



SAPIENZA
UNIVERSITÀ DI ROMA

UNIVERSITÀ "SAPIENZA" DI ROMA
FACOLTÀ DI INFORMATICA

BloodlineSeeker

Wikipedia web scraper per gli imperatori romani

Authors

Alessio Bandiera
Matteo Benvenuti
Simone Bianco
Andrea Ladogana

12 giugno 2022



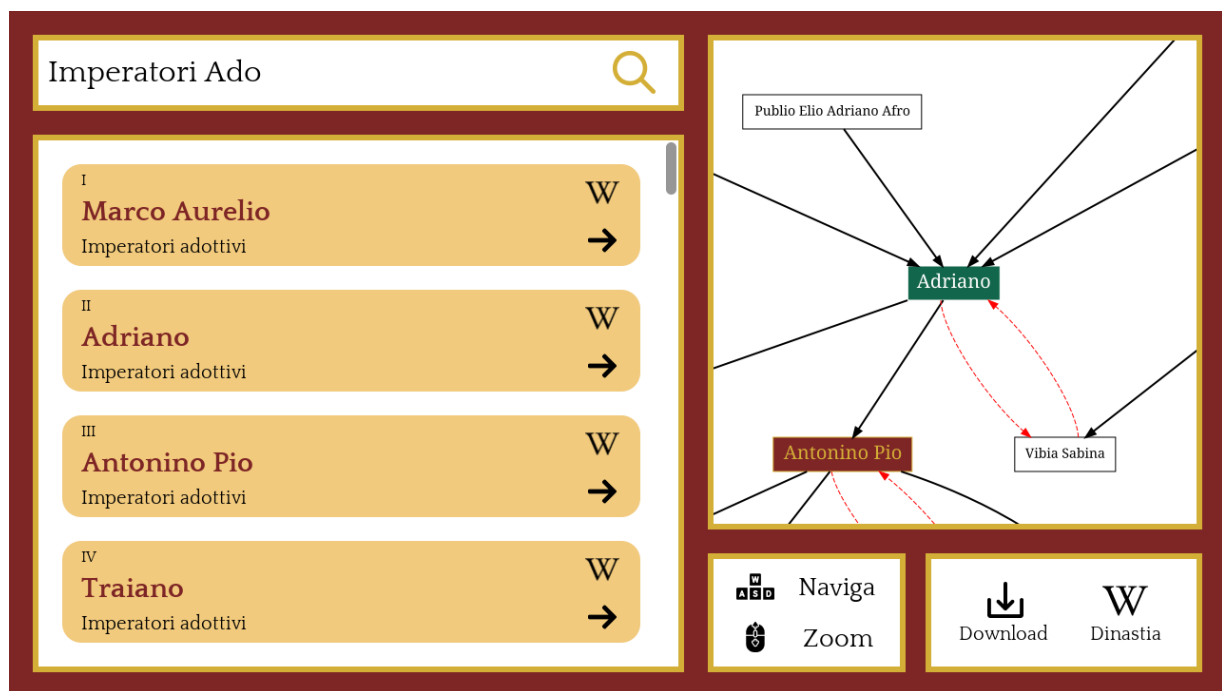
Indice

1	Introduzione	1
2	Progettazione e struttura del programma	2
2.1	[+++] Scelte di progettazione	2
2.1.1	[+++] File JAR	2
2.2	[+++] Progettazione delle classi	2
2.2.1	[+++] Descrizione delle classi	2
2.2.2	[+++] Gerarchia delle classi (UML)	2
2.3	Librerie utilizzate	2
3	Funzionalità del programma	4
3.1	Configurazione del programma	4
3.2	Estrazione dei file	4
3.3	Interpretazione degli argomenti	4
3.4	Antialiasing	5
3.5	Ricerca delle dinastie in locale	5
3.6	[+++] Scraping delle dinastie	6
3.6.1	[+++] Da Wikipedia utilizzando Selenium	6
3.6.2	[+++] Da Wikipedia utilizzando richieste HTTP	6
3.6.3	[+++] Console di download in tempo reale	6
3.7	[+++] Serializzazione delle dinastie	6
3.7.1	[+++] Algoritmo di serializzazione	6
3.7.2	[+++] Salvataggio e caricamento delle dinastie	6
3.8	[+++] Ricerca di persone e dinastie	6
3.8.1	[+++] Algoritmo di ricerca	6
3.8.2	[+++] Suggerimenti di ricerca	6
3.8.3	[+++] Visualizzazione dei risultati	6
3.9	[+++] Grafico delle dinastie	6
3.9.1	[+++] Struttura del grafico	6
3.10	[+++] Referenti dello sviluppo	6
4	Manuale della grafica	7
4.1	Finestra di avvio	7
4.1.1	Dinastie non trovate	7
4.1.2	Download delle dinastie	8
4.1.3	Dinastie trovate	9
4.2	Finestra di ricerca	10
4.2.1	Struttura generale	10
4.2.2	Suggerimento di ricerca	11
4.2.3	Risultati di ricerca e visualizzazione del grafico	12

Capitolo 1

Introduzione

La seguente relazione riguarda il progetto del corso di Metodologie di Programmazione che ha l'obiettivo di facilitare la visualizzazione dei legami di parentela che intercorrono tra gli imperatori romani e i membri delle loro dinastie. A tale scopo l'applicazione utilizza un web scraper per ottenere da [Wikipedia](#) i dati necessari a generare gli alberi genealogici dei membri, fornendo inoltre una interfaccia grafica per un utilizzo facile e intuitivo, che permette all'utente di navigare il grafico direttamente nella finestra.



Da un punto di vista tecnico, il progetto ha richiesto l'utilizzo di [Selenium](#), un framework open-source impiegato per l'automazione delle operazioni sui browser, e di [Graphviz](#), un programma open-source multi-piattaforma per generare grafi. Infine sono stati utilizzati strumenti per facilitare il lavoro in gruppo e aumentare la produttività generale del team quali: [git](#), [GitHub](#) e la funzione "Live Share" di [Visual Studio Code](#) che ha permesso la collaborazione per la realizzazione delle funzionalità più complesse.

Capitolo 2

Progettazione e struttura del programma

2.1 [+++] Scelte di progettazione

2.1.1 [+++] File JAR

2.2 [+++] Progettazione delle classi

2.2.1 [+++] Descrizione delle classi

2.2.2 [+++] Gerarchia delle classi (UML)

2.3 Librerie utilizzate

- [Selenium 4.1.4](#)

La libreria di Selenium per Java è stata utilizzata per effettuare lo scraping delle pagine da Wikipedia, in particolare per eseguire le richieste del codice HTML delle pagine.

Inoltre, il programma dispone di alcuni driver per Selenium, sia per Windows che per Unix:

- Microsoft Edge
- Google Chrome
- Mozilla Firefox

- [Log4j 1.2.17](#) e [SLF4J-Log4j 1.6.4](#)

Le librerie Log4j e SLF4J vengono utilizzate per monitorare gli output di Graphviz.

- [Graphviz 0.18.1](#)

La libreria Graphviz per Java viene utilizzata per la generazione degli alberi genealogici a partire dai dati ottenuti dallo scraping.

- [Flatlaf 2.2](#)

Flatlaf è una libreria multi-piattaforma per modificare il Look and Feel di Swing, e viene utilizzata per aumentare la personalizzazione dell'interfaccia

grafica. Alcuni dei suoi utilizzi sono: la possibilità di impostare bordi curvi ai componenti, migliorare l'aspetto grafico del JFileChooser, modificare la posizione e le dimensioni della barra di scorrimento del JScrollPane.

- **Jsoup 1.14.3**

La libreria Jsoup viene utilizzata dal programma per effettuare il parsing del codice HTML delle pagine di Wikipedia.

- **JSON-Java 20220320**

La libreria JSON-Java viene utilizzata dal programma sia per effettuare il salvataggio in locale delle informazioni delle dinastie in file JSON, sia per interpretare questi ultimi e caricare le dinastie in memoria.

- **Thumbnailator 0.4.8**

La libreria fornisce dei metodi efficienti per il ridimensionamento dell'immagine, indispensabili per la navigazione del grafico in tempo reale.

- Ricerca delle dinastie salvate in locale
 - Adattamento dell'interfaccia secondo l'esito della ricerca ● ●
- Avvio del programma
 - Utilizzando le dinastie salvate in locale ● ●
 - Scaricando le dinastie
 - Da Wikipedia utilizzando Selenium
 - Da Wikipedia utilizzando richieste HTTP
 - Interfaccia che mostra in tempo reale il download delle dinastie
- Ricerca di persone e dinastie
 - Filtraggio della chiave di ricerca
 - Suggerimento in caso di ricerca senza risultati
 - Visualizzazione dei risultati in ordine di importanza
 - Collegamento alla pagina Wikipedia della persona
 - Caricamento del grafico della persona
- Visualizzazione del grafico
 - Grafico centrato sulla persona cercata
 - Grafico esplorabile dinamicamente tramite spostamento
 - Grafico ridimensionabile dinamicamente
- Dinastia caricata
 - Scaricare il grafico della dinastia
 - Visitare la pagina Wikipedia della dinastia

Capitolo 3

Funzionalità del programma

Di seguito verranno esposti tutti gli accorgimenti presi in fase di progettazione, volti a migliorare l'esperienza finale dell'utente.

3.1 Configurazione del programma

All'interno di questa fase, vengono configurate alcune librerie al fine di ottenere un comportamento in linea con le esigenze del programma. In particolare, il Logger di Log4j viene impostato in modo tale da non restituire alcun messaggio di log nel terminale, e viene inizializzata la libreria AWT con la modalità `headless` abilitata per evitare conflitti tra Swing e la libreria Graphviz.

3.2 Estrazione dei file

In accordo con le scelte progettuali esposte nella sezione [File JAR](#), al fine di ottenere un file JAR indipendente da file esterni, è stato fondamentale sfruttare alcuni accorgimenti per importare le risorse necessarie al corretto funzionamento del programma. In particolare, all'avvio dell'applicazione viene effettuato un controllo preliminare all'interno della cartella di sistema dedicata ai file temporanei; questo serve a verificare l'eventuale presenza di risorse del programma precedentemente importate che, in tal caso, vengono rimosse per evitare conflitti. Successivamente, all'interno di questa cartella, vengono decompressi i file necessari per il funzionamento di Graphviz.

3.3 Interpretazione degli argomenti

Conclusa la fase preliminare, vengono interpretati gli argomenti passati al programma, attraverso i quali è possibile:

- visualizzare un messaggio sul terminale che elenca i possibili argomenti e le loro funzioni, e successivamente terminare l'esecuzione
- abilitare la modalità di debug, che permette di mostrare sul terminale dei messaggi relativi all'esecuzione del programma, utili per la risoluzione dei problemi
- avviare esclusivamente l'esecuzione dei test
- mostrare la versione del programma, e terminarne l'esecuzione
- abilitare la modalità `no-headless`, che modifica le impostazioni di Selenium per mostrare il browser usato dalla libreria

3.4 Antialiasing

L'antialiasing è una tecnica che permette di ridurre l'aliasing, l'effetto che si ottiene quando un segnale a bassa risoluzione viene mostrato ad una risoluzione maggiore. Per ottenere tale risultato su ogni piattaforma, abbiamo applicato diverse procedure:

- abbiamo modificato delle proprietà di sistema
- abbiamo creato dei componenti che ereditano da quelli predefiniti di Swing, effettuando l'override il metodo `paintComponent`
- abbiamo sfruttato l'implementazione dell'antialiasing di default di Flatlaf

3.5 Ricerca delle dinastie in locale

All'avvio del programma, verrà controllata la presenza della cartella `data/jsons` nella stessa directory dell'applicazione; nel caso in cui venga trovata, il programma proverà a caricare i file JSON contenuti in essa, se presenti. Se il caricamento è avvenuto con successo, la finestra permetterà all'utente di avviare il programma utilizzando le dinastie salvate in locale. Nell'eventualità in cui la cartella sia assente, o contenga file corrotti, la finestra di avvio del programma verrà modificata per permettere all'utente di avviare l'applicazione solo dopo aver effettuato il download delle dinastie in locale.

3.6 [+++] Scraping delle dinastie

3.6.1 [+++] Da Wikipedia utilizzando Selenium

3.6.2 [+++] Da Wikipedia utilizzando richieste HTTP

3.6.3 [+++] Console di download in tempo reale

3.7 [+++] Serializzazione delle dinastie

3.7.1 [+++] Algoritmo di serializzazione

3.7.2 [+++] Salvataggio e caricamento delle dinastie

3.8 [+++] Ricerca di persone e dinastie

3.8.1 [+++] Algoritmo di ricerca

3.8.2 [+++] Suggerimenti di ricerca

3.8.3 [+++] Visualizzazione dei risultati

3.9 [+++] Grafico delle dinastie

3.9.1 [+++] Struttura del grafico

3.10 [+++] Referenti dello sviluppo

Capitolo 4

Manuale della grafica

4.1 Finestra di avvio

All'avvio del programma, la prima schermata che viene mostrata permette all'utente di scegliere se **caricare** le dinastie **salvate in locale**, qualora ve ne fossero, oppure **scaricarle dal web** tramite **scraping**.

4.1.1 Dinastie non trovate

Nel caso in cui i dati delle dinastie **non siano** salvati in locale, ad esempio al primo avvio, verrà mostrata la seguente schermata:

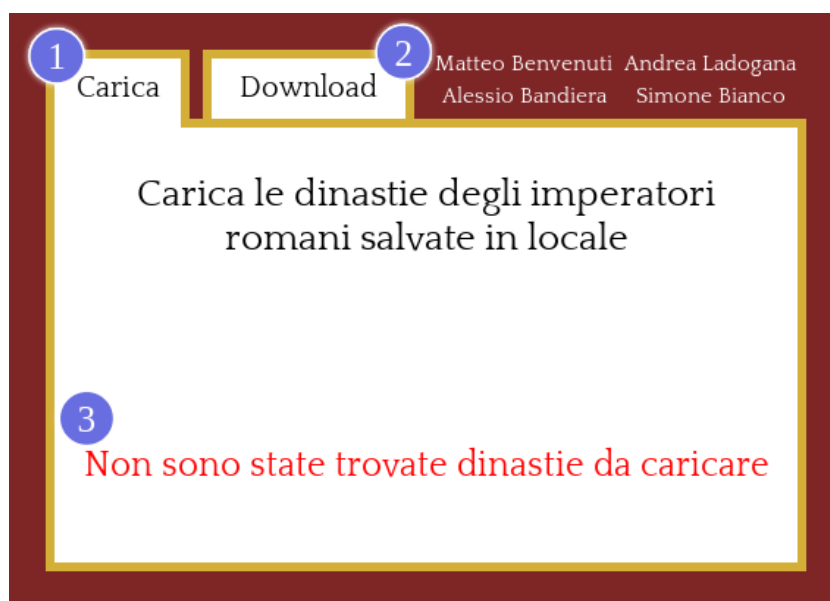


Figura 4.1: Schermata "Carica" con file delle dinastie non trovati.

1. Il pulsante "Carica" mostra la schermata di caricamento.
2. Il pulsante "Download" mostra la schermata di download.
3. Questo messaggio indica che non sono state trovate dinastie nella directory del programma.

4.1.2 Download delle dinastie

Nella pagina di download, invece, si potranno **scaricare** i dati delle dinastie e **salvarli in locale**:

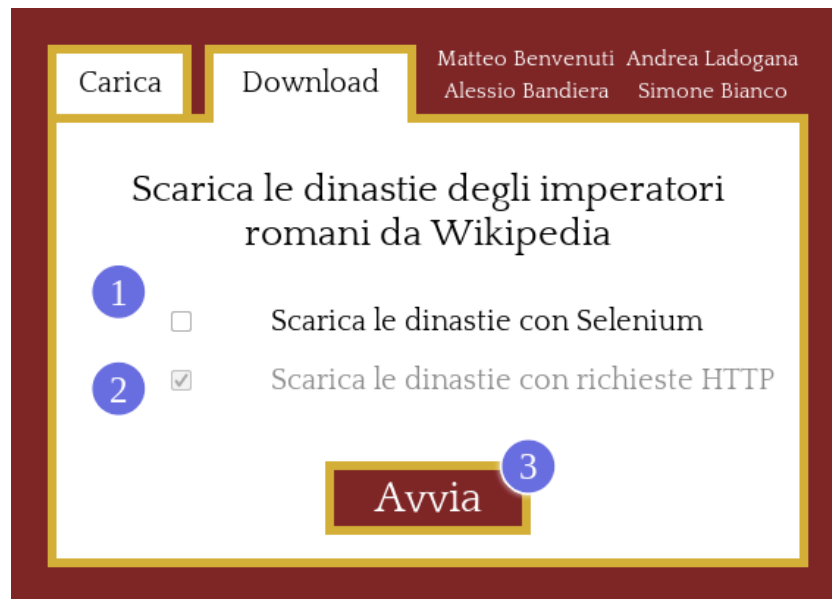


Figura 4.2: Schermata di download delle dinastie.

1. Questa casella, se spuntata, indica che il download verrà effettuato utilizzando Selenium.
2. Questa casella, se spuntata, indica che il download verrà effettuato utilizzando le richieste HTTP.
3. Il pulsante "Avvia" farà partire la finestra principale del programma, mostrando il download ed il salvataggio delle dinastie in locale tramite una console:



Figura 4.3: Console per la visualizzazione del download delle dinastie.

4.1.3 Dinastie trovate

Nel caso in cui i dati delle dinastie **vengano trovati** nella directory del programma, la schermata "Carica" apparirà in questo modo:



Figura 4.4: Schermata "Carica" con file delle dinastie caricati.

1. Questo messaggio indica che le dinastie sono state caricate correttamente.
2. Il pulsante "Avvia" farà partire la finestra principale del programma utilizzando i dati trovati in locale, quindi senza eseguire nessun download.

4.2 Finestra di ricerca

Una volta caricate le dinastie all'interno del programma, dopo aver premuto il pulsante "Avvia", apparirà la **finestra di ricerca**, tramite la quale sarà possibile **cercare le persone** all'interno delle dinastie caricate e visualizzarne l'**albero genealogico**.

4.2.1 Struttura generale

Una volta avviata, la finestra principale si presenta divisa in diverse sezioni:



Figura 4.5: Finestra di ricerca con grafico non caricato (all'avvio).

1. La barra di ricerca, nella quale si potrà inserire il nome o la dinastia che si vuole cercare; per effettuare la ricerca premere il tasto invio, oppure cliccare la lente di ingrandimento.
2. Dopo aver effettuato la ricerca, il suo esito verrà mostrato nel pannello dei risultati; se il numero dei risultati è troppo elevato, sulla destra apparirà una barra di scorrimento che, se trascinata verso il basso, permetterà di visualizzare ulteriori risultati; il pannello può essere navigato anche usando la rotella del mouse.
3. Se non è stato caricato nessun grafico, il pannello di visualizzazione mostrerà una schermata contenente suggerimenti riguardanti l'utilizzo del programma.

4.2.2 Suggerimento di ricerca

In caso di ricerca senza alcun risultato, il programma mostrerà un suggerimento di ricerca:

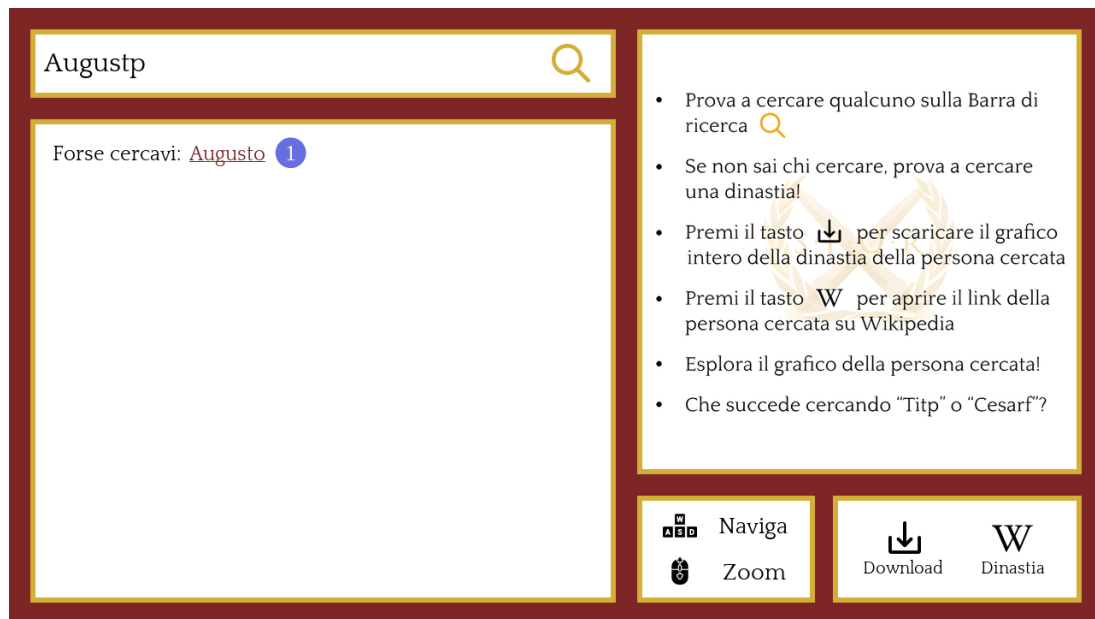


Figura 4.6: Pannello dei risultati con suggerimento.

1. Una volta cliccato il suggerimento (in rosso), il programma lo ricercherà e ne mostrerà i risultati.

In caso non sia possibile generare suggerimenti, verrà mostrata la seguente schermata:



Figura 4.7: Finestra di ricerca con grafico caricato

4.2.3 Risultati di ricerca e visualizzazione del grafico

La finestra di ricerca del programma si presenta in questo modo:

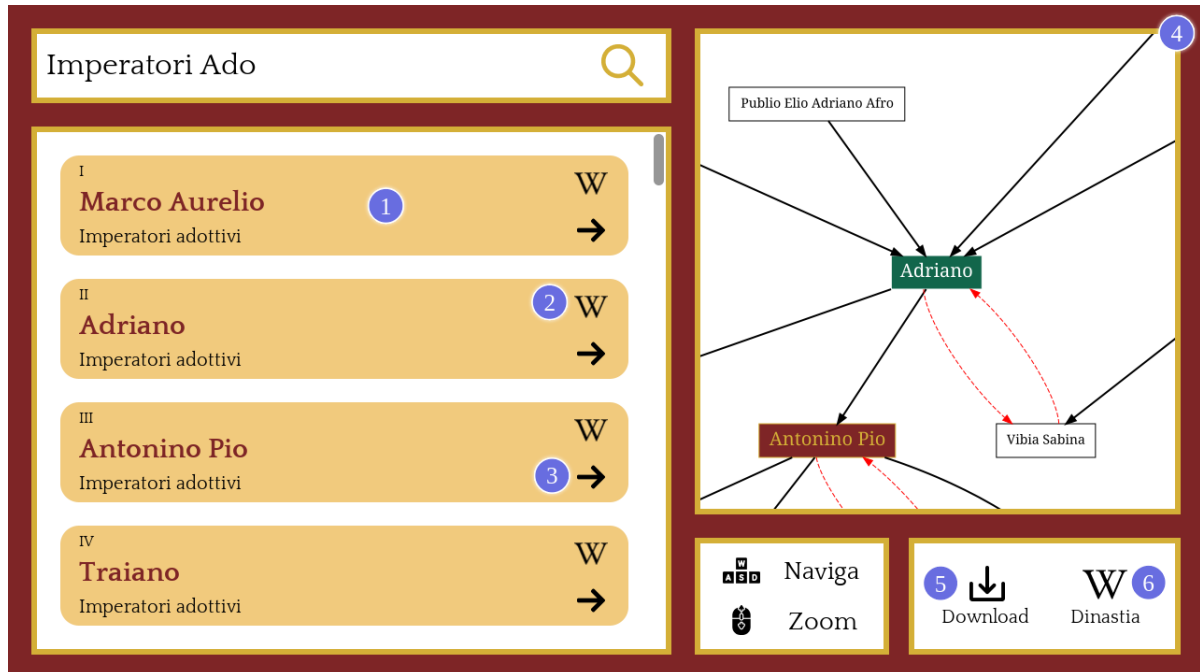


Figura 4.8: Finestra di ricerca con grafico caricato.

1. I risultati appaiono in delle card che mostrano i nomi delle persone e le rispettive dinastie di appartenenza; il nome degli imperatori è colorato in rosso, inoltre i risultati sono ordinati per importanza.
2. Il pulsante raffigurante una "W", se cliccato, porterà alla pagina di Wikipedia della persona corrispondente.
3. Il pulsante raffigurante una freccia caricherà nel pannello di visualizzazione l'albero genealogico della persona cercata.
4. Il pannello di visualizzazione mostra il grafico della persona cercata, e gli imperatori appartenenti all'albero genealogico sono mostrati in rosso; i legami di parentela tra genitori e figli è indicato con frecce nere (anche per genitori adottivi), mentre i legami coniugali sono rappresentati con frecce rosse tratteggiate; infine, il grafico viene mostrato centrato sulla persona cercata, colorata in verde.
5. Questo pulsante permette di salvare l'immagine della dinastia *completa* a cui appartiene la persona cercata.
6. Questo pulsante, raffigurante una "W", se cliccato, porterà alla pagina di Wikipedia della dinastia a cui appartiene la persona cercata.