Nguyễn Tiến Thành. Bm Công nghệ phần mềm

sdt: 0987768886 hoặc thanhnt.fit@gmail.com

LT: Ninh Mạnh Dũng, sdt: 0987845668,

dunglunkl0508@gmail.com

Chính sách:

+ Sẽ có điểm danh, buổi cuối sẽ tính số ngày đi trung bình của lớp. Chẳng hạn lớp có 3 bạn đi với số ngày là 3, 4, 5 => trung bình là 4. Bạn nào đi ít hơn số trung bình này 30% => bị trừ 1.

Bạn đi 3 ngày => ít hơn 4-3/ 4 = 25% => vẫn không bị trừ. Tất nhiên là trừ vào quá trình

+ Nếu trả lời được 1 câu hỏi => sẽ được cộng điểm vào QT. Chú ý rằng nếu một bạn đã được cộng N điểm thì nếu muốn cộng thêm => bạn đó sẽ phải trả lời được câu hỏi khó hơn các câu trước.

+ Các em sẽ chấm BTL => lấy đó làm điểm CK. Điểm QT = CK cộng với điểm cộng (nếu có) và trừ đi điểm phạt (nếu đi ít hơn 30% số ngày trung bình).

+ Nhóm làm BTL có tối đa 2 người. Nếu lớp có số lẻ => chỉ có duy nhất 1 nhóm có 3 người. Khuyến khích nhóm chỉ 1 người.

+ Ngày chấm BTL sẽ là ngày cuối cùng của môn học.

Ngày 1

Giới thiệu BTL:

Viết chương trình Java có chức năng như sau:

+ Có thể cho người dùng chọn file để đọc. Chỉ cần xử lý file .txt thôi. Bắt buộc file phải có tiếng Việt

+ Sẽ hiện ra tất cả các từ có khả năng sai chính tả trong file kia. Ví dụ: tăp, bob,

+ Hệ thống phải nói rõ từ sai chính tả đó là sai theo quy luật nào?

+ Có thể một số từ tên riêng, tên viết tắt tổ chức trong văn bản => ta cần chức năng đưa những từ đó vào danh sách các ngoại lệ.

+ Có thể search từ hoặc cụm từ trong file văn bản.

+ Để đảm bảo tính bền vững, chương trình phải cho phép xử lý với file lớn và nhiều file cùng một lúc. Một file không cần quá lớn < 1MB.

ĐỀ XUẤT CÁC LUẬT KIỂM TRA CHÍNH TẢ TIẾNG VIỆT

1) Có một số phụ âm chỉ được phép đứng đầu từ: q, s, d, đ, k, l, x, v, b

2) Một số chữ trên bàn phím không xuất hiện trong tiếng Việt: w, z, j, f

3)Thực ra luật 2 có thể mở rộng thành luật 3 như sau. Trong tiếng Việt chỉ có các chữ cái viết thường sau: qrtpsd đ g h k l x c v b n m, aeiuoy, â ấ ă ắ ơ ư ...

string VietnameseCharacters = "qertyuiopasdghklxcvbnm"

+ "èéẹẻẽ" + "ê" + "ềếệểễ"

+ "ùúụủũ" + "ư" + "ừứựửữ"

+ "ìíịỉĩ"

+ "òóọỏõ" + "ô" + "ồốộổỗ" + "ơ" + "ờớợởỡ"

+ "àáạảã" + "â" + "ầấậẩẫ" + "ă" + "ằắặẳẵ"

+ "đ"

+ "ỳýỵỷỹ"

;

4)

Chú ý về việc tiền xử lý (pre-process): sau khi đọc file và trước khi xử lý kiểm tra chính tả của từ => thì ta cần lọc bỏ các ký tự đặc biệt, các con số trong văn bản.

A! Môn đó không dễ... => ta phải lọc bỏ các dấu . và !

A!Mèo đen kìa!!! => A và Mèo là 2 từ

Chú ý thêm là xử lý bộ Kim Dung.

Quay lại với luật 4. Một từ trong tiếng Việt có tối thiểu 1 nguyên âm (tính cả có dấu) và y là nguyên âm

5) Một từ tiếng Việt tối đa là 3 nguyên âm

6) Các nguyên âm trong từ phải đứng cạnh nhau, không có phụ âm chen vào giữa

7) Một từ tiếng Việt có tối đa 5 phụ âm: nghiêng

8) Nếu một từ có phụ âm c đứng đầu thì phụ âm sau nó (nếu có) phải là H.

9) Luật 8 có thể tổng quát như sau: Nếu một từ có 2 phụ âm cạnh nhau thì các phụ âm đó phải là: tr, th, ph, gh, kh, ch, nh, ng,

10) Nếu một từ có 3 phụ âm cạnh nhau => đó phải là ngh

11) Trong tiếng Việt, không phải nguyên âm nào cũng đứng trước t, c, p, ch được. Hãy liệt kê các nguyên âm được phép đứng trước chúng. => Chỉ có các từ có dấu sắc hoặc dấu nặng được phép.

~~12) Trong tiếng Việt có một số phụ âm không được đứng cuối từ: q v b d l k s x r đ~~

13) Có một số phụ âm được phép đứng cuối nhưng nó phải đi kèm với một phụ âm khác. Hãy xác định các cặp này: nh, ch, ng,

→ chu yeu chua xet den chieu dai xau

14) ~~Có những phụ âm đứng cuối nhưng trước nó không được phép chứa phụ âm khác: n m t p c~~

Có một số nguyên âm không được đi sau "a" để tạo thành cặp nguyên âm, hãy liệt kê chúng: tất cả những nguyên âm nào có dấu thì không được đi sau "a", aa, aô, aê, aư, aơ, aă, aâ, ae,

15) Chỉ có một số nguyên âm được phép đứng đằng sau "á" để tạo thành cặp nguyên âm. Hãy liệt kê: ái, áu, áo, áy

16) Chỉ có một số nguyên âm được phép đứng đằng sau "ã" để tạo thành cặp nguyên âm. Hãy liệt kê: ãi, ão, ãy

17) Chỉ có một số nguyên âm được phép đứng đằng sau "ạ" để tạo thành cặp nguyên âm. Hãy liệt kê: ại, ạo, ạy

18) Chỉ có một số nguyên âm được phép đứng đằng sau "ả" để tạo thành cặp nguyên âm. Hãy liệt kê: ải, ảo, ảy.

19) Nếu có "ảu" thì từ này bắt buộc phải nằm trong từ "nhảu"

19B) Chỉ có một số nguyên âm được phép đứng đằng sau "à" để tạo thành cặp nguyên âm. Hãy liệt kê: ài, àu, ào, ày.

20) Chỉ có một số nguyên âm được phép đứng đằng sau "â" để tạo thành cặp nguyên âm. Hãy liệt kê: âu, ây

21) Chỉ có một số nguyên âm được phép đứng đằng sau "ấ" để tạo thành cặp nguyên âm. Hãy liệt kê: ấu, ấy

22) Chỉ có một số nguyên âm được phép đứng đằng sau "ẩ" để tạo thành cặp nguyên âm. Hãy liệt kê: ẩu, ẩy

23) Chỉ có một số nguyên âm được phép đứng đằng sau "ẫ" để tạo thành cặp nguyên âm. Hãy liệt kê: ẫu, ẫy

24) Chỉ có một số nguyên âm được phép đứng đằng sau "ậ" để tạo thành cặp nguyên âm. Hãy liệt kê: ậu, ậy

25) Chỉ có một số nguyên âm được phép đứng đằng sau "ầ" để tạo thành cặp nguyên âm. Hãy liệt kê: ầu, ầy

26) KHÔNG nguyên âm được phép đứng đằng sau "ă" để tạo thành cặp nguyên âm.

27) Cũng tương tự như thế, không có nguyên âm nào đằng sau "ằ", "ắ", "ẵ", "ẳ", "ặ"

28) Các nguyên âm được phép sau nguyên âm "i": ia, iu, iê, iế, iệ, iể, iễ, iề,

29) Nguyên âm i được phép đi cùng "ữ", nhưng đó phải là từ "giữ"

29B) Nguyên âm i được phép đi cùng "ó", "ố", "ỗ", "ọ", "ỏ", "ò", "á", "ấ", "ặ" nhưng đó phải là từ "gió" , "giống", "giỗ", "giọng", "giỏ", "giò", "giá", "giấc", "giặc",

30) Các nguyên âm được phép sau nguyên âm "ì": ìu, ìa,

31) Các nguyên âm được phép sau nguyên âm "ỉ": ỉu, mỉa,

32) Các nguyên âm được phép sau nguyên âm "ị": ịu, ịa,

33) Các nguyên âm được phép sau nguyên âm "ĩ": ĩu, ĩa,

34) Các nguyên âm được phép sau nguyên âm "í": íu, ía,

35) Các nguyên âm được phép sau "e", "é", "è", "ẻ", "ẹ", "ẽ" thì chỉ có: "o"

36) Các nguyên âm được phép sau "ê", "ế", "ề", "ệ", thì chỉ có: "u"

37) Nguyên âm "ễu" chỉ được phép đi trong từ "Tễu" và "phễu"

38) Nguyên âm "ể" không có nguyên âm khác đi sau.

39) Nguyên âm sau "o" để tạo thành cặp nguyên âm: "oa", "oi", "oe",

39B) Cặp nguyên âm "oo" chỉ có thể thuộc về xoong, boong,

40) Ta có thể tổng quát hóa lên rằng các từ "ó", "ò", "ỏ", "ọ" có các nguyên âm "e", "i", "a" đằng sau được.

41) Nguyên âm "ơ", "ớ", "ờ", "ở", "ợ" thì chỉ có đằng sau là nguyên âm "i"

41B) Nguyên âm "ỡ" nếu đi sau nó là "i" thì từ đó phải là "hỡi"

42) Nguyên âm "ô", "ố", "ồ", "ổ", "ộ", "ỗ" thì chỉ có đằng sau là nguyên âm "i"

43) Nguyên âm "u" đi đằng sau sẽ là các nguyên âm "i", "a", "y", "ê", "ở", "ế"

43B) Nguyên âm "ý", "e" và "ơ", "á", "ả", "ạ" "à" cũng đi sau "u" được nhưng đó phải là các từ "quý" "que", quơ, "huơ", "quá" quả, quạ, quà

44) Nguyên âm "ú", "ù", "ủ", "ụ" đi đằng sau sẽ là các nguyên âm "i", "a", "y",

45) Nguyên âm "ũ" đi sau là y được. Nhưng đi sau là "a" với các trường hợp "đũa", "dũa"

45B) Nguyên âm "ư" đứng đầu và nguyên âm khác đi sau....

46) Các nguyên âm ba có thể xuất hiện được trong tiếng Việt mà nguyên âm đầu tiên là "i" là: iữa, iễu, iếu, iều, iểu, iêu, iệu, iai, iải, iới, iỏi, iươ, iườ, iòi,

47) Các nguyên âm ba có thể xuất hiện được trong tiếng Việt mà nguyên âm đầu tiên là "u" gồm: uôi, uối, uồi, uỗi, uội, uổi, uai (phải đi với q), quái, quại, quay, quẩy, quậy, quầy, uây (đi được với q và kh), uấy (đi kèm với q và kh), quẫy, uya, uỷu (đi với kh)

48) Các nguyên âm ba có thể xuất hiện được trong tiếng Việt mà nguyên âm đầu tiên là "ư" gồm: ươi, ười, ưới, ưỡi, ưởi, ượi, ượu, ườu, ươu, ướu,

49) Các nguyên âm ba có thể xuất hiện được trong tiếng Việt mà nguyên âm đầu tiên là "o" gồm: oai, oài, oái, oải, oại (oãi thì gần như không thấy), oay (chỉ đi với x), oáy (đi với x và ng), oẹo (đi với ng),

============================================

Cài đặt luật

1) Bố trí lớp:

Để cài đặt các luật (vốn khá nhiều) thì ta cần bố trí sao cho các luật ấy không bị kết dính với nhau - Tức là mã nguồn của luật này hầu như không ảnh hưởng đến mã nguồn của luật khác.

Mà vẫn phải thỏa mãn được yêu cầu đơn giản, gọn nhẹ, tái sử dụng được của OOP.

abstract class Rule {

//Có một số các phương thức sau (có thể thêm pt khác)

public abstract boolean IsValid(string x) ;

public abstract void ShowError(string x) ;

}

Các lớp con sẽ kế thừa Rule và cài đặt từng luật một trong số các luật kia.

Bên cạnh đó, các đối tượng của các lớp con đó sẽ được gom vào một mảng, mảng đó sẽ nằm trong lớp Program (hoặc một lớp nào đó tùy tên đặt)

List<Rule> allOfRules = new ArrayList<Rule>( );

allOfRules.add(new Rule1( ) );

...

allOfRules.add(new RuleN( ) );

//Có một đoạn mã để tách ra các từ trong file.

//Các từ sẽ được lưu vào mảng arrayWords

foreach(string word : arrayWords)

{ foreach(Rule rule : alllOfRules)

{ if( ! rule.IsValid(word))

{ rule.ShowError( ); break; }

}

}

public class Program

{

public static void main(String args[])

{

string [] arrayWords = new string[]{

""

};

List<Rule> allOfRules = new ArrayList<Rule>( );

allOfRules.Add(new Rule1());

///Sẽ thêm vào các câu lệnh bổ sung luật

**for(string word : arrayWords)**

**{ for(Rule rule : alllOfRules)**

{ if( ! rule.IsValid(word))

{ System.out.print("Tu sai la: " + word);

rule.ShowError( ); break;

}

}

}

}

}

class Rule1 extends Rule{

public void ShowError()

{ Sysout("Có một số phụ âm chỉ được phép đứng đầu từ: q, s, d, đ, k, l, x, v, b");

}

public boolean IsValid(string x)

{ string firstConsonants = "q s d đ k l x v b";

for(int j = 1; j < x.length( ); j++)

{

if(firstConsonants.contains(x.charAt(j) + ""))

{

return false;

}

}

return true;

}

}

luật 3 như sau. Trong tiếng Việt chỉ có các chữ cái viết thường sau: qrtpsd đ g h k l x c v b n m, aeiuoy, â ấ ă ắ ơ ư ...

string VietnameseCharacters = "qertyuiopasdghklxcvbnm"

+ "èéẹẻẽ" + "ê" + "ềếệểễ"

+ "ùúụủũ" + "ư" + "ừứựửữ"

+ "ìíịỉĩ"

+ "òóọỏõ" + "ô" + "ồốộổỗ" + "ơ" + "ờớợởỡ"

+ "àáạảã" + "â" + "ầấậẩẫ" + "ă" + "ằắặẳẵ"

+ "đ"

+ "ỳýỵỷỹ"

;