



UNIVERSITÀ DI PADOVA
DIPARTIMENTO DI MATEMATICA
CORSO DI LAUREA MAGISTRALE IN INFORMATICA

Corso di Intelligenza Artificiale

Marco Romanelli, [1106706]

28 marzo 2017

Capitolo 1: Introduzione

1.1 Fase iniziale

La fase iniziale del progetto consiste in una prima analisi delle principali piattaforme che offrono servizi di *Cognitive Computing*, considerando vantaggi e svantaggi di ognuna. Data la diversificazione di proposte all'interno di ogni piattaforma, per l'analisi e la comparazione ci siamo focalizzati sul riconoscimento di immagini.

Procederemo nel seguente modo: nel Capitolo 2 verrà presentata una panoramica dei servizi presi in esame e seguirà poi nel Capitolo 3 l'analisi dettagliata per ogni servizio. Il Capitolo 4 contiene una sintesi sulle tariffe richieste per ogni servizio. Il Capitolo 5 riassume le conclusioni della squadra di lavoro. Il capitolo 6 concluderà con alcuni possibili sviluppi di questa analisi.

Capitolo 2: Cognitive Computing

2.1 Introduzione

2.2 Servizi disponibili

Le maggiori piattaforme per il *Cognitive Computing* sono offerte da alcune fra le maggiori aziende nell'ambito informatico e tecnologico e sono:

- Microsoft Cognitive Services[?] (Microsoft Corporation),
- Watson Developer Cloud (IBM: International Business Machines Corporation),
- Amazon Artificial Intelligence (Amazon.com, Inc)
- Google Cloud Platform (Google Inc.)

Capitolo 3: Analisi dei servizi

3.1 Microsoft Cognitive Services: Computer Vision API

3.1.1 Panoramica

Prerequisiti

- Input: dati grezzi (stream application/octet) o url;
- Formati supportati: JPEG, PNG, GIF, BMP;

- Dimensione file massima: 4 MB;
- Dimensione immagine minima: 50x50 pixel.

Le API¹ sono molteplici, a seconda dello scopo finale dell'analisi visiva.

Tagging Le API ritornano un insieme di etichette (in formato JSON) che descrivono gli oggetti presenti nell'immagine, come oggetti, esseri viventi, azioni, paesaggi; per ogni etichetta viene anche fornito il livello di *confidence* (affidabilità). I tag non sono in alcun modo organizzati fra loro e non esiste nessun tipo di ereditarietà. Nel caso un tag sia ambiguo viene fornito in aggiunta un *hint* che ne spiega il contenuto. Al momento la sola lingua supportata è l'inglese.

Classificazione L'immagine viene classificata in categorie che seguono una tassonomia con ereditarietà di tipo padre-figlio. Questa tassonomia prevede 86 categorie² e classifica gli elementi visivi in modo più o meno specifico.

Identificazione del tipo E' possibile classificare l'immagine come in bianco o nero o a colori, se è un disegno o se è del tipo *clip-art*; in quest'ultimo caso viene fornito un livello di qualità dell'immagine, compreso fra 0 e 3.

Riconoscimento volti Riconosce i volti umani e restituisce la posizione (coordinate) di questi all'interno dell'immagine, come anche età e sesso della persona.

Contenuto personalizzato Ideato per raffinare la tassonomia a 86 categorie utilizzando informazioni specifiche sul dominio. Attualmente è supportato solamente il riconoscimento dei volti delle persone famose.

Generazione di descrizioni Genera una lista di frasi (in lingua inglese) che descrivono il contenuto dell'immagine, ordinate secondo un livello di affidabilità calcolato per ogni descrizione.

Estrazione colori Identifica i colori analizzandoli in tre contesti: di sfondo, in primo piano e d'insieme; i colori sono raggruppati in 12 colori predominanti. Classifica le immagini fra in bianco e nero e a colori.

Riconoscimento contenuti non adatti ai minori Riconosce materiali pornografici e contenuti osé in generale. Può essere impostato un livello per il filtro.

Riconoscimento del testo (OCR) Rileva il testo presente nell'immagine e lo trasforma affinché sia leggibile da una macchina, ruota l'immagine se necessario per rendere il testo orizzontale e fornisce le coordinate per ogni parola. Al momento sono supportati 21 linguaggi, fra cui l'inglese, l'italiano, il francese, il tedesco e lo spagnolo.

L'accuratezza del riconoscimento dipende dalla qualità dell'immagine ed eventuali errori possono essere causati da immagini sfuocate, scrittura a mano, testo troppo piccolo, ecc.

Creazione anteprime Un'anteprima è una rappresentazione dell'immagine in scala ridotta. L'immagine viene prima analizzata e poi ritagliata secondo la "regione di interesse" (ROI); il rapporto dell'immagine (*aspect ratio*) può essere impostato secondo le proprie preferenze.

¹<https://www.microsoft.com/cognitive-services/en-us/computer-vision-api>

²<https://www.microsoft.com/cognitive-services/en-us/Computer-Vision-API/documentation/Category-Taxonomy>

3.1.2 Tariffe

Due tipologie di piani:

- Gratuito: fino a 5000 chiamate al mese, massimo 20 chiamate al minuto;
- Standard: fino a 10 TPS, 10\$ ogni 1000 chiamate.

3.1.3 Prova

La prova è stata eseguita utilizzando la funzione OCR.



Figura 3.1: default