

Wikibase e Semantic MediaWiki: perché e come

itWikiCon



Luca **Mauri** | 2022-10-01



Di cosa si tratta

- **Wikibase**

è il nome comune di un gruppo di estensioni di MediaWiki che permettono di gestire dati memorizzati in maniera cosiddetta "semi-strutturata", ovvero non secondo le regole di un comune RDMBS

- **Semantic MediaWiki**

una estensione di MediaWiki che consente di aggiungere dati semantici direttamente all'interno del wikitesto di una pagina

- Vedete anche

https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki_and_Wikidata

- Queste due estensioni **non** sono *nemiche*

Puntate precedenti

- **itWikiCon** 2020-10-25
"Wikibase, partendo da zero"
- **SMWCon** 2020-11-25
"SMW and WikiBase: blasphemy or feasible?"
- Wikimedia **Hackathon** 2021-05-22
"Wikibase, starting from scratch"
- **WikidataCon** 2021-10-31
"Wikibase and SMW: why and how"

Contesto, molto brevemente

- WikiTrek è la più completa guida a to Star Trek in **Italiano**
 - È nata come il **porting** di un guida chiamata *HyperTrek*, basata su una applicazione PHP personalizzata come *frontend* di un database
 - Alla **prima** importazione, ho generato wikipresto completamente statico
 - Molto presto mi sono reso conto che era necessaria una soluzione **dinamica**, per molte ragioni (flessibilità, velocità, facilità di manutenzione e via dicendo)
 - Dopo essermi chiarito i meccanismi di funzionamento di Wikidata insieme a Wikipedia, ho realizzato la necessità di una istanza privata di **Wikibase**
- **DataTrek** è una banca dati collaborativa specifica per informazioni relative a Star Trek
 - L'obiettivo primario è fornire dati grezzi a **WikiTrek**

Tutto bene all'inizio...

- Dopo aver installato e configurato l'istanza di Wikibase Repository, tutto funzionava **bene**
- Per fare un esempio, i dati degli episodi potevano essere facilmente **estratti** dall'elemento di DataTrek e presentati in WikiTrek come infobox generati **automaticamente**



The screenshot shows the Wikidata infobox for "Stagione 3 di *Discovery*". It includes the following information:

- Numero di produzione:** 305
- Date e emittenti della prima:**
 - glovedì 12 novembre 2020 su *CBS All Access* (USA)
 - venerdì 13 novembre 2020 su *Netflix* (Resto del Mondo)
- Durata:** 56min
- Titolo italiano:** Provarci fino alla morte
- Personaggi e interpreti DT:**
 - Attori co-protagonisti:**
 - Tenente Keyla Detmer: Emily Coutts
 - Tenente Gen Rhys: Patrick Kwok-Choon
 - Tenente Joann Owosekun: Oyin Oladejo
 - Tenente Ronald A. Bryce: Ronnie Rowe Jr
 - Tenente Nilsson: Sara Mitich
 - El: Brendan Belser
 - Dottor Attis: Jake Epstein
 - Tenente Audrey Willa: Vanessa Jackson
 - Mamma Barzani: Ana Sani
 - Figlia Barzani 1: Ava McKinnon
 - Figlia Barzani 2: Shazdeh Kapadia
 - Attori ospiti:**
 - Ammiraglio Charles Vance: Oded Fehr
 - Adira Tal: Blu del Barrio
 - Kovich: David Cronenberg
 - Jett Reno: Tig Notaro
 - Attori in ruoli non accreditati:**
 - Linus: David Benjamin Tomlinson
 - Attori ospiti straordinari:**
 - Philippa Georgiou (specchio): Michelle Yeoh

At the bottom, there is a "Navigatore episodi" section with buttons for "< Precedente", "Successivo >", "Non ti scordar di me", and "Sfruttatori". A small link at the bottom right says "Modifica i dati nella pagina della entità su DataTrek".

...e in vari ambiti

- Oltre all'*infobox*, elementi della pagina come i link esterni hanno beneficiato della generazione automatica
 - **Collegamenti esterni** gestiti manualmente a qualsiasi sito esterno
 - **SiteLinks** verso altre wiki
 - **Identificativi esterni** per creare link a siti basati su database

The screenshot displays the 'Collegamenti esterni' (External links) section of a Wikipedia article. It includes a list of external reviews and articles, an 'Interwiki' section with links to Wikidata and Memory Beta, and an 'Identificativi esterni' (External identifiers) section with a link to IMDb. Below this, a detailed view of the IMDb identifier 'tt10677382' is shown, including options to modify, add references, or add values. At the bottom, there are boxes for 'Wikipedia (0 siti)', 'WikiTrek (1 sito)', 'Piattaforma Fandom (2 siti)', and 'Altri siti (1 sito)', each with a 'modifica' (edit) link and a list of associated external links.

Collegamenti esterni [edit source]

- *Review: 'Star Trek: Discovery' Passes The Test In "Die Trying"*[?], TrekMovie
- *RECAP: Star Trek: Discovery - Die Trying*[?], StarTrek.com
- *How Nhan's Journey Connects Discovery to The Next Generation*[?], StarTrek.com
- *[REVIEW]: Star Trek: Discovery – Episode 305 "Die Trying"*[?], TrekNews
- *STAR TREK: DISCOVERY Review: "Die Trying"*[?], TrekCore
- *Die Trying*[?], Ex Astris Scientia

Interwiki

- *Q100327254*[?][?], Pagina della entità su Wikidata
- *Die Trying*, Memory Beta (inglese)
- *Die Trying (episode)*[?], Memory Alpha (inglese)
- *Q4285*[?], Pagina della entità su DataTrek

Identificativi esterni

- *tt10677382*[?], Identificativo su IMDb

Identificativi esterni

Identificativo su IMDb tt10677382 modifica

0 riferimenti + aggiungi riferimento

+ aggiungi valore

+ aggiungi dichiarazione

Wikipedia (0 siti) modifica

WikiTrek (1 sito) modifica

wikitrek Die Trying

Piattaforma Fandom (2 siti) modifica

enma Die Trying (episode)

enmb Die Trying

Altri siti (1 sito) modifica

wikidata Q100327254

Finché non sono incappato nei problemi

- Il fatto che Wikibase sia *operativamente* molto **diverso** da un RDBMS mi si è manifestato improvvisamente in seguito a un paio di problemi pratici
- In maniera **automatica**
 - Serviva un modo semplice per **elencare** tutti gli episodi di una serie o di una stagione
 - Bisognava risolvere un semplice problema multi-a-uno: **raccogliere** le apparizioni degli attori e i relativi ruoli

The screenshot shows the Wikibase item page for 'Die Trying' (Q4285). The page is titled 'Die Trying (Q4285)' and has a 'Discussion' tab. Below the title, there is a table of language labels and a list of statements. The 'Series' statement is circled in red, and the 'Performer' statement is also circled in red.

Language	Label	Description	Also known as
English	Die Trying	No description defined	
Italian	Provarci fino alla morte	No description defined	
British English	No label defined	No description defined	
French	No label defined	No description defined	

Statements

- Instance of: Episode (+ 0 references)
- Series: Star Trek: Discovery (+ 0 references)
- Season: 3 (+ 0 references)
- Production number: 305 (+ 0 references)
- Original airdate: 12 November 2020 (broadcast: CBS All Access, Geographic area: USA) (+ 0 references); 13 November 2020 (broadcast: Netflix, Geographic area: Resto del Mondo) (+ 0 references)
- Performer: Michelle Yeoh (Character: Philippa Georgiou (mirror), Type of participation: SpecialGuest) (+ 0 references); Oded Fehr (Prefix: Ammiraglio, Character: Charles Vance, Type of participation: GuestStar)

La soluzione "classica"

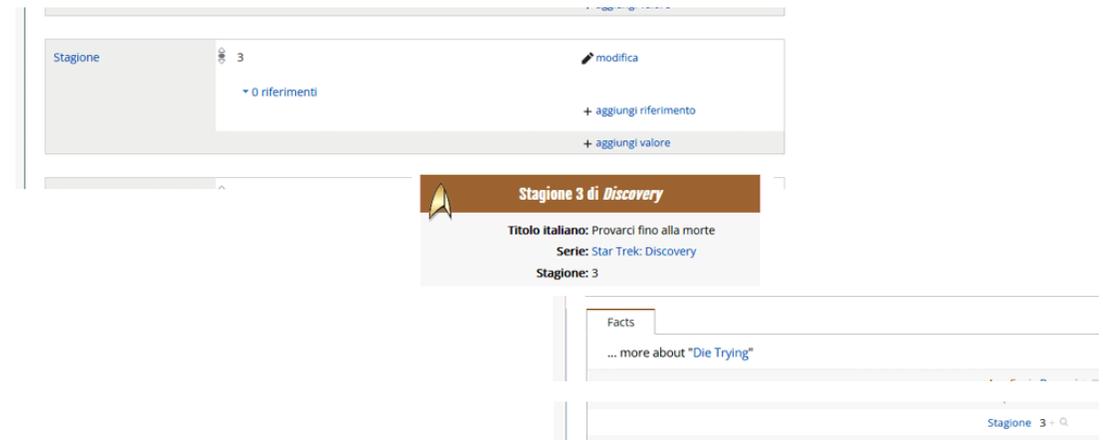
- Con una applicazione **tradizionale** basata su RDBMS, avrei potuto risolvere il problema con uno *statement* SQL per poi **esportare** l'output opportunamente formattato
 - Per fare un **esempio**, una query per ottenere tutti gli episodi di una specifica stagione di una data serie potrebbe assomigliare a
 - `SELECT * FROM Episodes WHERE Series = 'Discovery' AND Season = 3;`
- Una estrazione del genere, per molta sfortuna, è molto sconveniente da eseguire in Wikibase:
 - Mettere in funzione un *query service* (come QS di Wikidata) avrebbe permesso di eseguire **interrogazioni** con SPARQL, ma non di inserirle semplicemente nelle pagine

Arriva Semantic MediaWiki

- Per risolvere questi problemi operativi, ho **ripercorso** tutto il processo decisionale sulla infrastruttura del progetto e riconsiderato l'uso di **Semantic MediaWiki**
 - Era una delle soluzioni valutate prima di decidere per Wikibase, che sembrava un approccio più **moderno** e aperto
- SMW permette interrogazioni più **flessibili** e, ripensandoci, diventava una soluzione semplice ai problemi contingenti
 - Tuttavia, passare a SMW avrebbe significato **buttare** alle ortiche tutto il lavoro fatto su DataTrek e **rinunciare** all'idea di fornire una base dati aperta e standardizzata a potenziali terze parti

Facciamoli lavorare *insieme*

- Smettendo di pensare in maniera dicotomica, la soluzione è presto detta: WB e SMW *possono* funzionare **insieme**. Nulla vieta di memorizzare i dati in Wikibase **e** comunque di presentarli come dati semantici nella wiki:
 - Invece di estrarre i dati in una infobox *e basta*, con il wikitesto
 - Stagione: `{{#property:P18}}`
 - Gli stessi dati possono essere presentati all'utente **e** arricchiti con significato semantico:
 - Stagione: `[[Stagione::{{#property:P18}}]]`
- Quindi, i dati sono memorizzati DataTrek, ma WikiTrek può **anche** eseguire delle query su di loro tramite SMW



Interrogazioni semplificate

- Visualizzare la tabella degli episodi adesso è molto semplice con una query:

```
• {{#ask: [[Istanza::
{{#property:P90}}]]
[[Stagione::{{#property:P18}}]]
|?Titolo italiano
...
|format=table
|sort=Numero di produzione
|order=asc
|class=sortable wikitable
smwtable
}}
```

Page Discussion Read Edit Edit source View history More Search WikITrek

Stagione 3 di Discovery

Stagione 3 di Discovery è un Stagione di una serie

La Discovery e il suo equipaggio compiono un viaggio in un futuro lontano in cui si troveranno ad affrontare numerose complicazioni anche con chi pensavano dovesse essere loro alleati.

Dovranno rispondere a molte questioni e risolvere un misterioso enigma che ha compromesso la possibilità di viaggiare nello spazio.

Contents ^[hide]

- 1 Elenco episodi
- 2 Riferimenti
 - 2.1 Riferimenti diretti
 - 2.2 Riferimenti inversi
- 3 Collegamenti esterni
 - 3.1 Interwiki
 - 3.2 Identificativi esterni
- 4 Annotazioni

Elenco episodi

Titolo originale	Titolo italiano	Numero di produzione	Data di pubblicazione originale	Data di pubblicazione in Italia
That Hope Is You, Part 1	Quella speranza sei tu, parte 1	301	15 ottobre 2020	16 ottobre 2020
Far From Home	Lontano da casa	302	22 ottobre 2020	23 ottobre 2020
People of Earth	Popolo della Terra	303	29 ottobre 2020	30 ottobre 2020
Forget Me Not	Non ti scordar di me	304	5 novembre 2020	6 novembre 2020
Die Trying	Provarci fino alla morte	305	12 novembre 2020	13 novembre 2020
Scavengers	Sfruttatori	306	19 novembre 2020	20 novembre 2020
Unification III	Unificazione III	307	26 novembre 2020	27 novembre 2020
The Sanctuary	Il Rifugio	308	3 dicembre 2020	4 dicembre 2020
Terra Firma, Part 1	Terraferma: Parte 1	309	10 dicembre 2020	11 dicembre 2020
Terra Firma, Part 2	Terraferma: Parte 2	310	17 dicembre 2020	18 dicembre 2020
Su'Kal	Su'Kal	311	24 dicembre 2020	25 dicembre 2020
There Is A Tide...	C'è una marea...	312	31 dicembre 2020	1 gennaio 2021
That Hope Is You, Part 2	Quella speranza sei tu: Parte 2	313	7 gennaio 2021	8 gennaio 2021

Stagione di una serie

In originale: Discovery Season 3
Serie: Star Trek: Discovery
Stagione: 3

Navigatore

< Precedente Successivo >
[Stagione 2 di Discovery](#) [Stagione 4 di Discovery](#)

Modifica i dati nella [pagina delle entità su DataTrek](#)

Serie Star Trek: Discovery

Categorie

Serie • Personaggi • Episodi • Ordine di produzione • Titoli italiani

Episodi

Tutti • Stagioni: 1 • 2 • 3 • 4

Altre pagine

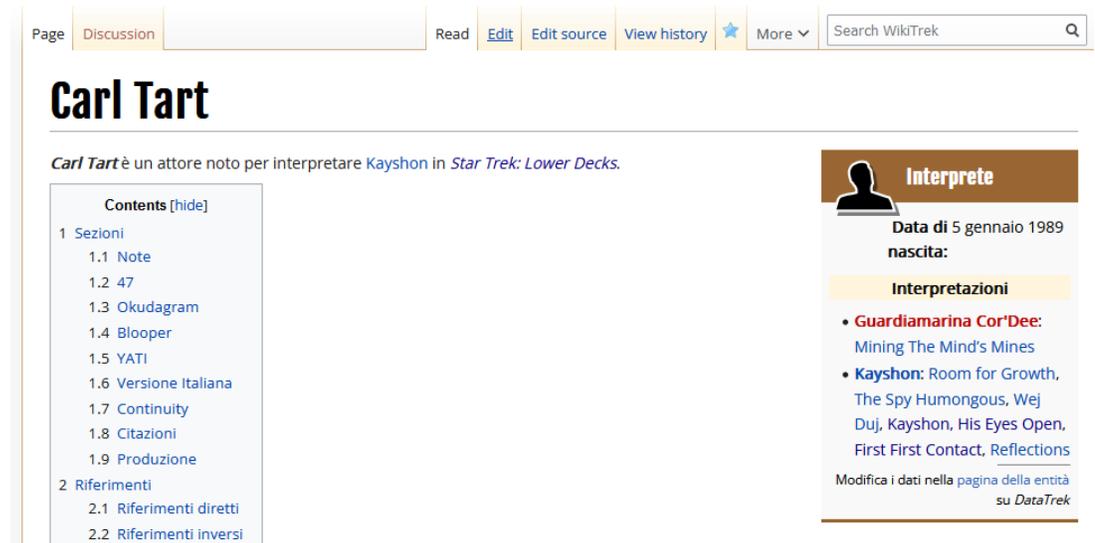
Personaggi ricorrenti di Discovery

Tutte le serie

Enterprise • Discovery • Short Treks • Strange New Worlds • Serie Classica, Serie Originale • Khan • Serie Animata • The Next Generation • Deep Space Nine • Voyager • Lower Decks • Prodigy • Picard

Caso interessante

- Un caso interessante e il calcolo automatico delle **apparizioni** di un interprete
 - **Aggrega** i personaggi interpretati
 - Elenca di conseguenza **tutte** le apparizioni dell'interprete in quel personaggio



The screenshot shows the WikiTrek page for Carl Tart. At the top, there is a navigation bar with tabs for 'Page', 'Discussion', 'Read', 'Edit', 'Edit source', 'View history', and 'More'. A search bar for 'Search WikiTrek' is also present. The main heading is 'Carl Tart'. Below it, a short biography states: 'Carl Tart è un attore noto per interpretare Kayshon in *Star Trek: Lower Decks*.' To the left, there is a 'Contents' table of contents with sections for 'Sezioni' (1.1 Note, 1.2 47, 1.3 Okudagram, 1.4 Bloopers, 1.5 YATI, 1.6 Versione italiana, 1.7 Continuity, 1.8 Citazioni, 1.9 Produzione) and 'Riferimenti' (2.1 Riferimenti diretti, 2.2 Riferimenti inversi). On the right, there is a 'Interprete' sidebar with a profile picture, birth date 'Data di 5 gennaio 1989 nascita:', and a list of roles under 'Interpretazioni': 'Guardiamarina Cor'Dee: Mining The Mind's Mines' and 'Kayshon: Room for Growth, The Spy Humongous, Wej Duj, Kayshon, His Eyes Open, First First Contact, Reflections'. At the bottom of the sidebar, it says 'Modifica i dati nella pagina della entità su DataTrek'.

Una occhiata al risultato della query

- Il risultato di una interrogazione a SNW è un **output** in formato JSON contenente
 - Un **preambolo** con dettagli della query
 - **L'elenco** dei risultati con le relative proprietà

```
{
  "printrequests": [
    {
      "label": "",
      "key": "",
      "redi": "",
      "typeid": "_wpg",
      "mode": 2,
      "format": false
    },
    {
      "label": "Carl Tart",
      "key": "Carl_Tart",
      "redi": "",
      "typeid": "_wpg",
      "mode": 1,
      "format": ""
    }
  ],
  "results": {
    "Kayshon, His Eyes Open": {
      "printouts": {
        "Carl Tart": [
          {
            "fulltext": "Kayshon",
            "fullurl": "https:\\\\wikitrete.org\\wiki\\Kayshon",
            "namespace": 0,
            "exists": "1",
            "displaytitle": ""
          }
        ]
      },
      "fulltext": "Kayshon, His Eyes Open",
      "fullurl": "https:\\\\wikitrete.org\\wiki\\Kayshon, His Eyes Open",
      "namespace": 0,
      "exists": "1",
      "displaytitle": ""
    },
    "The Spy Humongous": {
      "printouts": {
        "Carl Tart": [
          {
            "fulltext": "Kayshon",
            "fullurl": "https:\\\\wikitrete.org\\wiki\\Kayshon",
            "namespace": 0,
            "exists": "1",
            "displaytitle": ""
          }
        ]
      }
    }
  }
}
```

Un esempio di codice

- L'esempio di **codice** a lato è quello che si occupa di
 - Impostare la **query** ed eseguirla
 - Fare un **ciclo** lungo i contenuti del JSON
 - **Aggregare** i risultati in maniera compatta
 - **Presentarli** in maniera gradevole all'utente

```
function p.ListAppearances(frame)
  local FinalArray = {}
  local FinalString = ""
  local Actor = mw.title.getCurrentTitle().text
  local QueryResult = mw.smw.getQueryResult('[[Interprete::' .. Actor .. ']]?' .. Actor ..
'sort=Numero di produzione|order=asc')

  if QueryResult == nil then
    return "'Nessun risultato'"
  end

  if type(QueryResult) == "table" then
    for k, v in pairs(QueryResult.results) do
      -- v.fulltext represents EPISODE
      -- v.printouts[Actor][1].fulltext represents CHARACTER
      local Episode = v.fulltext
      local Character

      if v.printouts[Actor][1] == nil then
        Character = "'Senza pagina'"
      else
        for _, CurrChar in pairs(v.printouts[Actor]) do
          Character = CurrChar.fulltext
          if FinalArray[Character] == nil then
            FinalArray[Character] = {}
          end
          table.insert(FinalArray[Character], Episode)
        end
      end
    end
  end

  else
    return "'Il risultato non è una TABLE'"
  end

  for ID, Group in pairs(FinalArray) do
    FinalString = FinalString .. "* '" .. ID .. "'": [[[" .. table.concat(Group, ","), [{" ..
"']]" .. string.char(10)
  end

  return FinalString
end
```

Altro esempio significativo

- Le query di SMW consentono di generare con relativa **semplicità** box secondari anche complessi
 - Esempio a lato di un box che **sfrutta** tre proprietà dell'elemento per generare automaticamente i relativi **elenchi**



La **USS Enterprise NCC-1701-B**, prima nave della *Classe Excelsior* migliorata, nel 2293

[Modifica i dati nella pagina della entità su DataTrek](#)

USS Enterprise NCC-1701-B

Navi Enterprise

[Enterprise NX-01](#) • [USS Enterprise NCC-1701](#) • [USS Enterprise NCC-1701 \(reboot\)](#) • [USS Enterprise NCC-1701-A](#) • **[USS Enterprise NCC-1701-B](#)** • [USS Enterprise NCC-1701-C](#) • [USS Enterprise NCC-1701-D](#) • [USS Enterprise NCC-1701-E](#)

Personale

[Demora Sulu](#) • [John Harriman](#)

Classe Excelsior

[Astronave di classe Excelsior \(assfkexcelsior1\)](#) • [Astronave di classe Excelsior \(assfkexcelsior2\)](#) • [NCC-14427](#) • [NCC-45505](#) • [USS Al-Batani NCC-42995](#) • [USS Berlin NCC-14232](#) • [USS Cairo NCC-42136](#) • [USS Charleston NCC-42285](#) • [USS Crazy Horse NCC-50446](#) • [USS Crockett NCC-38955](#) • **[USS Enterprise NCC-1701-B](#)** • [USS Excelsior NCC-2000](#) • [USS Farragut](#) • [USS Fearless NCC-4598](#) • [USS Fredrickson NCC-42111](#) • [USS Gorkon NCC-40512](#) • [USS Grissom NCC-42857](#) • [USS Hood NCC-42296](#) • [USS Intrepid NCC-38907](#) • [USS Lakota NCC-42768](#) • [USS Lexington NCC-14427](#) • [USS Livingston NCC-34099](#) • [USS Malinche](#) • [USS Melbourne NCC-62043](#) • [USS Potemkin NCC-18253](#) • [USS Repulse NCC-2544](#) • [USS Roosevelt NCC-2573](#) • [USS Tecumseh NCC-14934](#) • [USS Valley Forge NCC-43305](#)

T • V • E

Generazione automatica delle proprietà

- Le stesse Proprietà di SMW sono definite da Dichiarazioni delle proprietà corrispondenti in Data Trek
 - "*Description*" corrisponde alla descrizione della proprietà
 - "*Has type*" al momento è impostato in maniera manuale, ma una miglioria futura è effettuare una conversione automatica dei **tipi**

Proprietà Discussion

Proprietà:Numero di produzione

Descrizione

Questa proprietà semantica:

- è di tipo *Text*
- ha descrizione
 - Numero di produzione (it)
 - Production number (en)

DataTrek [\[edit source \]](#)

Corrisponde alla proprietà Numero di produzione (P1) ★ di DataTrek

Usage 852

previous 20 20 50 100 250 500 next 20

Showing 20 pages using this property.

'		
	'Til Death Do Us Part +	568 + Q
1		
	11001001 +	116 + Q
	11:59 +	217 + Q
A		
	A Fistful of Datas +	234 + Q
	A Man Alone +	403 + Q
	A Matter of Honor +	124 + Q

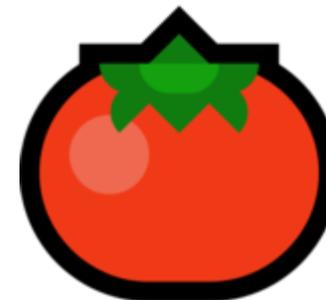
Acqua calda?

- Effettivamente esiste una estensione chiamata **Semantic Wikibase**
- A **prima** vista sembrerebbe fare tutto da sola, ma
- A **ben** vedere ha alcuni problemi
 - Funziona **solo** se tutti i componenti sono su una singola istanza di Mediawiki (per questa ragione non ho potuto provarla in prima persona)
 - Lo sviluppo è apparentemente bloccato e non c'è attività della **comunità**
 - Per quanto sia più semplice da integrare, sembra anche più **rigida** nell'operatività

Ci avviciniamo alla fine

- Molto è stato fatto, ma ancora **molto** lavoro resta da fare
- Nella presentazione ho cercato di mostrare i concetti più importanti e le implementazioni più **esemplificative**
- Tutto il codice LUA è disponibile nel nostro **repository** pubblico <https://github.com/WikiTrek> e viene sincronizzato *al volo* tramite estensione **PageToGitHub**
 - Il codice è funzionale, ma **brutto**, per piacere non siate troppo severi nel giudicare
 - Al contrario, chiunque volesse contribuire sarà **ben** accetto

Grazie, ora è il vostro turno



t.me/lucamauricom