



Laboratorio Innovazione.

Risultati e prospettive

**CONFRONTO DELLE TECNICHE DI WEB SCRAPING NELLA
RILEVAZIONE DEI PREZZI AL CONSUMO PER I PACCHETTI
VACANZA INTERNAZIONALI E NAZIONALI**

GIUSEPPINA NATALE | ISTAT / DCSW / SWA

24
NOVEMBRE
2023

#IstatWebinar

CONFRONTO DELLE TECNICHE DI WEB SCRAPING NELLA RILEVAZIONE DEI PREZZI AL CONSUMO PER I PACCHETTI VACANZA INTERNAZIONALI E NAZIONALI

GIUSEPPINA NATALE | ISTAT -DIPS-DCSW-SWA

MASSIMILIANO AMARONE | ISTAT -DIPS-DCRD-RDE

RICCARDO GIANNINI | ISTAT -DIRM-DCIT-ITB

Indice



- La rilevazione dei prezzi al consumo per i pacchetti vacanza
- Pacchetti vacanza (Internazionali – Nazionali)
- La rilevazione dei pacchetti vacanza
- L'idea da sviluppare - Scopo del progetto –Obiettivi- Risultati attesi
- Parte informatica
- Risultati - Conclusioni
- Ringraziamenti

La rilevazione dei prezzi al consumo per i pacchetti vacanza



I **pacchetti vacanza** partecipano all'indagine mensile sui **prezzi al consumo**, fanno parte della divisione di spesa **ECOICOP 09** (ricreazione, spettacoli e cultura), nel gruppo: **pacchetti vacanza** e nei due aggregati di prodotto: **pacchetti vacanza internazionali**, **pacchetti vacanza nazionali**.

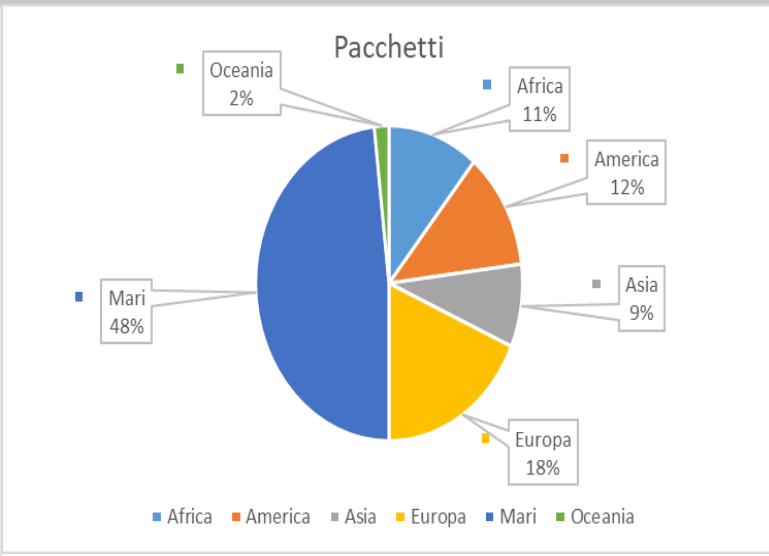
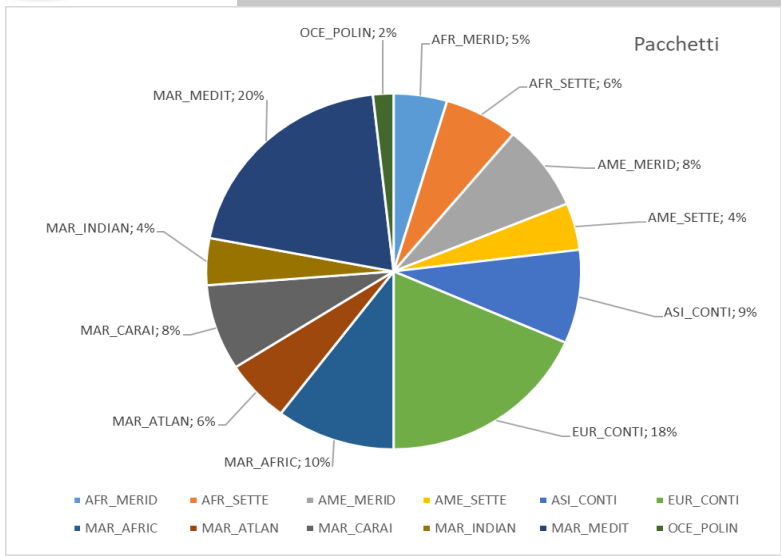


Pacchetti Vacanza Internazionali



Pacchetti Vacanza Internazionali			
282	21	41	12
Pacchetti vacanza	Tour Operator	Destinazioni estere	Macroaree geografiche estere
	ASTOI Confindustria Viaggi	(esempio: Brasile; Canada; Croazia; Francia; Spagna; Cina)	(esempio: Asia; Europa; America; Mare Med)

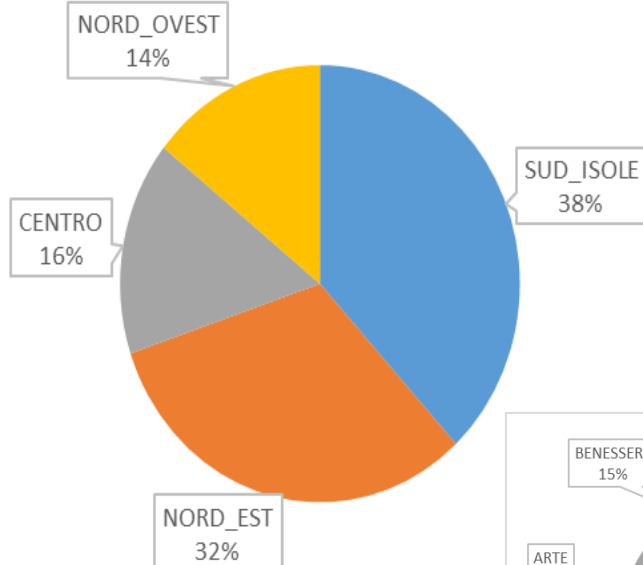
Macroarea	Pacchetti
AFR_MERID	5%
AFR_SETTE	6%
AME_MERID	8%
AME_SETTE	4%
ASI_CONTI	9%
EUR_CONTI	18%
MAR_AFRIC	10%
MAR_ATLAN	6%
MAR_CARAI	8%
MAR_INDIAN	4%
MAR_MEDIT	20%
OCE_POLIN	2%



Macroarea	Pacchetti
Africa	11%
America	12%
Asia	9%
Europa	18%
Mari	48%
Oceania	2%

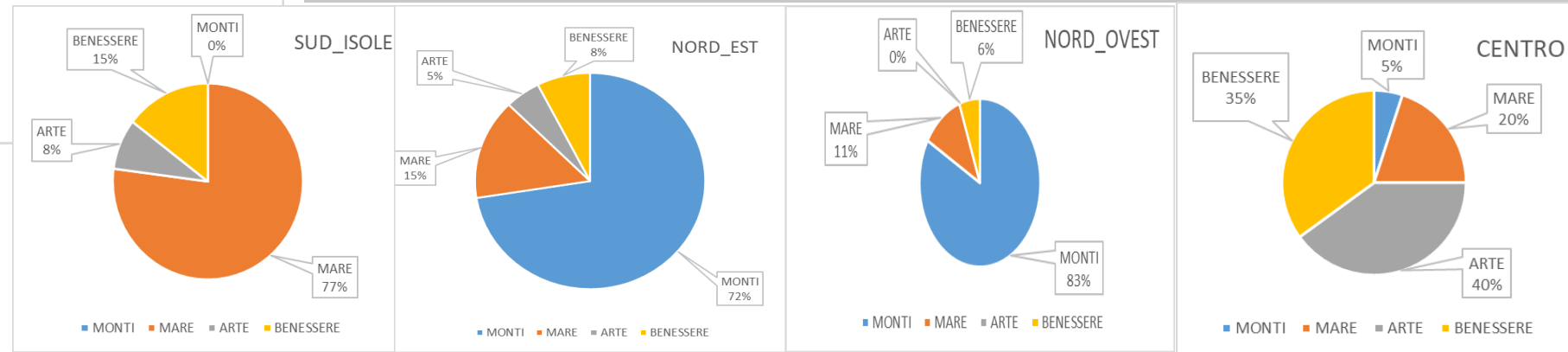
Pacchetti Vacanza Nazionali

Pacchetti per macroaree geografiche



Pacchetti Vacanza Nazionali

126	8	4	4
Pacchetti vacanza	Tour Operator	Macroaree geografiche	Tipologie di vacanza
	ASTOI Confindustria Viaggi	(Nord Est; Nord Ovest; Centro; Sud_Isole)	(Mare; Monti Benessere;Arte)



Tour operator nazionali

Macroaree geografiche

Tipologie di vacanza

Da notare: nello strato: **Sud Isole** non troviamo la tipologia di vacanza **Monti**, esistono nello strato «Sud isole» le montagne come ad esempio La Sila in Calabria, il parco del Pollino ma non è stato possibile recuperare informazioni utili «tour operator» che forniscono cataloghi on line per questo tipo di vacanza, analogo discorso per lo strato **Nord Ovest** per la tipologia di vacanza **Arte**.

La rilevazione dei pacchetti vacanza

La rilevazione dei prezzi si basa sulla consultazioni mensile dei listini/cataloghi on-line offerti dai tour operator presenti nel campione.

8 giorni / 7 notti - All Inclusive-Soft Drinks

QUOTE INDIVIDUALI DI PARTECIPAZIONE

Partenze Dal / Al	VOLI ITC			SUPPLEMENTI		RIDUZIONI		
	Quota Boom	Prenota Prima 30 giorni	Adulti 8 giorni/ 7 notti	Settimana supplementare	Supplemento Singola	Riduzione Tripla	1° Bambino 2/12 anni	2° Bambino 2/12 anni
01/12 - 17/12	470	560	700	240	no suppl.	50	gratis	30%
18/12 - 24/12	-	640	780	290	190	50	50%	30%
25/12 - 31/12	-	730	870	360	260	50	50%	30%
01/01 - 06/01	-	630	770	310	210	50	50%	30%
07/01 - 28/01	470	550	690	240	no suppl.	50	gratis	30%
29/01 - 18/02	490	580	720	240	170	50	gratis	30%
19/02 - 03/03	520	590	730	250	170	50	gratis	30%
04/03 - 16/03	550	630	770	270	170	50	gratis	30%
17/03 - 23/03	-	690	830	340	230	50	50%	30%
24/03 - 31/03	-	720	860	360	230	50	50%	30%
01/04 - 21/04	560	640	780	290	190	50	gratis	30%
22/04 - 27/04	590	670	810	290	190	50	50%	30%
28/04 - 26/05	550	630	770	290	190	50	gratis	30%
27/05 - 02/06	590	670	810	290	190	50	50%	30%
03/06 - 16/06	550	630	770	260	190	50	gratis	30%
17/06 - 30/06	570	650	790	260	190	50	50%	30%
01/07 - 21/07	590	660	800	290	210	50	50%	30%
22/07 - 27/07	610	680	820	290	230	50	50%	30%
28/07 - 03/08	620	690	830	320	230	50	50%	30%
04/08 - 07/08	640	720	860	390	290	50	50%	30%
08/08 - 15/08	-	810	950	490	330	50	50%	30%
16/08 - 21/08	-	740	880	460	310	50	50%	30%
22/08 - 31/08	640	710	850	390	270	50	50%	30%
01/09 - 08/09	540	620	760	290	190	50	gratis	30%
09/09 - 30/09	510	590	730	260	190	50	gratis	30%
01/10 - 31/10	530	610	750	260	190	50	gratis	30%
01/11 - 30/11	510	590	730	240	190	50	gratis	30%

Quota gestione pratica € 80. Oneri (Council Tax, PRM e tasse aeroportuali) € 42
 Visto € 33 (a passaporto o carta d'identità); Polizza annullamento viaggio obbligatoria pag. 103
 Supplemento per Hard All Inclusive € 45 per persona a settimana

Esempio listino

JAZ FANARA RESORT & RESIDENCE

All inclusive - camera standard

Stagionalità	QUOTA PACCHETTO (7 notti + volo a/r)			notte segg.	doppia uso singola	doppia usi singola superior
	START	EASY	BASE			
06/05-19/05	890	995	1.065	74	35	55
20/05-23/06	950	1.020	1.090	74	35	55
24/06-30/06	1.090	1.160	1.230	77	35	55
01/07-28/07	1.090	1.160	1.230	81	38	57
29/07-04/08	1.090	1.160	1.230	82	39	59
05/08-11/08	1.450	1.520	1.590	82	39	59
12/08-18/08	1.550	1.620	1.690	82	39	59
19/08-25/08	1.490	1.560	1.630	82	39	59
26/08-01/09	1.130	1.195	1.265	82	39	59
02/09-08/09	1.090	1.160	1.230	79	38	57
09/09-29/09	960	1.030	1.095	79	38	57
30/09-03/11	990	1.060	1.130	88	43	62

Occupazione camere: standard max 2 adulti (culla inclusa); superior e supe family deluxe max 2 adulti + 2 bambini (culla inclusa); family suite min 3 adu
Riduzioni soggiorno: 1° bambino 2-12 anni -100%, 2°/3° bambino 2-12 anr supplementari o in caso di solo soggiorno); 3° letto adulto -€ 12 a notte; 4° le

* LA QUOTA FORFAIT INCLUDE

L'importo del forfait è sempre dovuto sia per adulti, sia per bambini, per pre aeroportuali, la quota individuale di gestione pratica, gli oneri di trasporto.



Volo + 7 notti - All Inclusive - Quote individuali in camera doppia

Partenze Dal / Al	QUOTE		SUPPLEMENTI PER PERSONA PER NOTTE	
	Prima Quota a partire "DA"	Notte extra/solo hotel a partire "DA"	Notte in singola	
01/05	03/05	460	38	26
04/05	24/05	460	38	22
25/05	07/06	460	38	26
08/06	14/06	540	38	26
15/06	21/06	540	40	26
22/06	05/07	520	40	28
06/07	12/07	520	40	30
13/07	19/07	520	41	30
20/07	26/07	540	41	30
27/07	02/08	580	41	30
03/08	09/08	740	43	35
10/08	16/08	760	43	40
17/08	23/08	720	43	35
24/08	30/08	620	43	35
31/08	06/09	590	38	26
07/09	13/09	540	38	26
14/09	20/09	500	38	26
21/09	27/09	500	38	22
28/09	04/10	520	38	22
05/10	11/10	540	38	26
12/10	25/10	540	39	26
26/10	01/11	570	39	30
02/11	08/11	540	40	26
09/11	15/11	520	40	22
16/11	22/11	490	40	22
23/11	29/11	460	40	22
30/11	06/12	460	39	22
07/12	13/12	440	39	22
14/12	17/12	440	-	20

Quota gestione pratica € 85
 Oneri e tasse aeroportuali pag. 22-23

PROMO BIMBO FREE: 1° bambino (2-14 anni non compiu- ti): sconto fino al 100%

Riduzioni:
 Riduzione per 5° letto adulto € 9 al giorno

L'idea da sviluppare- Scopo del progetto - Obiettivi

Confronto tra le tecnologie di Web Scraping per l'acquisizione dei prezzi, considerando **4 campioni**

- 2 per i Pacchetti Vacanza Internazionali (campione fisso/variabile Last Minute)
- 2 per i Pacchetti Vacanza Nazionali (campione fisso/variabile Last Minute).



Pacchetti Vacanza Internazionali

Campione fisso (282) - Campione variabile (Last Minute)



Pacchetti Vacanza Nazionali

Campione fisso (126) - Campione variabile (Last Minute)

L'introduzione dei Pacchetti Vacanza «Last Minute» è suggerito da **Eurostat** nel manuale metodologico «Harmonised Index of Consumer (HICP)» novembre 2018 nel capitolo 12.5 Flights and package holidays.

Raccomandazione 2:

«I prezzi dovrebbero essere registrati con sufficiente anticipo rispetto alla partenza per garantire che siano rappresentativi della spesa e del comportamento dei consumatori. Inoltre, la quota di spesa delle offerte last minute dovrebbe essere esaminata e se significativa, inclusa nell'indice come voce aggiuntiva».



L'idea da sviluppare- Scopo del progetto - Obiettivi

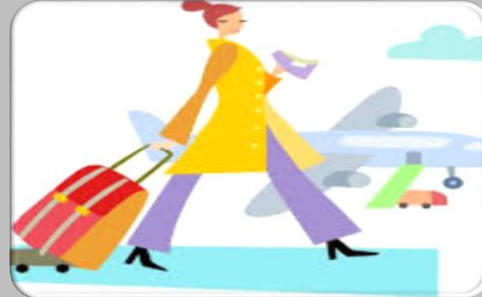


Risultati attesi

- Scelta della tecnologia di web scraping per l'acquisizione dei prezzi dei pacchetti vacanza internazionali e nazionali;
- Inclusione della tecnologia nel processo produttivo.



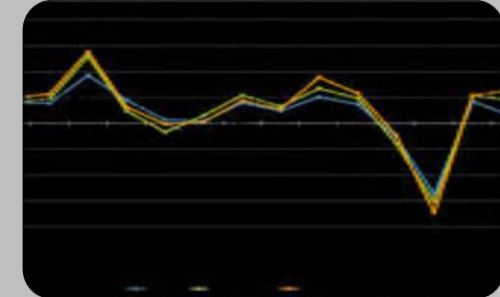
Ottimizzare i tempi della rilevazione con una consultazione più semplice dei listini



Ampliare il campo di osservazione con un numero maggiore di quotazioni rilevate mensilmente



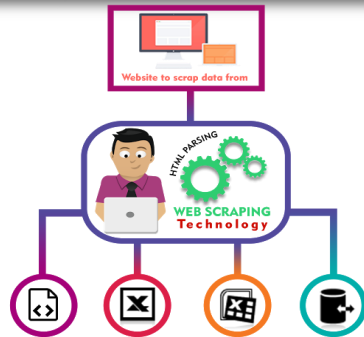
Migliorare la rappresentatività dei comportamenti di acquisto dei consumatori inserendo anche i pacchetti last minute per entrambi gli aggregati



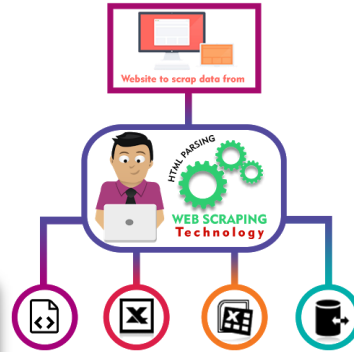
Ottenere una misura più accurata della stima dell'inflazione per i servizi turistici

Parte informatica

Riccardo Giannini
(Software iMacros)



Massimiliano Amarone
(Software scritto in Java
con utilizzo delle API
della libreria HtmlUnit)



Tecnologia - Architettura prototipo iMacros

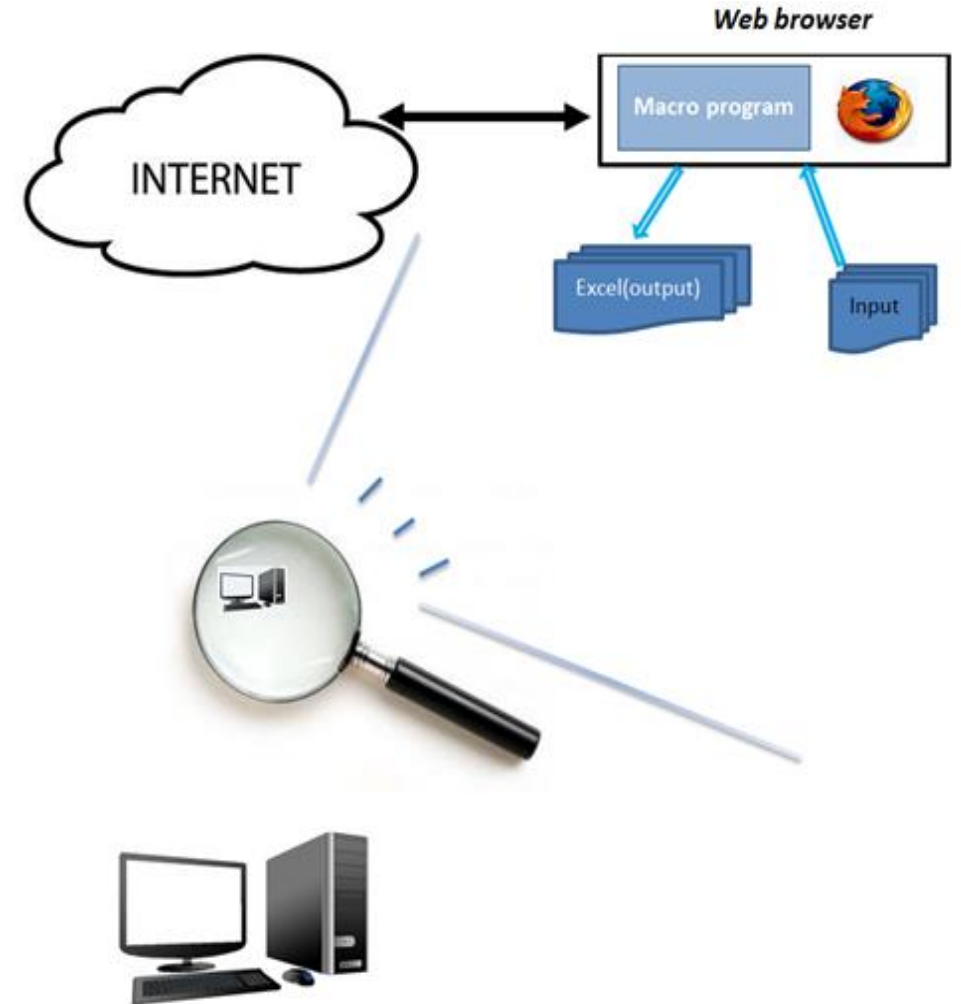
Ci si riferisce ad iMacros sia come software che come una tecnica di automazione webscraping per l'acquisizione di dati dal web tramite l'uso di un browser.

iMacros nasce nel 2001.

Il team che l'ha sviluppata è della società iOpus che a sua volta appartiene dal 2012 alla Ipswitch inc.

Il suo utilizzo può avvenire sia all'interno del browser che avvalendosi dei linguaggi di programmazione come Java, Php, C# etc..

- **Macro program:** programma eseguito all'interno del browser grazie ad un'estensione che coordina tutte le operazioni all'interno del browser (cleaning, scrittura, webscraping etc..)
- **Web browser:** Mozilla Firefox è il browser scelto all'interno del quale avviene il webscraping.
- **Output:** file testo dove vengono scritti i dati letti dai siti internet.
- **Input:** qui risiedono i dati utilizzati dalla macro per guidare la rilevazione.



Tecnologia - Architettura prototipo Java

Tecnologia principale: Java con l'estensione delle librerie Selenium per l'automazione dei browser web.

Ruolo del WebDriver di Selenium: Il *WebDriver* di Selenium è l'interfaccia chiave per automatizzare le interazioni con i browser web. Consente l'esecuzione di comandi automatici su pagine web.



Descrizione del progetto

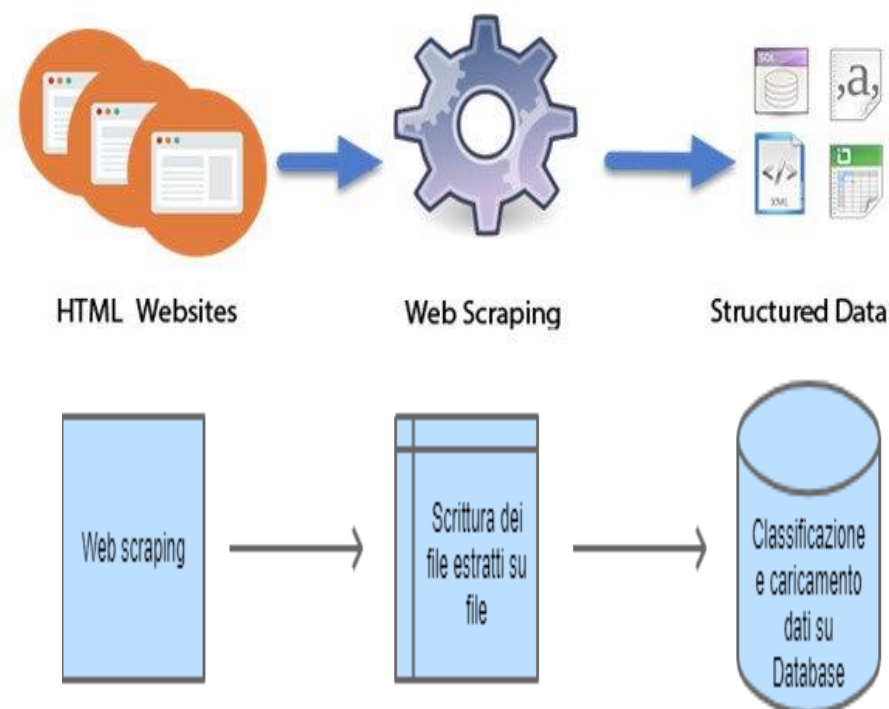
Composto da tre processi principali:

Processo 1: Scansione e scrittura delle informazioni relative alle offerte e ai last minute dai portali specificati. Automatizzato e schedato su un server Linux.

Processo 2: Lettura dei file scritti, normalizzazione delle classificazioni e caricamento su un database Oracle in forma strutturata. Eseguito quotidianamente su server Linux.

Processo 3: Calcolo di indicatori relativi alle statistiche sul turismo basati sulle informazioni salvate nel database.

Architettura Software



Risultati - Conclusioni

Per il **campione fisso è stato impossibile catturare i prezzi tramite web scraping** per la grande quantità di informazioni presenti nei listini pubblicati come PDF o JPEG, i quali necessitano di una continua manutenzione dal punto di vista informatico, al contrario per il **campione variabile formato dai prezzi «last minute» la scelta del web scraping è risultata ottimale.**



Periodo dati rilevati: **Luglio 2020 – ottobre 2021**
 Raccolti: **DB Oracle Last Minute Offerte**

Quotazioni Offerte: **Italia 6865**, **Estero 6068**
 Quotazioni Last minute: **Italia 1759**, **Estero 10408**



Tour operator Last Minute Offerte: **10**

Destinazioni: **Italia: Strato Sud Isole**, **Estero: mare Egitto, Grecia**

Sperimentazione: **Indice di Laysperes base di calcolo luglio 2020**



Risultati - Conclusioni

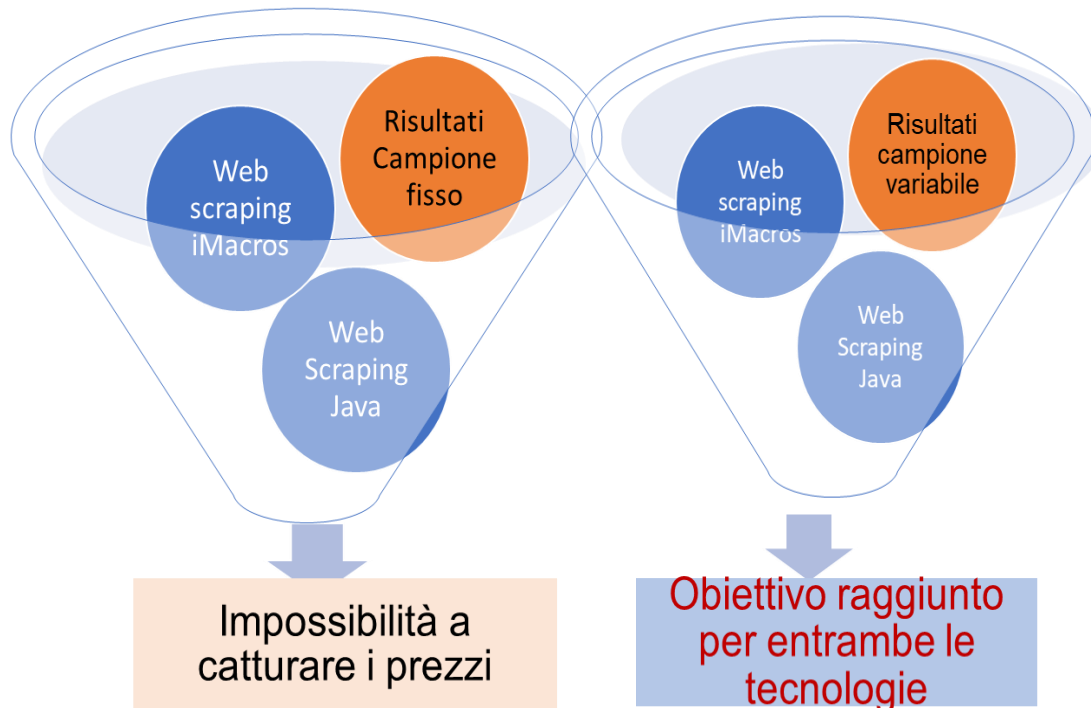
Confronto tra le tecnologie di Web Scraping

Tecnologia iMacros

- Raggiunge il risultato per last minute/offerte, richiede una maggiore manutenzione rispetto alla tecnologia Java, poiché prima di lanciare le singole macro è necessario predisporre i **file di input** da aggiornare ad ogni lancio.

Tecnologia Java

- Raggiunge il risultato per last minute/offerte, catturando **tutti i giorni** i prezzi pubblicati, senza creare nessun file aggiuntivo oltre a quello iniziale, con uno sforzo minimo ottenendo una grande quantità di informazioni.



Questo progetto analizzato durante la pandemia mondiale (Covid 19) ha messo in evidenza, inoltre, alcune difficoltà nell'utilizzare i prezzi, offerti da tour operator nei vari cataloghi, disponibili ma non acquistabili dagli acquirenti, per i vincoli esistenti dai vari paesi, al contrario dei prezzi, catturati tramite web scraping, disponibili all'acquisto perché proposti esclusivamente in luoghi senza vincoli di spostamenti.



Dedicato al ricordo di mia sorella
Caterina Natale

24
NOVEMBRE
2023

GRAZIE
PER L'ATTENZIONE

#IstatWebinar