

:: Projet Cindy ::

Chapitre I : Traitement Automatique du Langage (TAL)

Algorithme de Compréhension du Langage Humain

- Comment faire comprendre un ordre à un ordinateur ?
- Comment faire comprendre une question à un ordinateur ?
- Comment faire tenir une conversation à un ordinateur ?

A)Le Raisonnement Humain

1.Introduction

L'Homme à depuis tout temps, subjugué les esprits de par son génie à comprendre et à réaliser les choses, qui parfois, pouvait se voir révéler d'utilité public. De fait, comme il l'a été prononcé dans le passé par les Humanistes : "On ne naît pas Homme on le devient", ce qui ici pourrait se voir ramener à "On ne naît pas génie, on le devient".

Pour cause, les esprits les plus créatifs et les plus ingénieux ont du au préalable de leurs accomplissements, commencer par apprendre les choses les plus basiques qui soit, et en gravant les échelons, on réussi à faire ce qu'ils ont fait.

Ainsi, l'esprit humain ne peut concevoir des choses, que grâce à des connaissances acquise dans des domaines diverses. Et de fait, les génies ont du comme à chacun, commencer par apprendre à lire, puis à compter, et ainsi de suite pour finalement aboutir à ce qu'ils sont à l'heure actuelle.

2.Compréhension

De fait nous nous retrouvons ici dans la base de connaissance même de tout esprit arrivé à maturité, à savoir, savoir lire et écrire correctement dans sa langue maternelle, correspondant ainsi à la débouché même d'une compréhension et d'une certaine maîtrise effective de cette langue.

Ainsi, lorsque nous lisons une phrase, nous analysons sans le savoir la structure et la syntaxe de la phrase, ainsi que l'orthographe des mots de cette phrase. Puis, avec cet ensemble d'analyse intuitive, nous arrivons à comprendre la phrase, et à l'interpréter à notre manière.

3. Algorithme

Soit : "J'ai besoin d'un programme pour aller sur IRC."

On Analyse :

- > J' : Pronom Personnel (1er Personne du Sujet / Diminutif de Je)
- > ai : Auxiliaire Avoir (1er Personne du Singulier / Conjugué au Présent)
- > besoin : Nom Commun Masculin (Complément d'Objet Direct de J'ai)
- > d' : Déterminant (Diminutif de De)
- > un : Déterminant (Complément Direct de D')
- > programme : Nom Commun Masculin (Complément d'Objet Direct de J'ai besoin)
- > pour : Adverbe (Précise le Pourquoi du Besoin)
- > aller : Verbe Aller (Conjugué à l'Infinitif / Complément d'Objet de Pour)
- > sur : Déterminant (Détermine le Lieu / Précise le Verbe Aller)
- > IRC : Nom Propre (Complément Circonstanciel de Lieu de Aller)

On Interprète :

- > J' : Le Sujet Parlant est Sujet de la Phrase
- > ai besoin : Le Sujet Parlant Souhaite Quelque Chose
- > d'un programme : Le Sujet Parlant Souhaite un Programme
- > pour : Le Sujet Parlant Explique/Définit Son Besoin de Programme
- > aller : Le Sujet Parlant Détermine Son Programme
- > sur IRC : Le Sujet Parlant Précise Son Besoin de Programme Avec un But Définit

D'où :

- > J'ai besoin : De Quoi ?
- > d'un programme : Pour Quoi ?
- > pour aller : Pour Aller Où ?
- > sur IRC : Il ne reste plus qu'à savoir ce qu'est IRC (Connaissance Spécifique)

Donc :

- > Cette phrase est syntaxiquement une énonciation simple.
- > Cette phrase peut devenir un ordre en fonction de l'interprétation.

B)Algorithme de Raisonnement

1.Introduction

On a vu précédemment qu'un être humain était capable avec des connaissances acquise, d'interpréter une phrase et de la comprendre. Mais qu'en est-il d'un ordinateur ? Etant donné que ce dernier n'est qu'une simple machine dont les principes formels repose sur le concept de l'Algèbre Booléen, les capacités de compréhension d'un ordinateur sont très limité (0 et 1).

Par conséquent, il va nous falloir créer un algorithme émulant le système de raisonnement de la pensée humaine, sur la fonction de compréhension du langage, pour que l'ordinateur soit en mesure de tenir une conversation, de traiter un ordre ou de répondre à des questions simples.

Soit : "J'ai besoin d'un programme pour aller sur IRC."

On Analyse :

- > J' : Pronom Personnel (1er Personne du Sujet / Diminutif de Je)
- > ai : Auxiliaire Avoir (1er Personne du Singulier / Conjugué au Présent)
- > besoin : Nom Commun Masculin (Complément d'Objet Direct de J'ai)
- > d' : Déterminant (Diminutif de De)
- > un : Déterminant (Complément Direct de D')
- > programme : Nom Commun Masculin (Complément d'Objet Direct de J'ai besoin)
- > pour : Adverbe (Précise le Pourquoi du Besoin)
- > aller : Verbe Aller (Conjugué à l'Infinitif / Complément d'Objet de Pour)
- > sur : Déterminant (Détermine le Lieu / Précise le Verbe Aller)
- > IRC : Nom Propre (Complément Circonstanciel de Lieu de Aller)
- > '.' : Le Signe de Ponctuation '.' signifira ici la fin de la phrase.

Chaque mot sera ici comparé à une BDD spécifique à chaque type de mot (Pronoms, Noms, Adverbes ...). Le Bot aura également besoin d'une BDD spécifique de connaissance et de culture générale, comme l'être humain (Sciences, Histoire, Géographie, Langues, Informations ...), et également d'une Base de Données comportementale à l'image de l'Homme (Expressions, Humeur, Attitude ...).

Le Bot devra également posséder un moteur d'auto-apprentissage pour apprendre les nouveaux mots ou structures inconnues à son répertoire. Egalement, un système pro-actif devra lui être fourni pour analyser dynamiquement les entrées utilisateur (Parser & Moteur de Raisonnement).

C)BDD Envisageable

1.BDD Spécifiques

- > Informatiques (Programmation, Piratage, OS, Internet ...)
- > Mathématiques
- > Physiques (Mécanique Classique, Mécanique Quantique ...)
- > Chimie
- > Electronique
- > Histoire Mondiale
- > Géographie
- > Mécanique Automobile
- > Sports
- > Etat d'Esprit (Humeur, Gentillesse, Répondant, Politesse ...)
- > Biologie
- > Astronomie
- > Culture Générale (Monuments, People, Consommable ...)

2.BDD de Langue

- > Noms Communs (Masculin, Féminin)
- > Noms Propres (Noms, Sigles)
- > Adverbes
- > Verbes (Action, Auxiliaire)
- > Pronoms Personnel (Sujet)
- > Déterminants (Masculin, Féminin)

D)Notes Subsidières

- > Le Bot devra prendre les Verbes comme des actions.
- > Le Bot devra pouvoir effectuer des recherches sur Internet en cas de manque de connaissance.
- > Le Bot devra être capable d'entreprendre des actions commandé par l'utilisateur.
- > Le Bot devra être capable de reconnaître les erreurs d'orthographes et de conjugaison.

E)Exemple

User > Lance un BruteForcing sur 184.52.36.35 !
Cindy > Analyse de votre demande en cour...
Cindy > Sur quel port TCP ?
User > 21
Cindy > Analyse de vorte demande en cour...
Cindy > Sur quel compte utilisateur ?

User > root
 Cindy > Analyse de votre demande en cour...
 Cindy > Lancement du BruteForcing sur 184.52.36.35:21@root en cour.
 Cindy > Le processus est cour.
 User > Merci Cindy.
 Cindy > À vos services.
 User > Estimation du temps de Travail ?
 Cindy > Approximativement 5h47mn22s.
 User > Ok. Préviens-moi lorsque tu auras terminé.
 Cindy > Très bien.

Soit : "Lance un BruteForcing sur 184.52.36.35 !"

- On compte le nombre de mots (MaxLen = 4)
- On compte le nombre de signe de ponctuation (MaxSign = 4)
- Le point d'exclamation définit une ordre au Bot

On Analyse :

- Lance : Verbe / 1er Groupe / 3e Personne du Singulier
 ->Donc User s'adresse à Cindy / On relie avec l'occurrence suivante.
- un : Déterminant / Masculin
 ->Donc on relie avec l'occurrence suivante.
 -> Précise l'action du Verbe Lancer.
- BruteForcing : Connaissance Spécifique / Informatique /
 Nom Commum Masculin
 ->Donc User ordonne à Cindy d'établir un BruteForcing.
- sur : Déterminant de Lieu
 ->On relie à l'occurrence suivante.
- 184 : Nombre Entier Positif / Appartient à l'ensemble R
 ->Qu'est-ce que 184 ? On essaye de chercher avec l'occurrence suivante.
- '.' : Signe de Ponctuation / Complète le Nombre 184
 ->C'est peut-être un nombre flottant / Fin de Phrase ?
 ->On continue de chercher.
- 52 : Nombre Entier Positif / Appartient à R
 ->C'est peut-être la valeur décimale de 184.
 ->C'est peut-être le début d'un IPv4.
- '.' : Signe de Ponctuation
 ->Fin de Phrase ? D'où le lieu est 184.52 (GPS, Fréquence ?) / IPv4 ?
- 36 : Nombre Entier Positif / Appartient à R
 ->Trois Nombres entier positif séparé par un '.'
 ->C'est une IPv4. Besoin de la suite.
- '.' : Signe de Ponctuation
 ->C'est la séparation Ipv4.

- 35 : Nombre Entier Positif / Appartient à R
 - >IPv4 = 184.52.36.35 / Définit une machine distante.
- '!' : Signe de Ponctuation
 - >Définit un ordre / Cindy doit effectuer cette action.
- > Requête en BDD Spécifique Informatique
 - Besoin du Port
 - Besoin du Nom de Compte
 - Besoin d'un Logiciel Adéquate

E)Algorithme Nécessaire

- >Moteur d'Inférence
- >Moteur d'Auto-Apprentissage
- >Logique Flou
- >Algorithme Génétique

F)Problème Potentiel

1.Contexte

Bien entendu, un des premier problème qui se verra poser est sans nul doute le contexte de la phrase. Et pour preuve, dans mes exemples précédant, le verbe utilisé 'Lancer' pourrait être, avec la façon dont il a été écrit, une lance (Comparé au Javelot) si le contexte n'était pas préciser.

Ainsi, il va falloir intégrer un autre algorithme au logiciel Cindy, afin que ce dernier puisse débloquent une situation complexe. Il va falloir faire en sorte que lorsque le contexte lui est inconue, qu'il puisse demander des précisions à l'utilisateur sur son besoin.

Mais si ce mot est suivie d'une occurrence intéressante, tel que 'un' dans mes exemples ci-dessus, alors 'Lance un' ne pourra être qu'une action à effectuer. Autrement, 'lance' tout seul pourra être interpréter comme un Javelot ou autre.

Ce problème a été résolu par un algorithme intégré à Skaaz, un Bot MSN capable de demander des précisions lorsque ce dernier n'est pas en mesure de comprendre ce qu'il a analysé.