

# Normalizzatore

Uso del Modulo Territoriale

# Descrizione del Modulo Territoriale

Il Modulo Territoriale ha l'obiettivo di Riconoscere , Normalizzare , Capificare , Geocodificare e Georeferenziare un'indirizzo presente nel territorio italiano;

Per Riconoscere si intende applicare delle "Ricerche Intelligenti" capaci di abbinare un' indirizzo comunque trascritto (anche con vistosi errori formali) purchè riconducibile univocamente all'interno del D.B. Territoriale.

Le entità di base presenti nel D.B. territoriale sono le seguenti:

Province ; Comuni ; Frazioni – Dug (Denominazione Urbanistica Generalizzata - Via,Piazza,etc..) – Stradario Nazionale – Archi di Strada C.A.P., Archi di Strada Sezioni di Censimento , Coordinate Sezioni di Censimento.

Per Normalizzazione si intende "riportare" la descrizione della Località/Comune e del Toponimo Strada presente nella banca dati territoriale; le descrizioni riportate sono quelle comunicate dai comuni all'Istat per le attività di Censimento;

Per Capificazione si intende riportare il Cap Località o Strada (per le 27 città multicap il CAP), oltre alla descrizione della Località Postale;

Per Geocodifica si intende attribuire i seguenti Codici Territoriali : Codice Regione Istat , Codice Provincia Istat , Codice Comune Istat , Codice Sezione di Censimento ISTAT 1991/2001 , Codice Strada AMS;

Per Georeferenziazione si intende attribuire (solo dopo aver riconosciuto l'indirizzo completo) le coordinate spaziali X,Y (Sistema Metrico WGS84 UTM-32) del Civico e le coordinate spaziali X,Y (Sistema Metrico WGS84 UTM-32) del Centroide della Sezioni di Censimento ISTAT 1991/2001.

I dati in input necessari per eseguire i processi di abbinamento sono i seguenti:

Località (Comune o Località non Comune) – Obbligatorio;

Sigla Aci - Opzionale ma utile per discriminare eventuali ambiguità di località;

C.A.P – Opzionale ma utile per discriminare località e/o strade;

Indirizzo (Via + Toponimo + Civico) – Obbligatorio;

Il software è provvisto di tre livelli di algoritmo:

1 - Semplice;

2 - Complesso;

3 - Forzatura DUG;

Il livello "semplice" esegue i seguenti processi;

- a. Individuazione e separazione dal toponimo strada della Specifica D.U.G. e Numero Civico;
- b. eliminazione caratteri speciali;
- c. maiuscolizzazione;
- d. individuazione parole speciali;
- e. ricerca in banca dati per "uguaglianza totale";
- f. qualora non trovato per uguaglianza totale esegue la ricerca nelle liste invertite località e strade dove almeno un termine deve essere riscontrato univocamente;
- g. creazione matrice decisionale;

Il livello "complesso" esegue i seguenti processi;

- a. Individuazione e separazione dal toponimo strada della Specifica D.U.G. e Numero Civico;
- b. eliminazione caratteri speciali;
- c. maiuscolizzazione;
- d. individuazione parole speciali;
- e. ricerca in banca dati per "uguaglianza totale";

- f. qualora non trovato per uguaglianza totale esegue la ricerca nelle liste invertite località e strade dove almeno una termine deve essere riscontrato univocamente;
- g. qualora non trovato esegue la ricerca sempre nelle liste invertite per abbreviazioni Iniziali / Finali / Sincopi verificando eventuali errori ortografici di mancanza o aggiunta di una lettera o inversione di due lettere adiacenti;
- h. creazione matrice decisionale;

Il livello "Forzatura DUG" esegue i seguenti processi sulla DUG;

- a. in caso di mancanza della DUG in input viene forzato "Via"
- b. in caso di univocità del toponimo strada nel Comune la DUG di input viene modificata se diversa

Il livello da utilizzare dipende dalla qualità dell'indirizzo di input da elaborare; se è in buone condizioni (presenza di tutti i campi, mancanza di errori ortografici, etc) oppure si desidera normalizzare solamente quei record estremamente "certi" utilizzare il livello "Semplice" in quanto le funzioni di abbinamento sono abbastanza restrittive; se invece il livello qualitativo dell'indirizzo di input è medio/basso e si vuole "tentare lo stesso il riconoscimento" si può utilizzare il livello Complesso, magari dopo aver prima eseguito un livello Semplice. Il livello Complesso può generare degli errori di abbinamento (la massima % di abbinamento errata con il livello complesso è stata pari ad un 7%) ma risolve tutta una serie di casistiche (errori ortografici, forte abbreviazioni, etc..) che il livello Semplice non contempla.

## Regole da utilizzare per un migliore riconoscimento dell'indirizzo e della località

E' buona norma inserire l'indirizzo completo di tutte le informazioni che si possiedono su DUG, toponimo e civico, sia per aumentarne il riconoscimento, che per accelerarne i tempi di elaborazione e quindi aumentare l'efficienza operativa dell'utente.

Il normalizzatore considera alcuni caratteri come separatori di informazioni nel campo "indirizzo":

la virgola ( , ) per separare il civico dal toponimo (es. Via del Corso, 1); tutto cio' che sta a destra della virgola viene considerato come numero civico e ulteriore specifica del civico (se composto da piu' elementi), ma ai fini della geocodifica, e georeferenziazione viene considerata solo la parte numerica piu' a sinistra, mentre il resto viene riportato in output come "civico completo".

Inoltre la virgola permette al normalizzatore di non confondere un eventuale numero presente nel toponimo come un civico (es. Via 25 Maggio, 22).

il trattino alto ( ¯ ) per separare la frazione/localita' dall'indirizzo (es. Via del Corso, 1 – loc. Frattocchie)

la presenza in input del codice di avviamento postale (CAP) e' utile al normalizzatore per eliminare la presenza di eventuali ambiguita' sia per l'indirizzo che per la localita'.

La sigla provincia non e' obbligatoria, ma funziona come il CAP, restringe i possibili candidati in normalizzazione per quanto riguarda la localita' in input, eliminando eventualmente ambiguita'.

Cosa fare se l'indirizzo viene scartato :

- Cambiare il livello di riconoscimento da 1 (semplice) a 2 (complesso)
- Se la DUG e' mancante o diversa da "VIA" forzare quest'ultima
- Se il toponimo e' composto da piu' termini (es. Via Carlo Alberto Generale Dalla Chiesa) , provare ad alternare i piu' significativi escludendo quelli che possono essere nomi propri di persona

- Invertire nome e cognome nel toponimo (es. GIUSEPPE GARIBALDI -> GARIBALDI GIUSEPPE) per ovviare a mancanze di sinonimie nel database viario
- Inserire una virgola tra toponimo e civico, o aggiungere una virgola anche in mancanza di civico
- In caso di presenza di numero o date nel toponimo, provare ad inserire le versioni letterali (25 in VENTICINQUE, oppure XXV)
- Evitare l'immissione di doppie DUG, (es. VIA CLIVO di CINNA)
- In caso di presenza nel toponimo di nomi stranieri, provare ad inserire le versioni italiane (Richard Wagner in Riccardo Wagner)