**Công việc tuần 2 của nhóm 8 – lớp OOP 403 thầy Đạt.**

**Mục đích công việc tuần này:**

**Trích rút dữ liệu từ web, nhận biết nó là thực thể gì và in thông tin ra như trong file pdf bài tập lớn của thầy**

1. **Trích rút dữ liệu từ web: Tạo 1 public class tên: CrawData trong package: com.gr8.crawdata  
   Constructor gồm có 1 thuộc tính: path(public String). Các phương thức:**

**+ getTitle(): trả về title của bài báo**

**+ getEntity(): trả về 1 mảng các String là các thực thể trong đoạn title đó.  
+ getText(): trả về String là đoạn văn giữa, hoặc ngay phía sau các đối tượng trong title đó.  
+ getSentence(): trả về 1 array các đối tượng Sentence (entity(String[]), relation(String) ). -riêng phần này tôi muốn nghĩ lại là 1 cái title có thể có nhiều câu, nên tôi đã sửa 1 chút kết quả trả về ở buổi off**

1. **Tạo một database gồm các bảng:**
   1. **Các bảng thực thể: PERSON, LOCATION,….**
      1. **Chứa khóa chính: id**
      2. **Chứa các field là các thuộc tính bắt buộc như: Name, description, Source, Time**
   2. **Bảng quan hệ: RELATION**
      1. **Khóa chính: id**
      2. **Chứa các field khác: Name, Description (lúc đầu team định làm thêm mấy cái cột mang giá trị Boolean kiểu: Oganization\_Event để biết rằng quan hệ ấy chỉ có thể có ở đó thôi, nhưng mà thôi khỏi cần nữa, mình có thể liên kết id).**
   3. **Làm một cách nào đó để có data vào các bảng =))**
   4. **Tạo một đối tượng: Sentence trong package: com.gr8.object, gồm 2 thuộc tính: entity(Public String[]) và relation(Public String)**
2. **Tạo 1 public class tên: ProcessData, với Constructor bao gồm 2 thuộc tính là: (listSentence( public Sentence[]), dataBase(cái này các bạn lúc đầu có thể để file txt nếu cần test, nhưng khuyến khích là nên đổi sang file .db sớm vì việc tạo bảng cũng không lâu lắm))**
   1. **Riêng về phần này t sẽ không thể hướng dẫn chi tiết cần làm những gì, các thành viên nhận công việc này sẽ tạo ra các phương thức để:**
      1. **Nhận biết các Entity có trong các câu kia thuộc loại gì bằng cách so sánh với dữ liệu có sẵn, tìm mô tả của chúng (thử lấy mô tả bằng cách của Vũ là lên wifi đọc thông tin)**
      2. **Nhận biết quan hệ giữa các Entity**
      3. **Bắt buộc có 1 hàm getResult(): trả về 1 String in toàn bộ các thứ như thầy hướng dẫn trong file đề tài**
   2. **Các thành viên nhận phần này sớm update ý tưởng giải quyết bài toán này cho mình sau khi họp bàn với nhau, ghi lại dưới file word, tên là: processdata.docx update vào project của nhóm trong package: com.gr8.work.**
3. **Phân chia công việc:**

**Dựa vào kết quả buổi off, t sẽ phân chia công việc, tuy nhiên các công việc chưa thể định hình chính xác và chi tiết, nên deadline sẽ bàn thêm với từng cá nhân nhận việc để ghi lại vào file này sau:**

* **1. Trích rút dữ liệu từ web: Minh**
* **2. Tạo bảng, tạo đối tượng Sentence: Vương, Vũ – deadline: 22h chủ nhật ngày 28/10**
* **3. Xử lí dữ liệu: Phúc, Vũ, Tuấn Anh, Vương. Nhóm này sẽ có Phúc là người phụ trách, nhận việc tổ chức họp riêng, bàn bạc trao đổi đưa ra các giải pháp để lấy kết quả cuối cùng cho công việc, sau đó Tuấn Anh sẽ là người tạo file word như t đã nói ở trên cho vào project của nhóm và nhớ là up lên github nhé. Riêng phần này deadline cần bàn rất nhiều, nhưng cố gắng phải qua tuần 10 là xong.  
  (Tôi sẽ cố gắng join vào công việc số 3 để giúp đỡ ae một chút về nhân lực + phần tổ chức, định hướng công việc)**

**Anh em còn gì thắc mắc về công việc, hãy cmt trực tiếp dưới bài viết đính kèm file này trong nhóm fb nhé!**

***Chúc anh em làm việc hiệu quả, năng suất, sản phầm đạt chất lượng cao, quay tay vừa phải giữ sức khỏe nhé!!!***