

Bộ Giáo Dục & Đào Tạo  
Trường Đại Học Công Nghệ TP.HCM-HUTECH



Môn học: **KHAI THÁC THÔNG TIN**

Mã đề tài: **05**

Tên đề tài: **Nghiên cứu về các thuật toán học để xếp hạng (learn to rank – L2R) trong ngữ cảnh áp dụng các mô hình học máy cho bài toán xếp hạng các tài liệu trả về của một hệ thống truy hồi thông tin**

## **1. Nội dung thực hiện**

Nội dung của đề tài gồm các yêu cầu sau:

- **Về mặt lý thuyết:**

- Học viên trình bày các tìm hiểu và khái quát về hướng tiếp cận L2R trong lĩnh vực truy hồi thông tin.
- Học viên mô tả về các bài toán L2R (bài toán hướng dẫn giải quyết vấn đề gì trong truy hồi thông tin? Cách tiếp cận của các nhóm nghiên cứu ra sao? v.v.).
- Học viên trình bày & mô tả về các hướng tiếp cận: pointwise, pairwise và listwise trong việc xác định mục tiêu học và huấn luyện các mô hình L2R (bao gồm loss function và cơ chế tối ưu hóa - optimization).
- Học viên trình bày về các thuật toán L2R phổ biến, bao gồm:
  - RankNet [1]
  - LambdaRank [2]
  - LambdaMART [3]

- **Về mặt xây dựng ứng dụng:**

- Học viên có thể lựa chọn và cài đặt một trong các thuật toán L2R đã trình bày bên trên.
- Có thể cài đặt trực tiếp bằng Google Colab hoặc tại máy tính cá nhân – sau đó chạy demo – giải thích về các kết quả và các thức thực hiện của chương trình.

## **2. Sản phẩm yêu cầu**

Đề tài yêu cầu các sản phẩm sau:

- **Thuyết minh đề tài:** nộp dạng bài tiểu luận từ **20-25 trang**. Học viên nộp bài dưới cả 02 hình thức, gồm [**Bản mềm**]: Nộp trên hệ thống hỗ trợ học tập E-

learning HUTECH (<http://e-graduate.hutech.edu.vn/portal>) và [**Bản cứng** – 2 bản]: Nộp cho Giảng viên phụ trách học phần. Nội dung trình bày trong bài thuyết minh:

- Lý thuyết nền tảng L2R.
- Minh họa & demo về một trong các thuật toán L2R được chọn bên trên.
- **Thuyết trình trước lớp:** có slides – thời gian thuyết trình từ 30-45 phút.
- Về mặt cài đặt thử nghiệm thuật toán - có demo.

### 3. Hướng dẫn

- Trong đề tài này học viên chú trọng vào việc trình bày nội dung lý thuyết và các ứng dụng của L2R trong lĩnh vực truy hồi thông tin.
- Đối với việc cài đặt và huấn luyện mô hình L2R học viên cần trình bày được các hiểu biết về cơ chế hoạt động của thuật toán – bao gồm dữ liệu đầu vào của mô hình là gì – kết quả đầu ra là gì, v.v.

### 4. Tài liệu tham khảo chính

- [1] Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., & Hullender, G., "Learning to rank using gradient descent," in *ICML*, 2005.
- [2] Burges, C., Ragno, R., & Le, Q., "Learning to rank with nonsmooth cost functions," in *NeurIPS*, 2006.
- [3] Burges, C. J., "From ranknet to lambdarank to lambdamart: An overview," *Learning*, vol. 11, no. 23-581, p. 81, 2010.

### 5. Các tài liệu tham khảo thêm

- [https://everdark.github.io/k9/notebooks/ml/learning\\_to\\_rank/learning\\_to\\_rank.html](https://everdark.github.io/k9/notebooks/ml/learning_to_rank/learning_to_rank.html)
- <https://towardsdatascience.com/introduction-to-ranking-algorithms-4e4639d65b8>
- <https://www.youtube.com/watch?v=YroewVVp7SM>
- <https://practicaldatascience.co.uk/machine-learning/a-quick-guide-to-learning-to-rank-models>