CAPSTONE PROJECT |MINUTES

Meeting date | time 1/19/2016 1:00 PM | Meeting location FPT University R.503R

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | Meeting called by | Mr. LamPT | | Type of meeting | Introduction | | Facilitator | None | | Note taker | HoangDH60539 | | Timekeeper | None | |  |  | | |  | | --- | | Attendees  AnhPNSE90158  PhucVMSE60731  KhangNCSE61043  HoangDH60539 | |  | |

# Agenda topics

Time allotted | 9h00 | Agenda topic Review Report, Prototype | Presenter Anhpnse90158

-Review công việc tuần trước:

* Tìm hiểu về công nghệ N­Gram.
* Để phần mềm đáp ứng tốt nhu cầu người dùng cần có chức năng set thời gian hiệu lực của văn bản.
* Thực hiện cuộc tìm hiểu về cách thức tìm một văn bản của nhân viên trong một cơ quan, tổ chức.
* Sử dụng SQL 2008.

-Thầy Lâm yêu cầu:

* Mô tả thuật toán N-Gram bằng follow chart.
* Nghiên cứu Extraction Proprocessing
* Tìm hiểu thuật toán tương tự N-gram
* Convert file HTML và CHM qua text
* Demo validate bằng cách sử dụng cái dữ liệu từ các trang như vn-express … để sinh ra các keyword có giống với các keyword có sẵn không. Nếu giống đến trên 60% là đạt.
* Sau khi convert sang text cần có một bước tiền xử lý văn vản (Preprocessing) trước khi chay thuật toán N-gram
* Đưa Preprocessing và thuật toán N-gram thành 1 component modunt trong dó có các parameter, các parameter này được đặt trong 1 file config có thể edit được(calibrate).
* Các key word nên có 2 cách chiêt suất.( Manual và Algorithms)