

Millora de rendiment d'un sistema d'avaluació de traductors automàtics

Ibai Gilabert Rodríguez

9 de març de 2015

Resum

Informe de seguiment del TFG "Millora de rendiment d'un sistema d'avaluació de traductors automàtics".

Autor: Ibai Gilabert Rodríguez

Director: Jordi Turmo

Índex

1	Introducció	4
1.1	Context	4
1.2	Abast	4
2	Estat de l'Art	6
3	Planificació	7
3.1	Fase 1	7
3.2	Fase 2	7
3.3	Fase 3	7
3.4	fase puta	7
4	Anàlisi d'alternatives	8
5	Metodologia	9
6	Estat actual	10

1 Introducció

El camp de la traducció automàtica i, en definitiva, el tractament del llenguatge natural està en alça. La seva rellevància augmenta dia a dia. En la mesura que la nostra capacitat tecnologia creix els horitzons en aquesta àrea són cada vegada més i més enlluernadors.

El NLP és perse molt complexe, computacionalment parlant. Requereix de molts recursos, temps i memòria principalment, el que implica haver de disposar d'entorns d'execució d'altres prestacions.

L'objectiu d'aquest projecte és fer-ho real. És aprofitar tota la capacitat tecnològica de la que disposem de manera que faci del NLP (*Natural language processing*) una empresa més tractable.

1.1 Context

Prenem com a *baseline* un sistema plenament funcional: ASIYA[1]. ASIYA és una eina que posa a disposició un ventall molt ampli i heterogeni de mètriques d'avaluació i meta-avaluació, la majoria d'elles són usables i perfectament vàlides per a ser executades de manera singular. Així que ASIYA actua més aviat com una interfície per tota aquesta amalgama de mètriques.

En l'actualitat no hi ha cap sistema, per quantitat i qualitat, com ASIYA, com a mínim d'aquestes característiques. No obstant, el seu principal handicap són els recursos que exigeix. Tot i que algunes mètriques fan us d'eines externes com ara processadors lingüístics que, per motius de manteniment i flexibilitat no tocarem i tractarem com a caixes tancades i completament opaques, sí és cert que es deixa entreveure una certa concurrència en el processament de les dades.

Aquí és on el projecte agafa sentit. Com podem fragmentar l'execució d'ASIYA de manera que mitjançant alguna arquitectura paral·lela millorem l'ús dels recursos?

1.2 Abast

ASIYA és una eina flexible, adaptable i de fàcil integració. Això és una característica que no volem perdre com ja s'ha esmentat. Això vol dir que les tasques de manipulació de l'eina com afegir una mètrica, modificar un tipus de format, etc han de ser el més fàcils possible. Es persegueix una filosofia o idea de senzillesa en el seu ús i modificació. Adaptabilitat al cap i a la fi.

Així doncs si no volem modificar el còmput, el que ens queda es modificar les dades. Si tenim una entrada de mida N la partirem en $n \in N$ fragments

perquè ASIYA pugui executar els $n \in N$ fragments de manera paral·lela, aquesta és la idea.

Per portar-ho a terme, es plantejarà un redisseny íntegre de tot el sistema. Aquest serà en un llenguatge més adient per a les nostres necessitats (eficiència i concurrència). S'ha escollit C++ per la seva estructura de classes, tipificació i compatibilitat amb diferents arquetips de programació paral·lela que veurem més endavant.

2 Estat de l'Art

¿?

3 Planificació

S'ha produït canvis respecte la planificació inicial.

La planificació definitiva. No afecten als costos del projecte. Ens trobem en la fase 3 d'experimentació dels prototips.

3.1 Fase 1

3.2 Fase 2

3.3 Fase 3

3.4 fase puta

Inicialment es plantejava la proposta de fonamentar la implementació, pel que fa a concurrència, en GPUs, degut a la disposició d'aquest tipus de tecnologia al sistema de clustering on es preveu l'execució. Aquesta idea va ser descartada degut a ...

4 Anàlisi d'alternatives

Aprofitarem la tipologia del cluster (nodes independents), per a implementar el paral·lelisme de manera *recursiva*. La idea és que

També s'ha hagut de fer ús de llibreries de tercers a fi de solventar les mancances del llenguatge nadiu. Per exemple, boost[ref], per a l'ús d'expressions regulars i tractament de fitxers a nivell de sistema operatiu.

5 Metodologia

Continuem amb la metodologia proposada a la fita inicial. Un esquema de prototipatge basat en els següents passos.

6 Estat actual

Què hi ha fet?

Referències

- [1] <http://nlp.lsi.upc.edu/asiya/>
- [2] <https://www.boost.org>