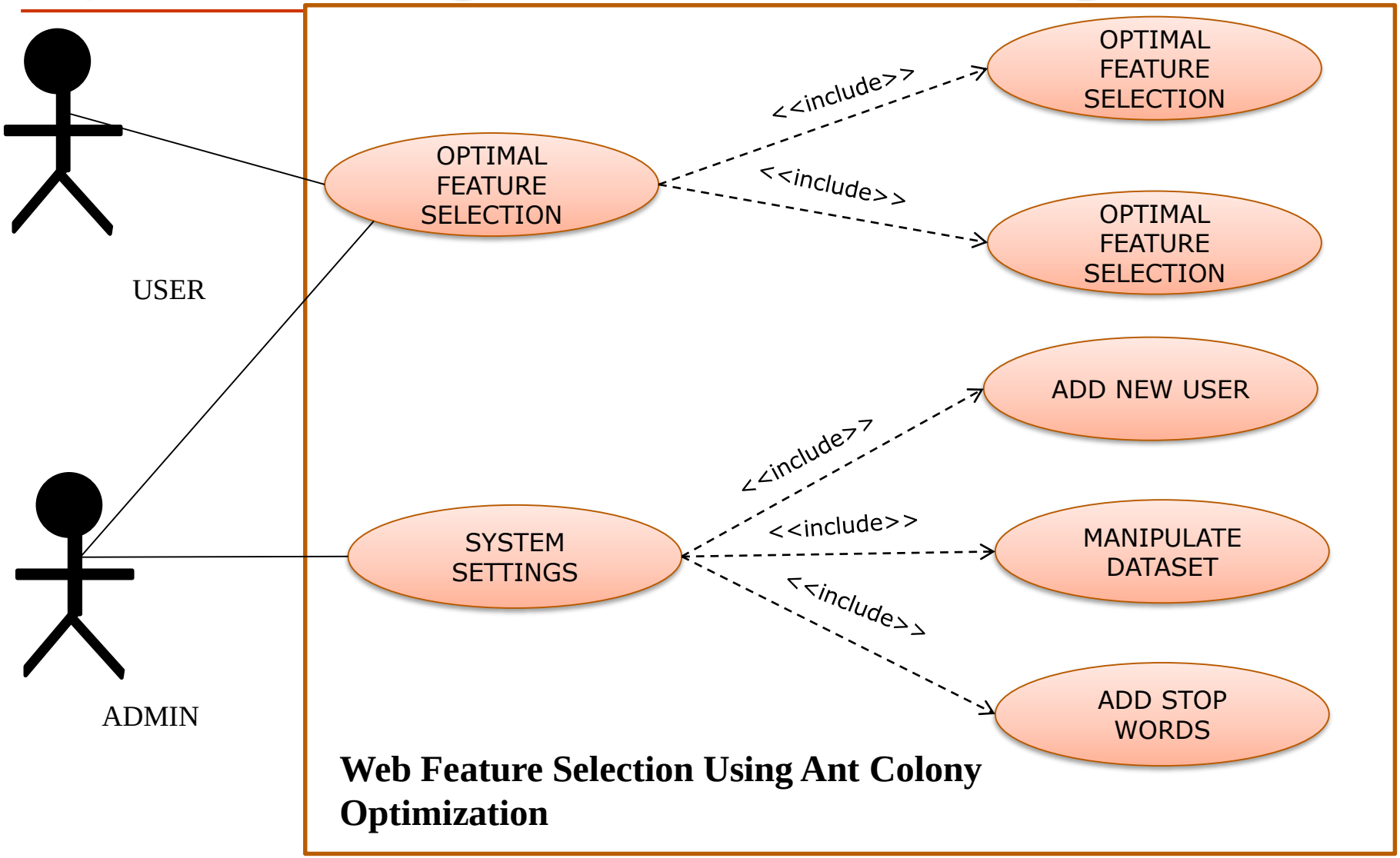


System design : Data Flow Diagram

- Level 2 DFD



System design : Use Case Diagram



Implementation Details and Related Docs

- SRS-Software Requirement Specification
- Overview of User interfaces

Conclusion

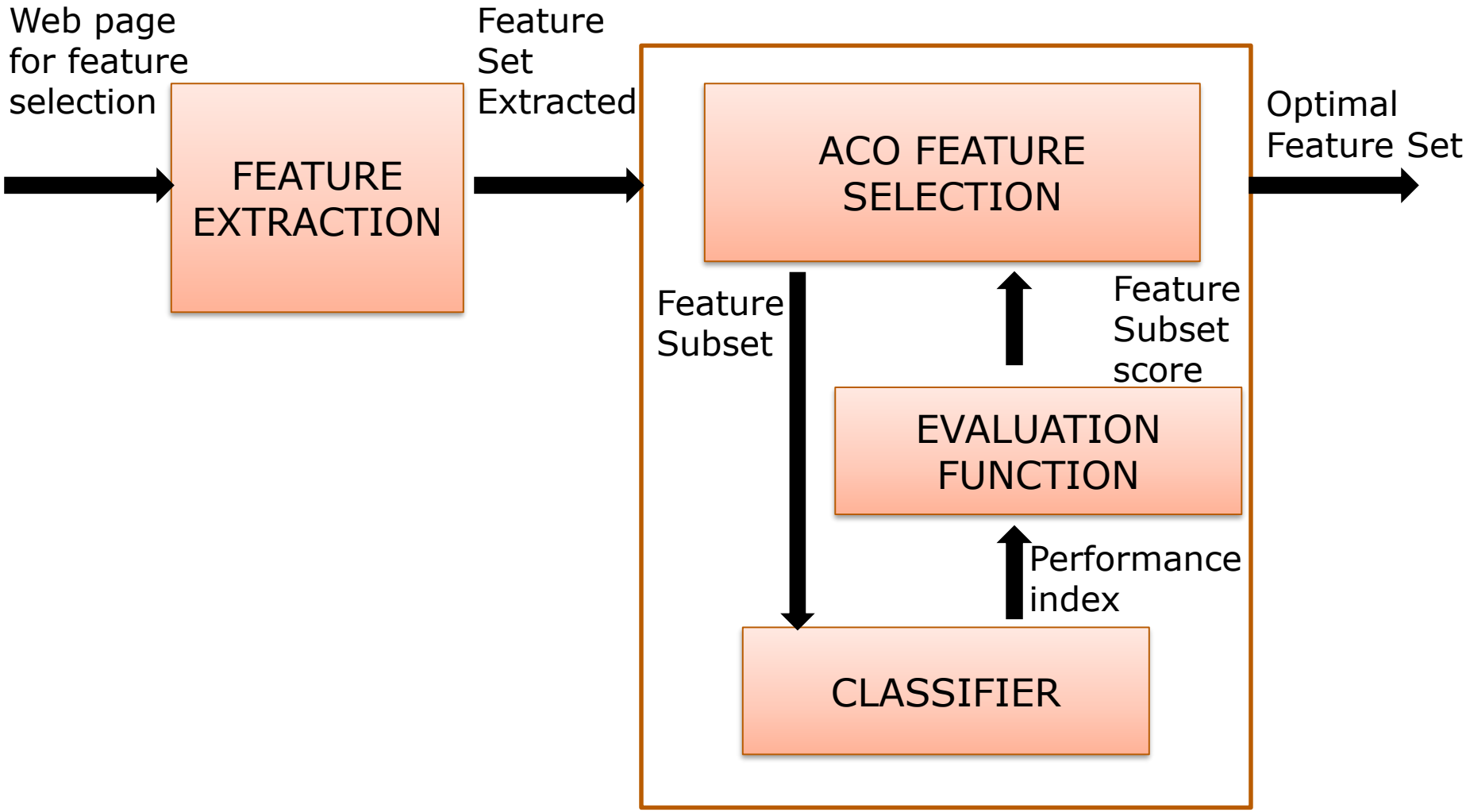
- The project WEB PAGE FEATURE SLECTION USING ANT COLONY OPTIMIZATION is intended to generate an optimal feature set.

- Web page Categorization is emerging technology today the project aim is to provide an optimized feature set for a classification system and henceforth increasing the classification system performance.

Future work

- As a future enhancement of system we would like to extended the system to a Web Page Classification System with a help of an hybrid algorithm, that is combination of ACO and Mutual Information.
- Which would allow the feature selection system to select more optimized feature set on the bases of performance index from the classifier and try to find mutual information among the feature and classified class in between each iteration of feature selection system.

Overview of the Enhanced System



References

1.M.Dorigo,Ant colony optimization webpage
http://iridia.ulb.ac.be/mdorigo/ACO/ACO.html.

2. J.L. Bentley (1992) Fast algorithms for geometric travelling salesman problem, *ORSA Journal on Computing, 4:387-411*

3. C. Blum, M. Sampels (2002) Ant colony optimization for FOP shop scheduling: *Evolutionary Computation, Honolulu, USA, pp 1558-1563*

4.A. Colorni, M. Dorigo, V. Maniezzo (1991) Distributed optimization by ant colonies
In *Proceedings of ECAL'91 European Conference on Artificial Life, Elsevier Publishing*

5. O. Cordon, I. Fernandez de Viana, F. Herrera, L. Moreno (2000) A new ACO model
In *Proceedings of ANTS'2000 - From Ant Colonies to Artificial Ants: Second International Workshop*

6. D. Camara, A.A.F. Loureiro (2000) A GPS/ant-like routing algorithm for ad hoc networks,
In *Proceeding of 2000 IEEE Wireless Communications and Networking Conference,* Chicago, USA, 3:1232-1236

Questions ?

Thank you !