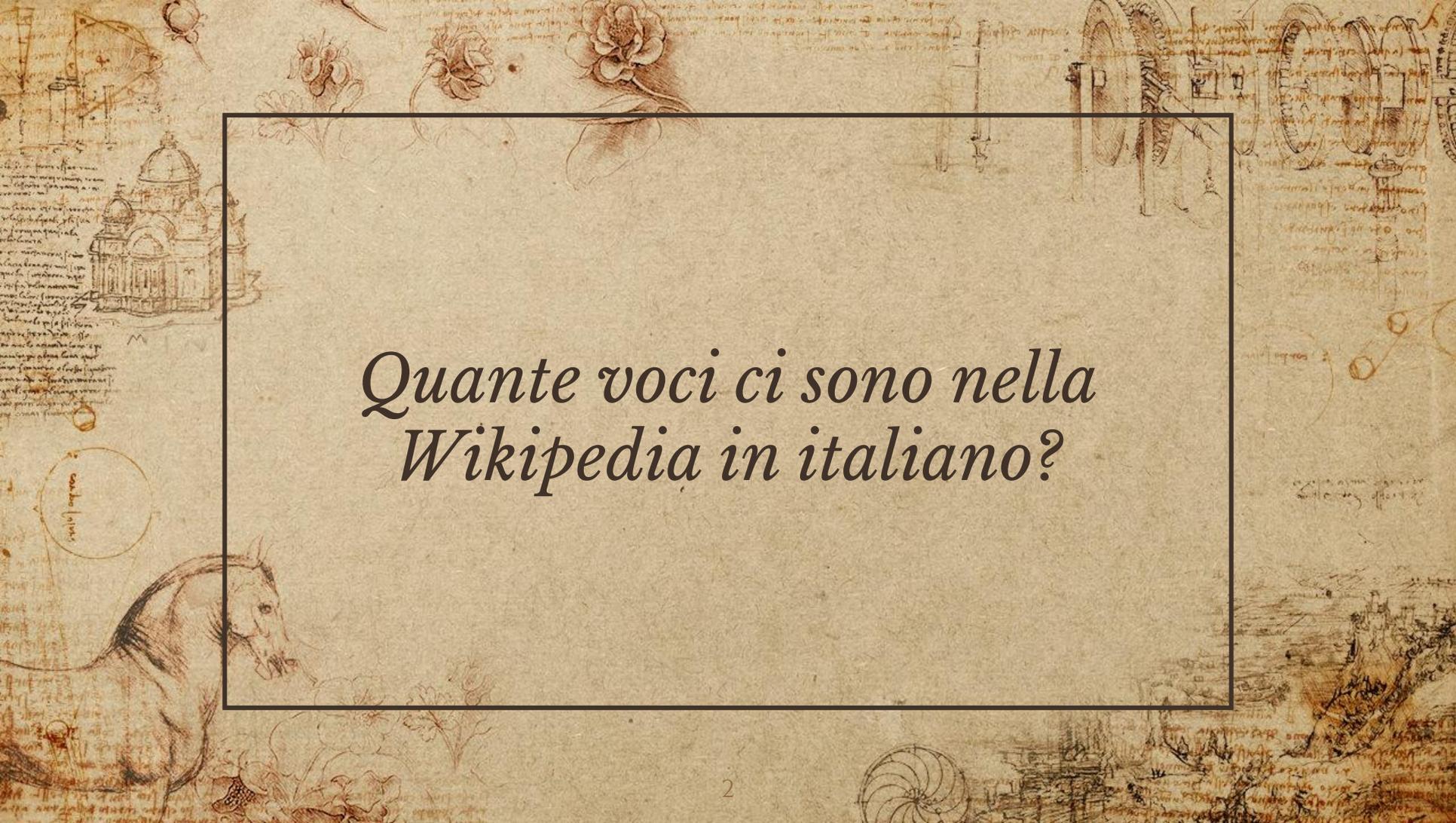




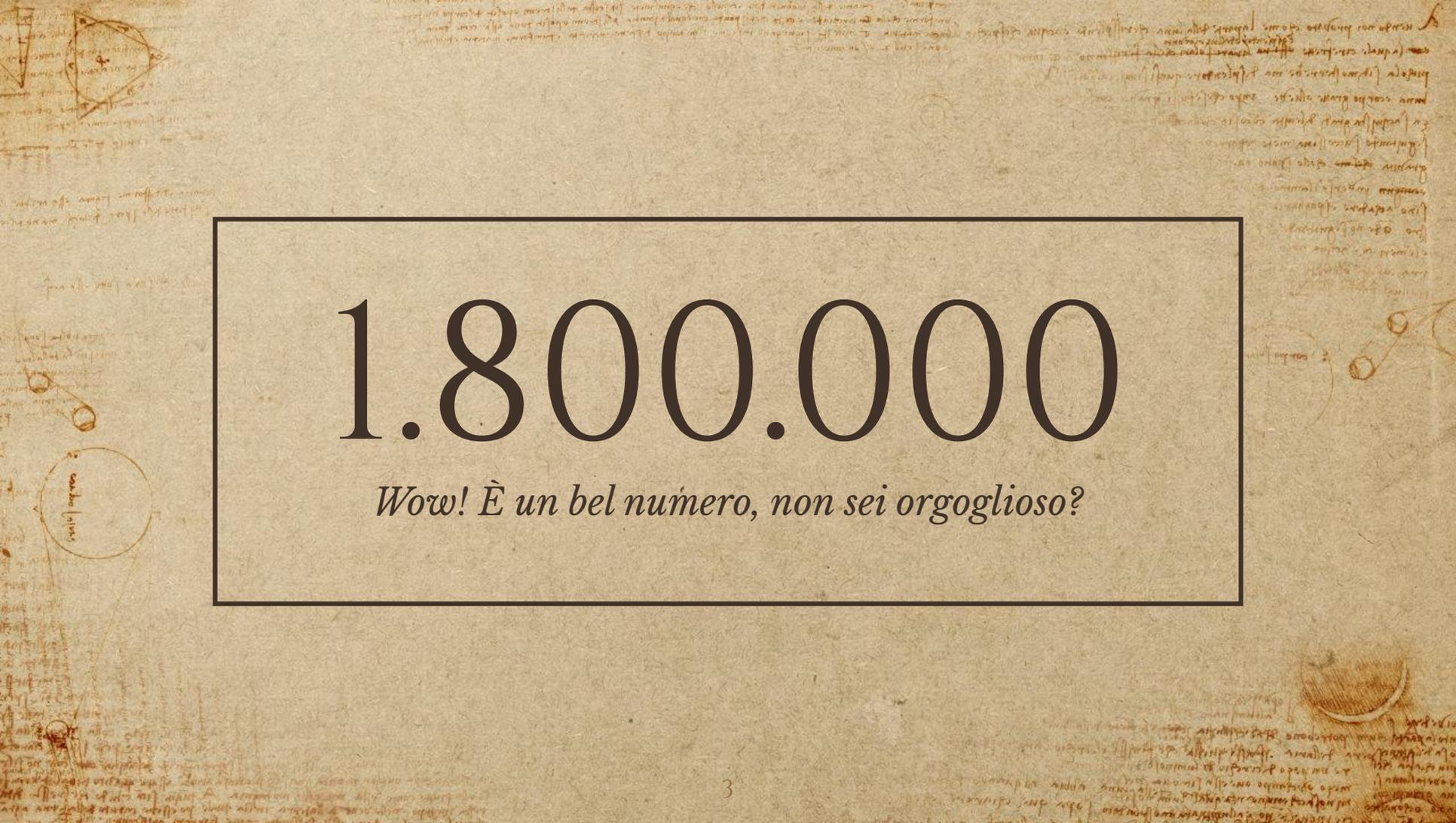
LA SOMMA DI TUTTA LA CONOSCENZA UMANA

A che punto siamo.

ITWIKICON
SETTEMBRE 2022

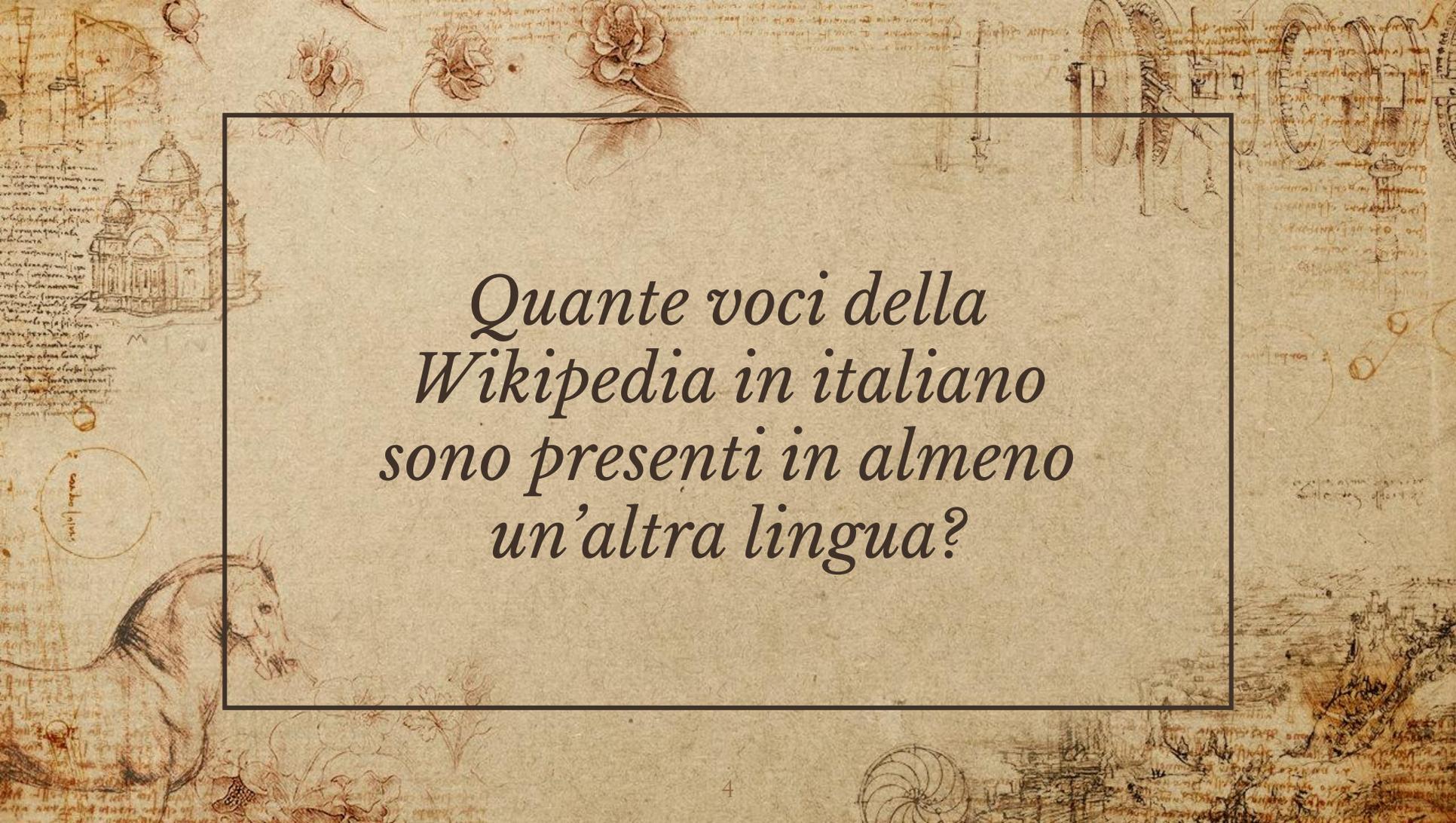


*Quante voci ci sono nella
Wikipedia in italiano?*

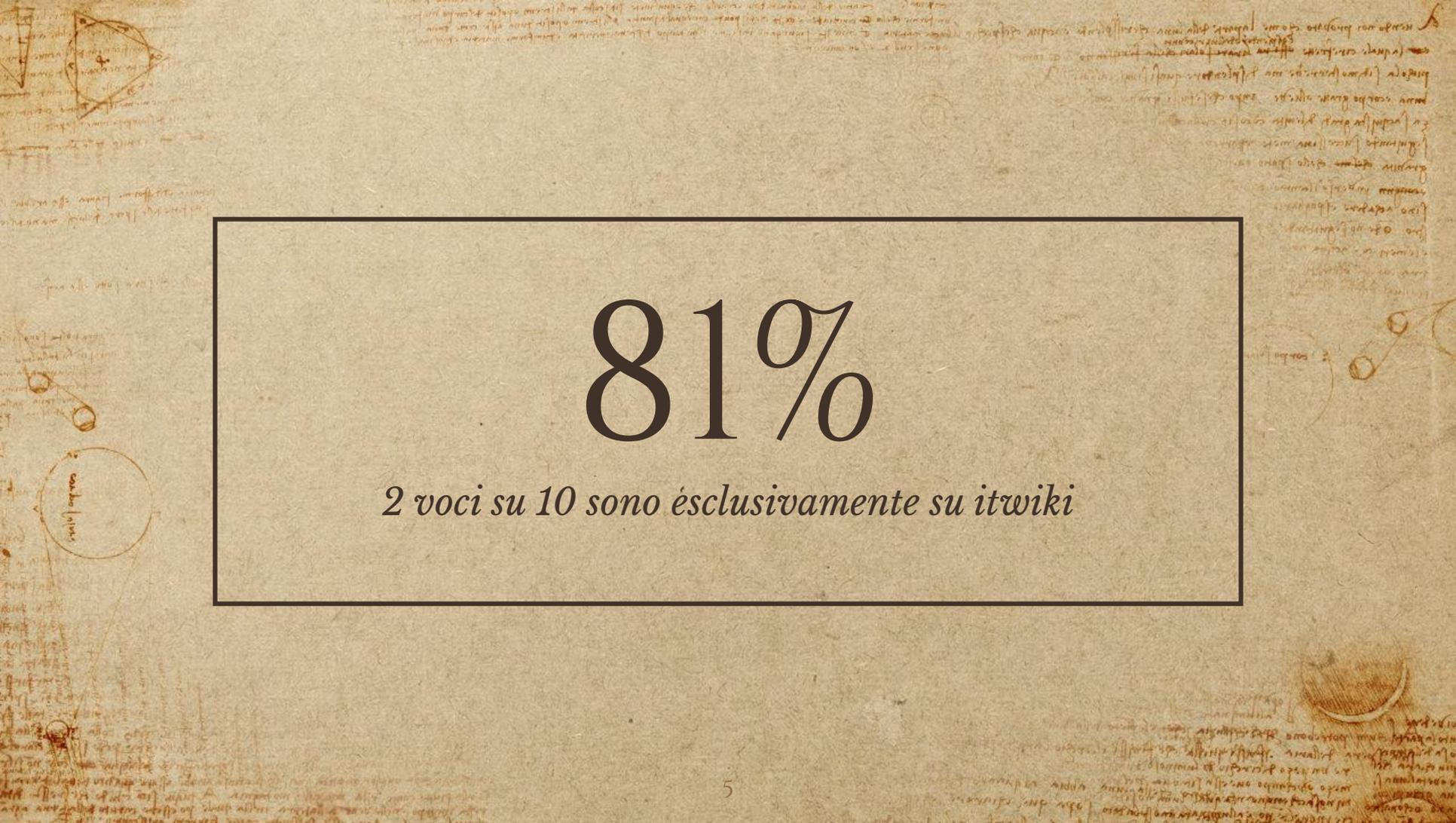
The background is a page from an old manuscript, likely a notebook or a treatise, featuring handwritten text in a cursive script and several diagrams. One diagram in the top left shows a complex geometric shape, possibly a triangle with internal lines. Another diagram in the middle left shows a circle with a vertical line through its center. The text is written in a dark ink, and the paper has a yellowish, aged appearance.

1.800.000

Wow! È un bel numero, non sei orgoglioso?



*Quante voci della
Wikipedia in italiano
sono presenti in almeno
un'altra lingua?*



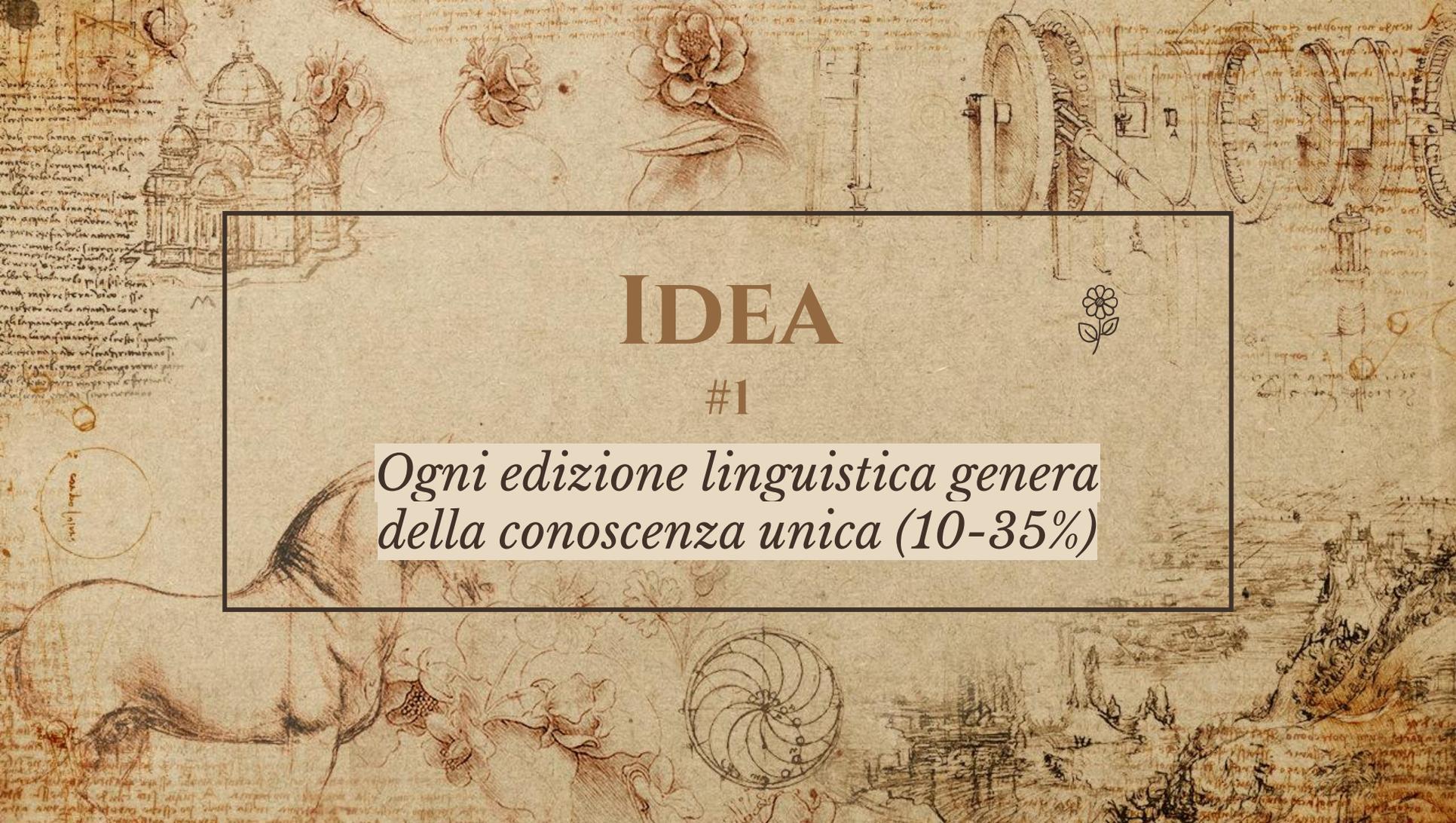
81%

2 voci su 10 sono esclusivamente su itwiki



E LE ALTRE EDIZIONI LINGUISTICHE?

EDIZIONE	VOCI <u>ANCORA</u> UNICHE	VOCI ANCHE IN ALTRE LINGUE
dewiki - 3°	34%	66%
enwiki - 1°	33%	67%
plwiki - 10°	23%	77%
ruwiki - 7°	21%	79%
itwiki - 9°	19%	81%
frwiki - 5°	19%	81%
eswiki - 8°	16%	84%
nlwiki - 6°	16%	84%
svwiki - 4°	10%	90%
cebwiki - 2°	55%	45%

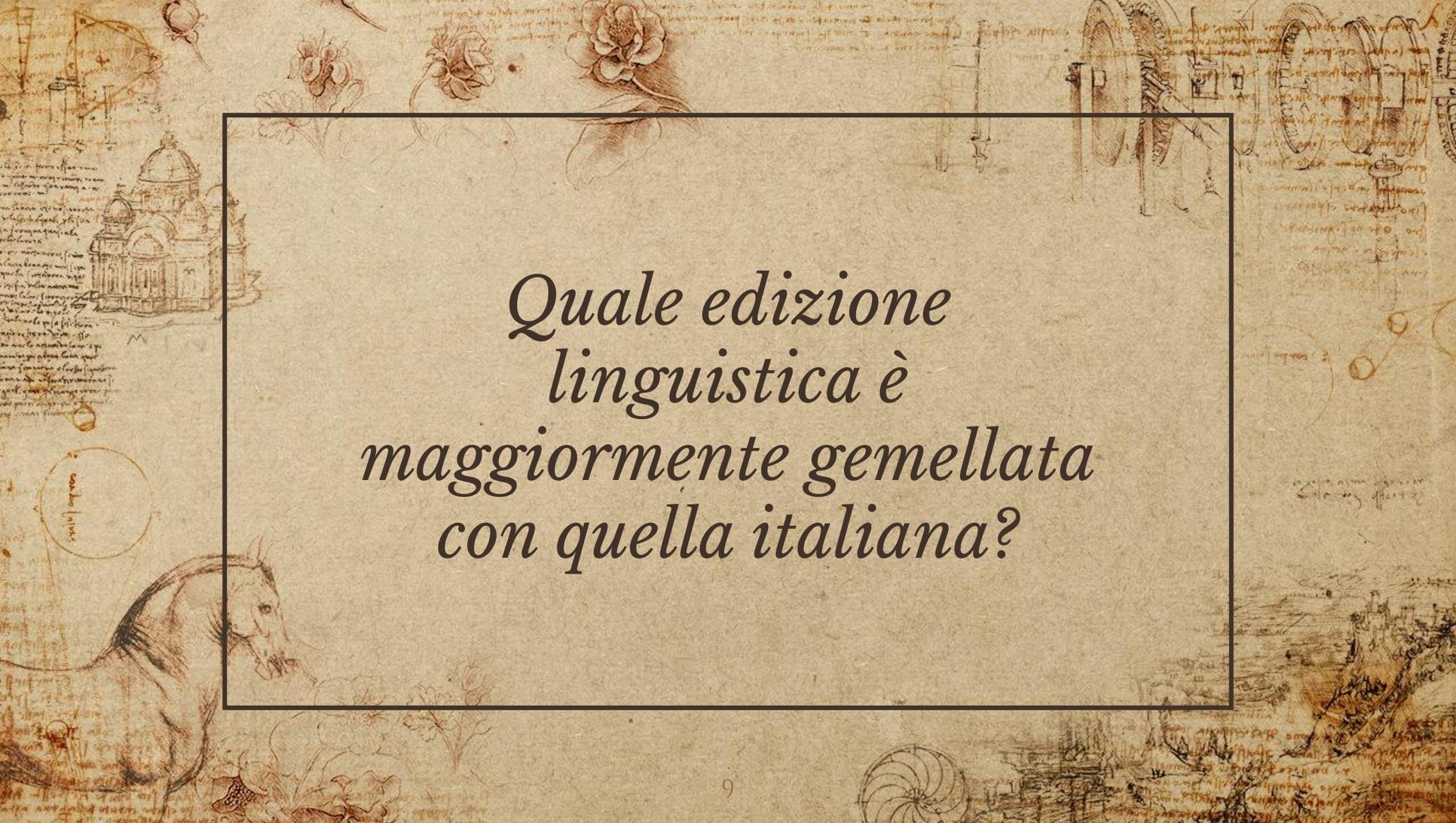


IDEA

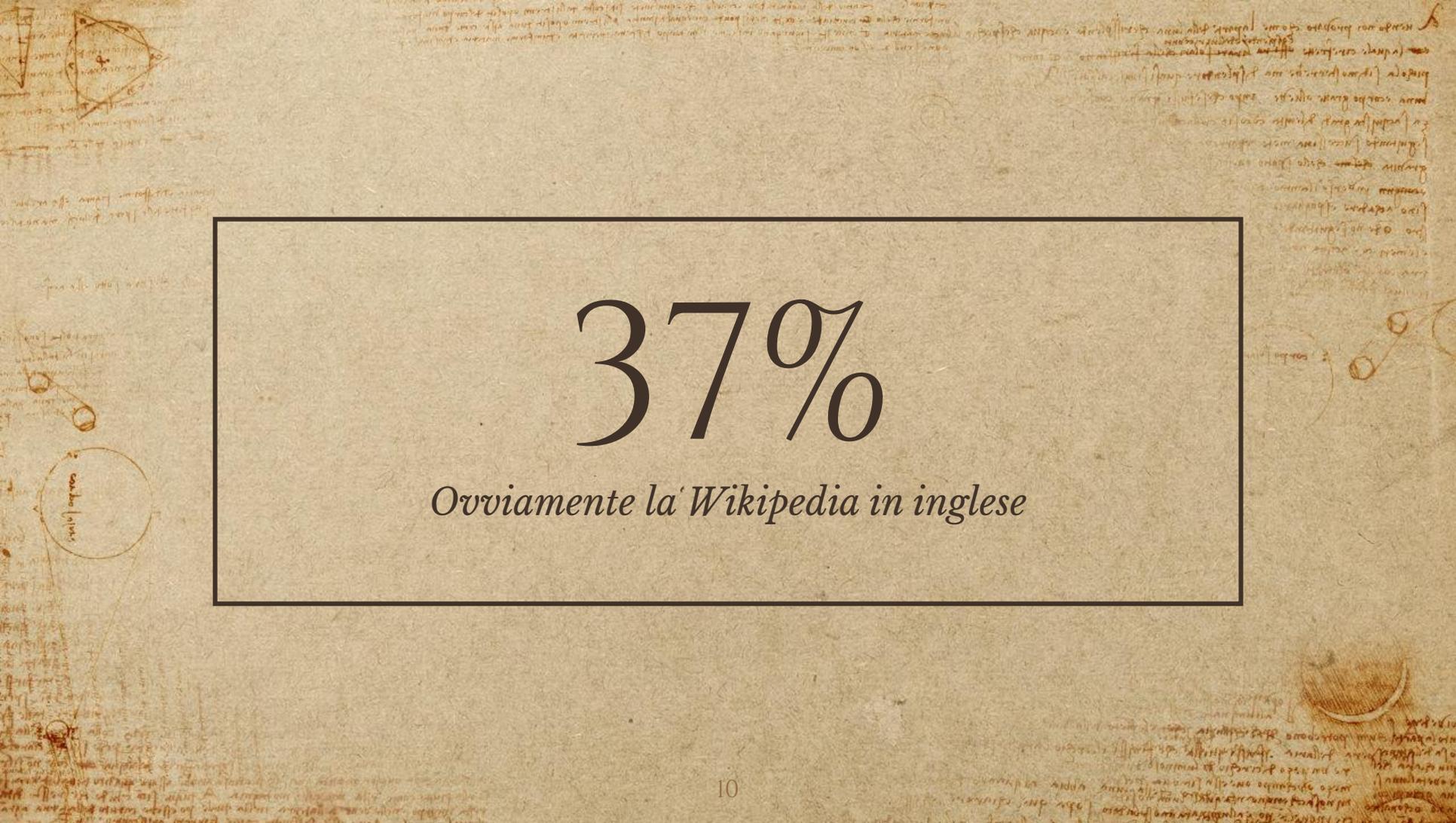
#1



*Ogni edizione linguistica genera
della conoscenza unica (10-35%)*

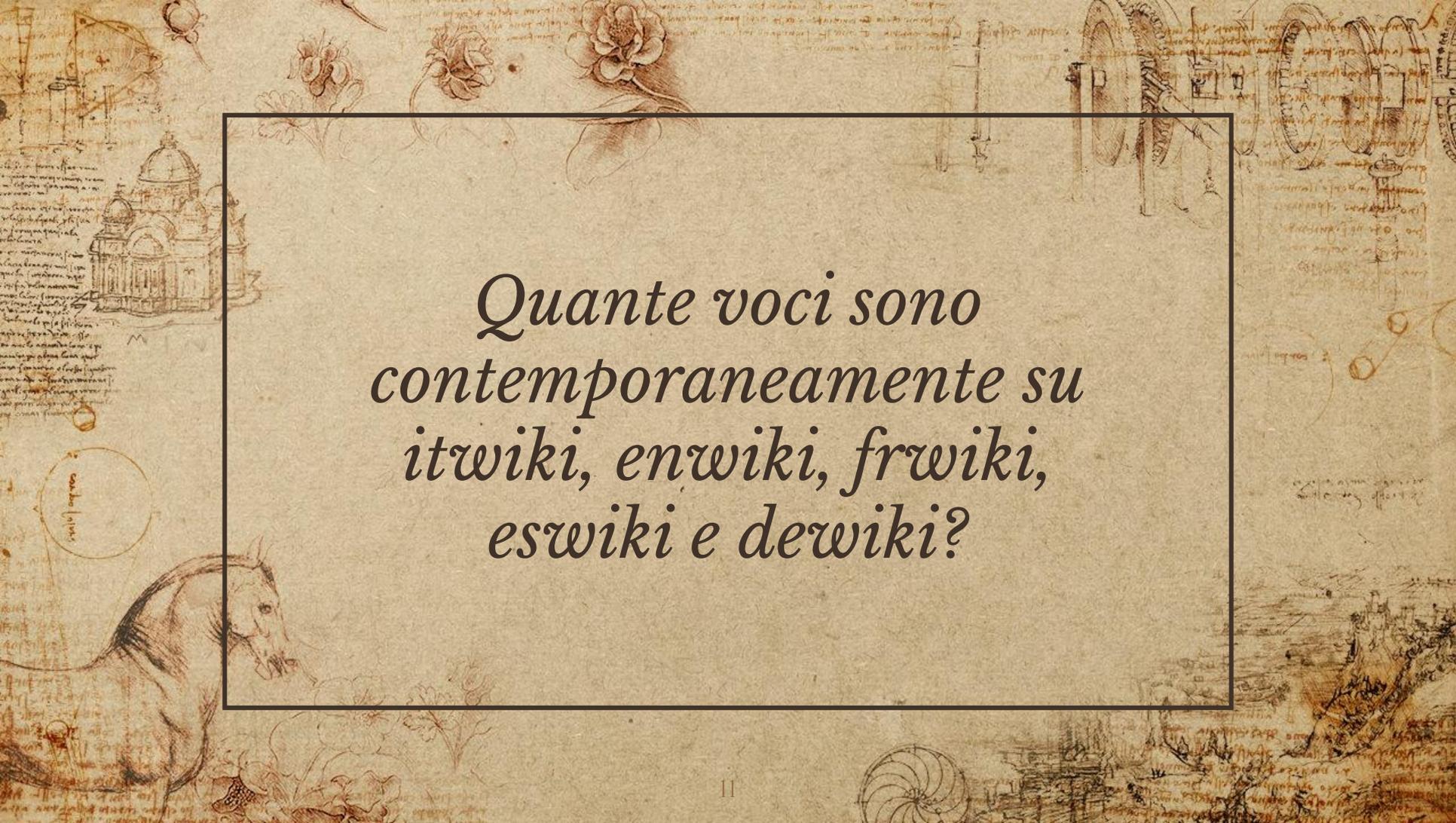


*Quale edizione
linguistica è
maggiormente gemellata
con quella italiana?*



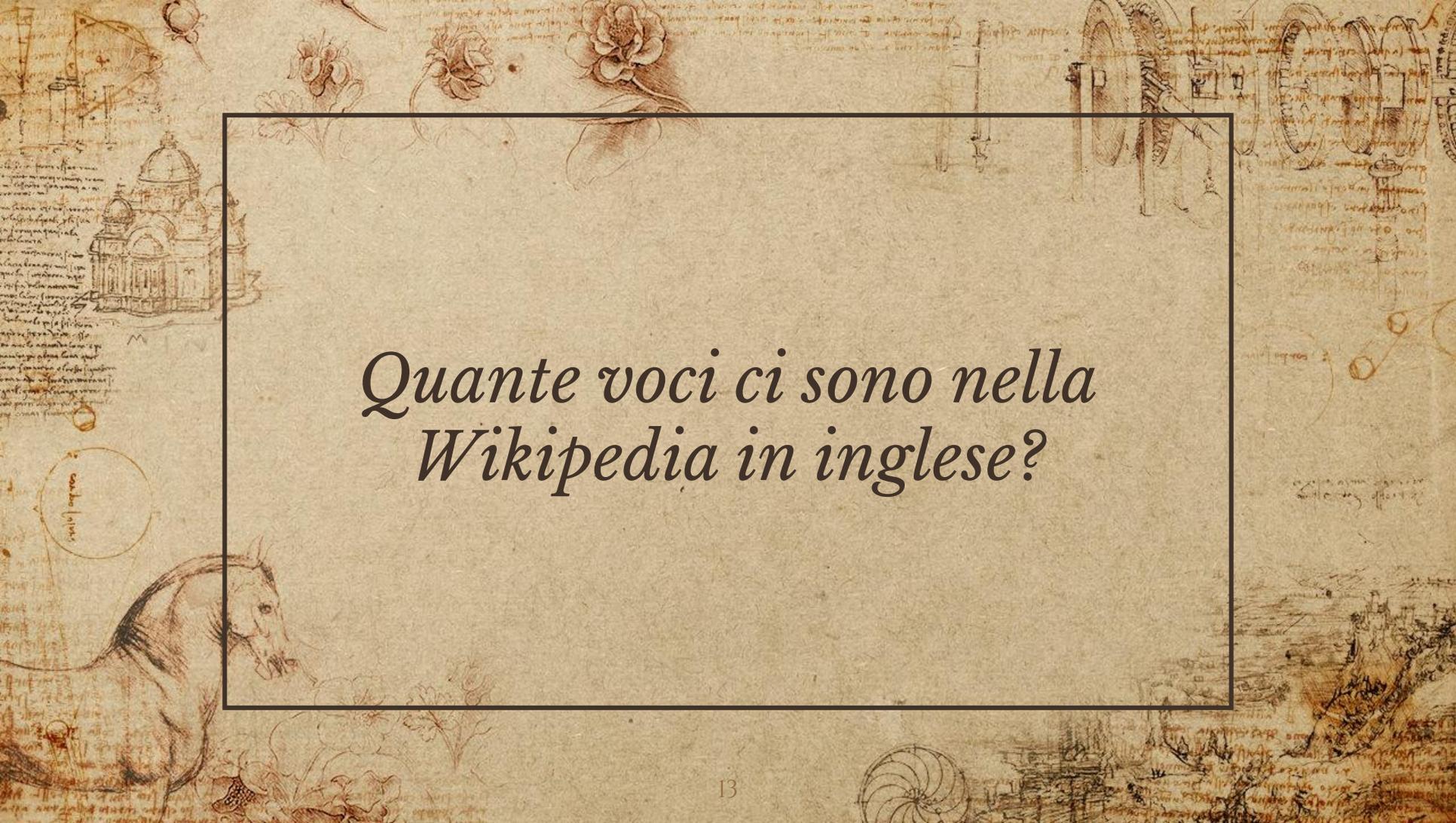
37%

Orvviamente la Wikipedia in inglese



*Quante voci sono
contemporaneamente su
itwiki, enwiki, frwiki,
eswiki e dewiki?*

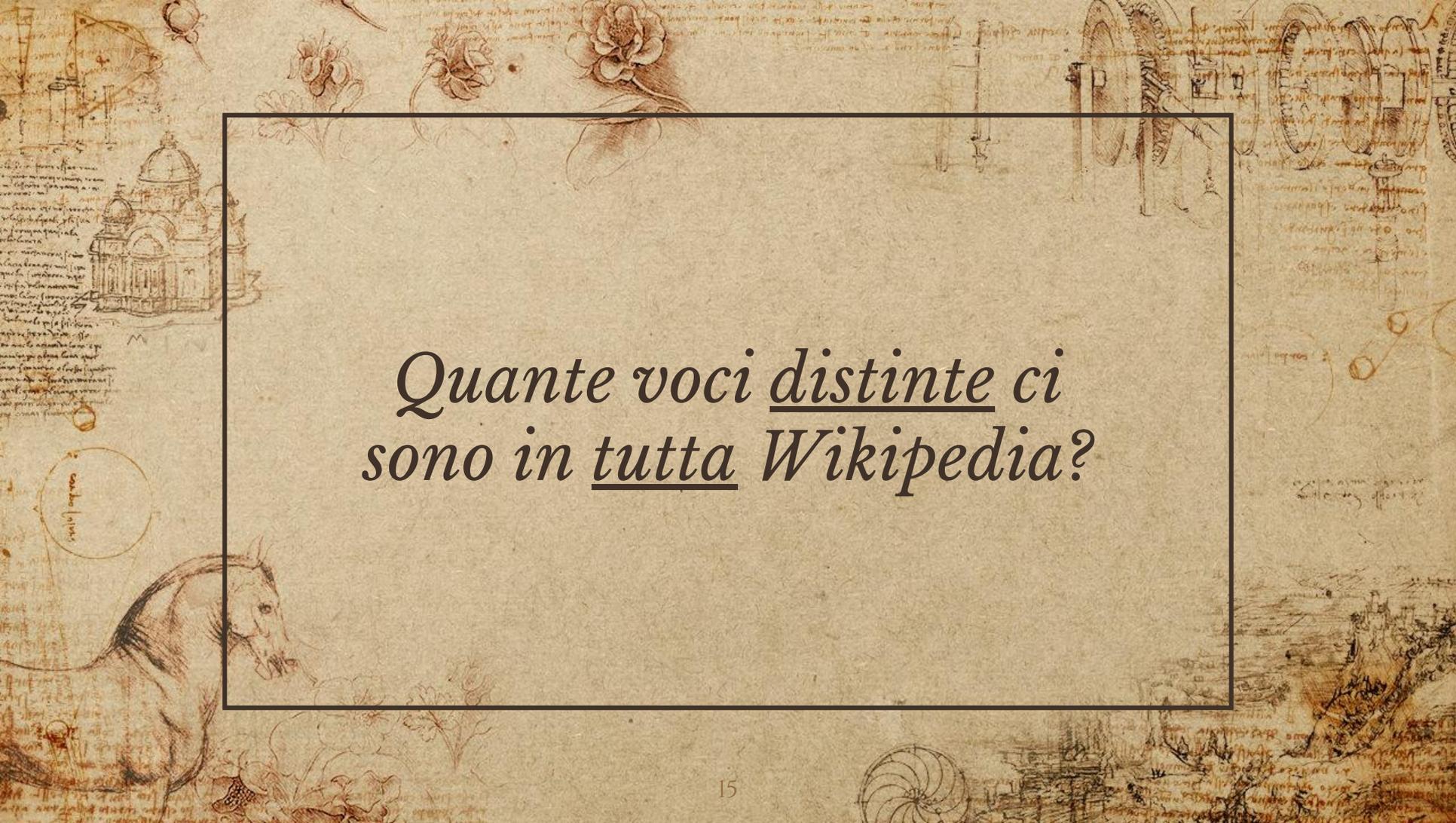
<i>COPPIA</i>	<i>VOCI IN COMUNE</i>	<i>% ITWIKI</i>	<i>% ALTRA EDIZIONE</i>
itwiki - enwiki	670.000	37%	10%
itwiki - frwiki	490.000	27%	20%
itwiki - dewiki	420.000	23%	16%
itwiki - eswiki	400.000	22%	22%
itwiki - nlwiki	330.000	19%	16%
itwiki - plwiki	330.000	19%	22%
itwiki - svwiki	290.000	16%	12%
itwiki - ruwiki	45.000	3%	3%
it - en - fr	410.000	22%	N.D.
it - en - de - fr - es	210.000	12%	N.D.



*Quante voci ci sono nella
Wikipedia in inglese?*

6.600.0000

Whoa! That's a big number, aren't you proud?



*Quante voci distinte ci
sono in tutta Wikipedia?*

The background is a scan of a manuscript page, likely from a historical scientific or mathematical text. It features handwritten text in a cursive script, possibly Italian or Latin, and several diagrams. In the top left, there is a complex geometric diagram with lines and points. In the bottom right, there is a circular diagram with internal lines. The overall appearance is that of an aged, historical document.

13.000.000

Considerando le top 9 edizioni

23.000.000

Considerando tutte le edizioni edizioni linguistiche

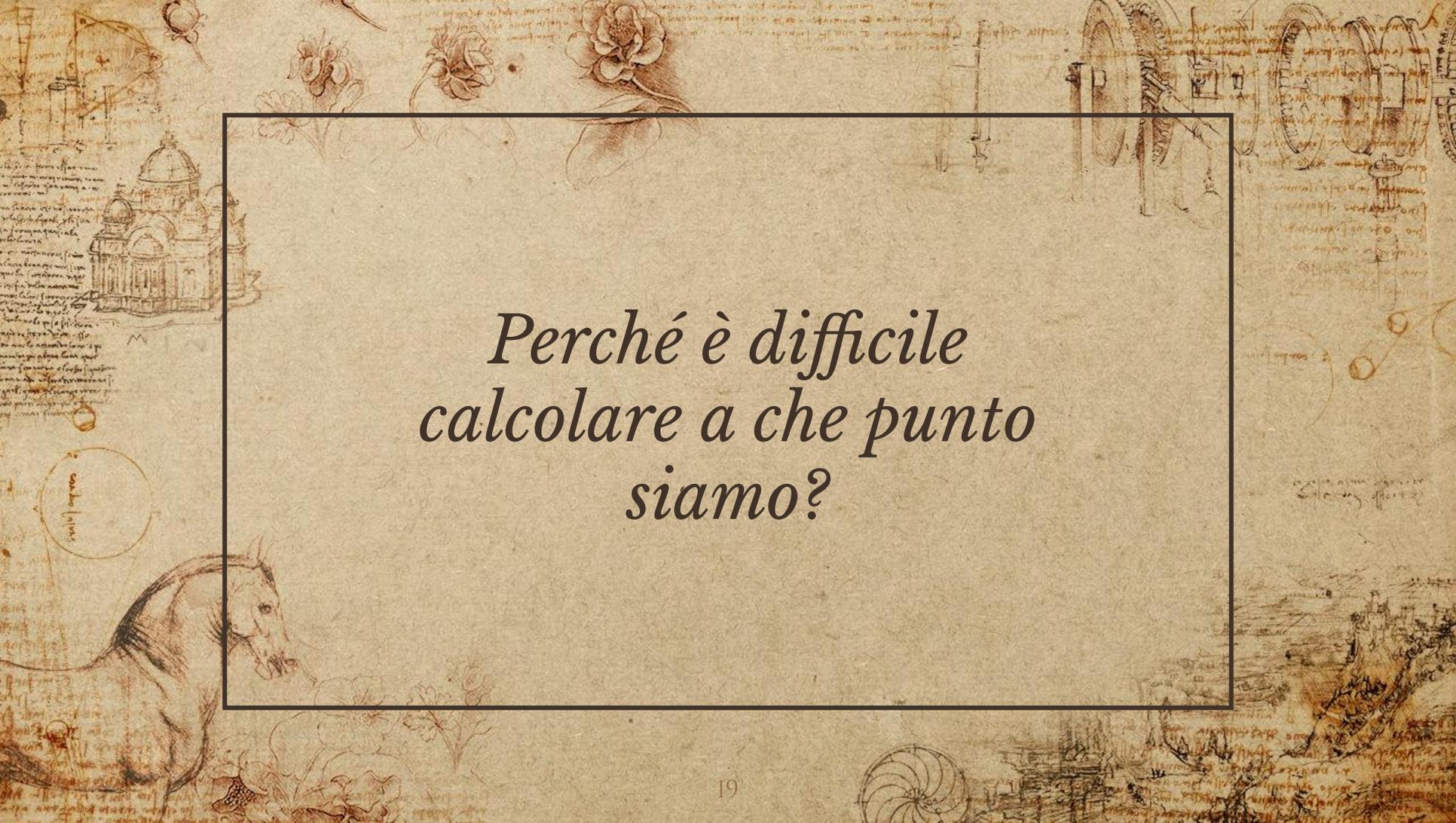
N.B. non è la somma dei vari counter!

IDEA

#2

Esiste una Wikipedia “complessiva & poliglotta” grande 3 volte quella inglese



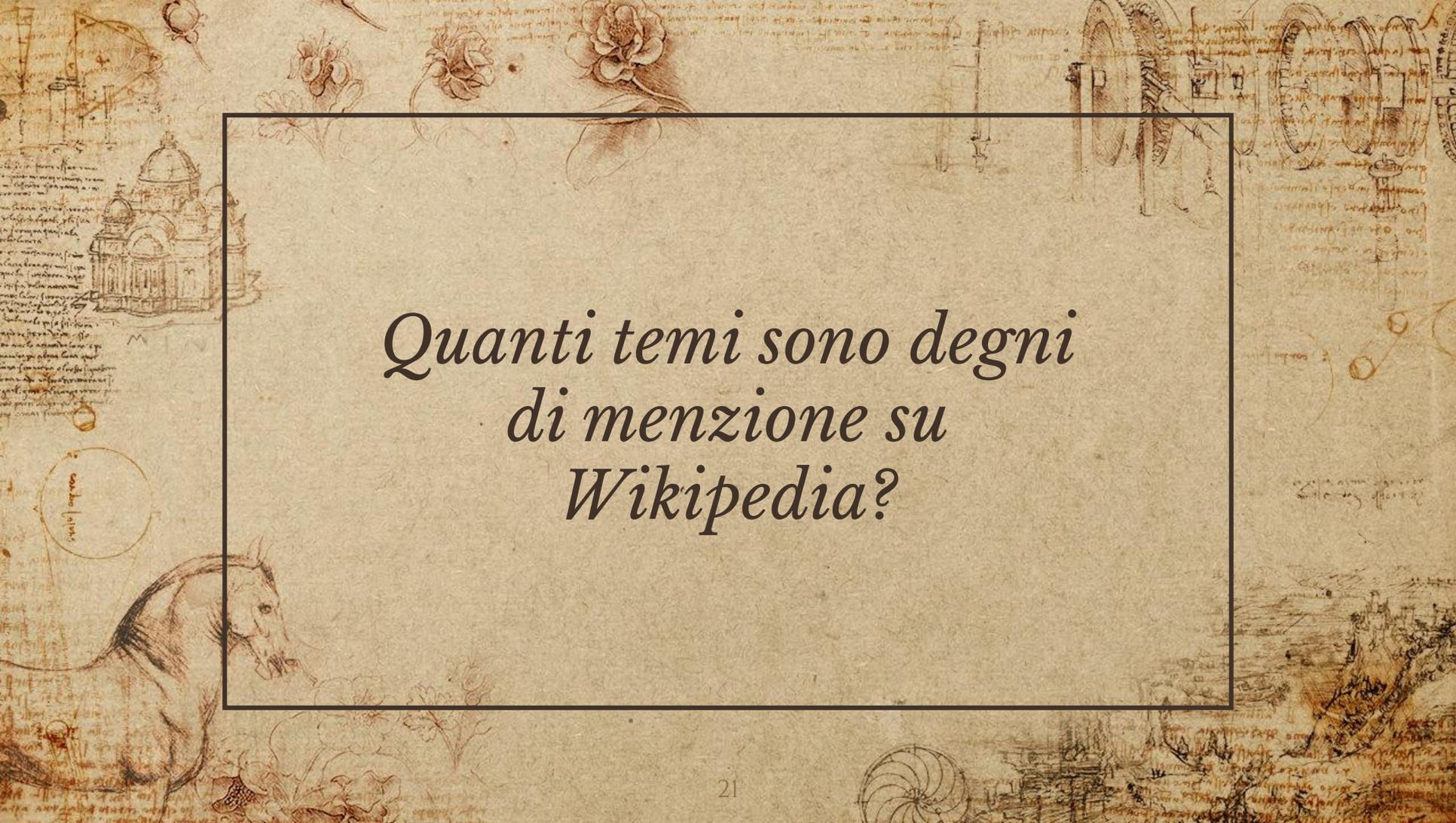
The background of the slide is a collage of Leonardo da Vinci's sketches. At the top, there are several detailed drawings of flowers, including a large rose and smaller blossoms. To the left, there is a sketch of a domed building, possibly a church or cathedral. In the bottom left corner, a horse is depicted in profile, facing right. The right side of the background features various mechanical sketches, including gears and a complex device with a large wheel. The entire background is filled with faint, handwritten text in Leonardo's characteristic cursive script.

*Perché è difficile
calcolare a che punto
siamo?*

PER SAPERE A CHE PUNTO SIAMO OCCORRE

- ✓ Saper contare
- ✦ Conoscere il totale

Le cose che contano di più nella vita sono le cose che non possono essere contate.



*Quanti temi sono degni
di menzione su
Wikipedia?*

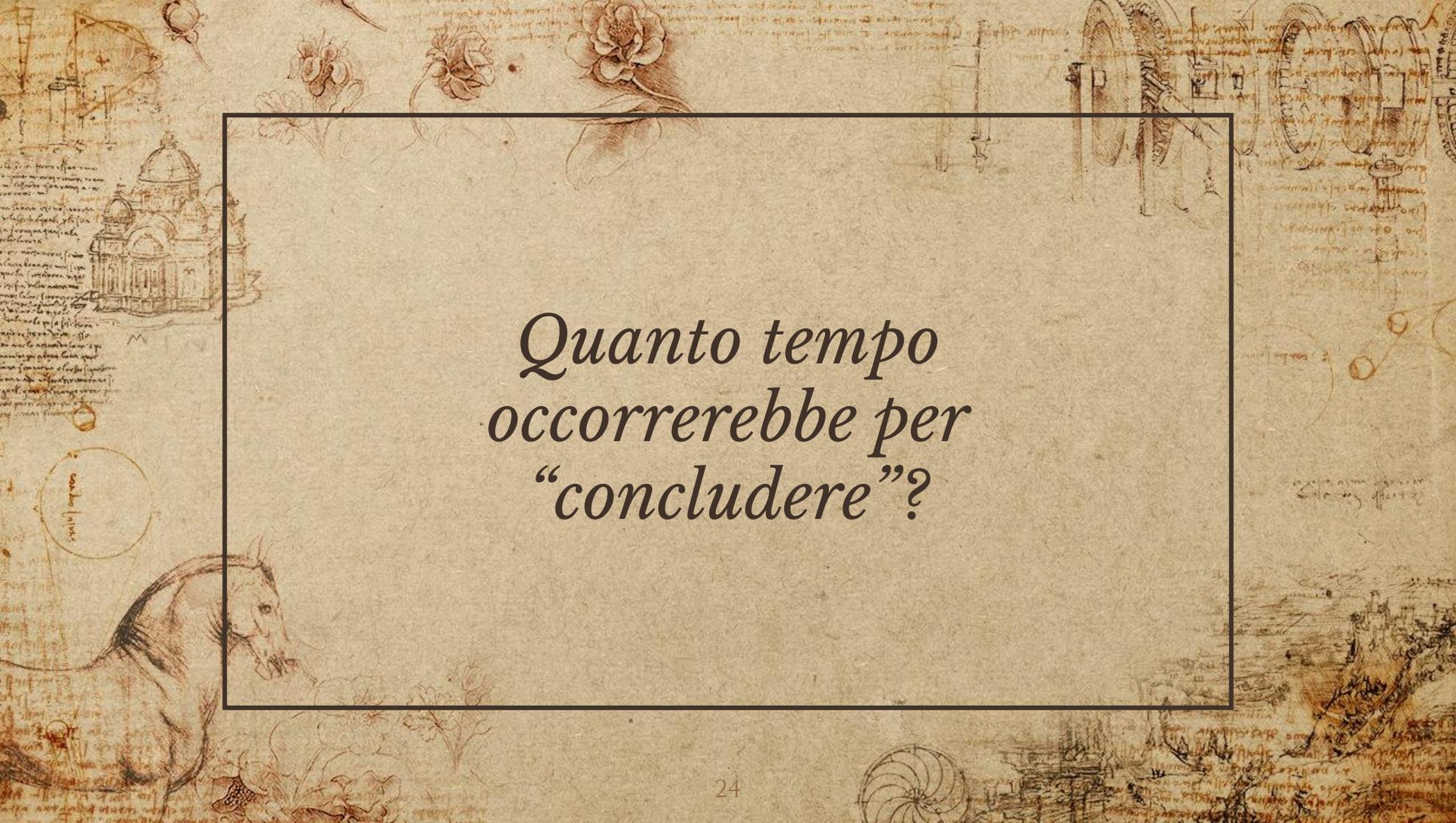
120.000.000+

Fonte:

https://www.wikidata.org/wiki/User:Emijrp/All_Human_Knowledge

LOADING... 19%

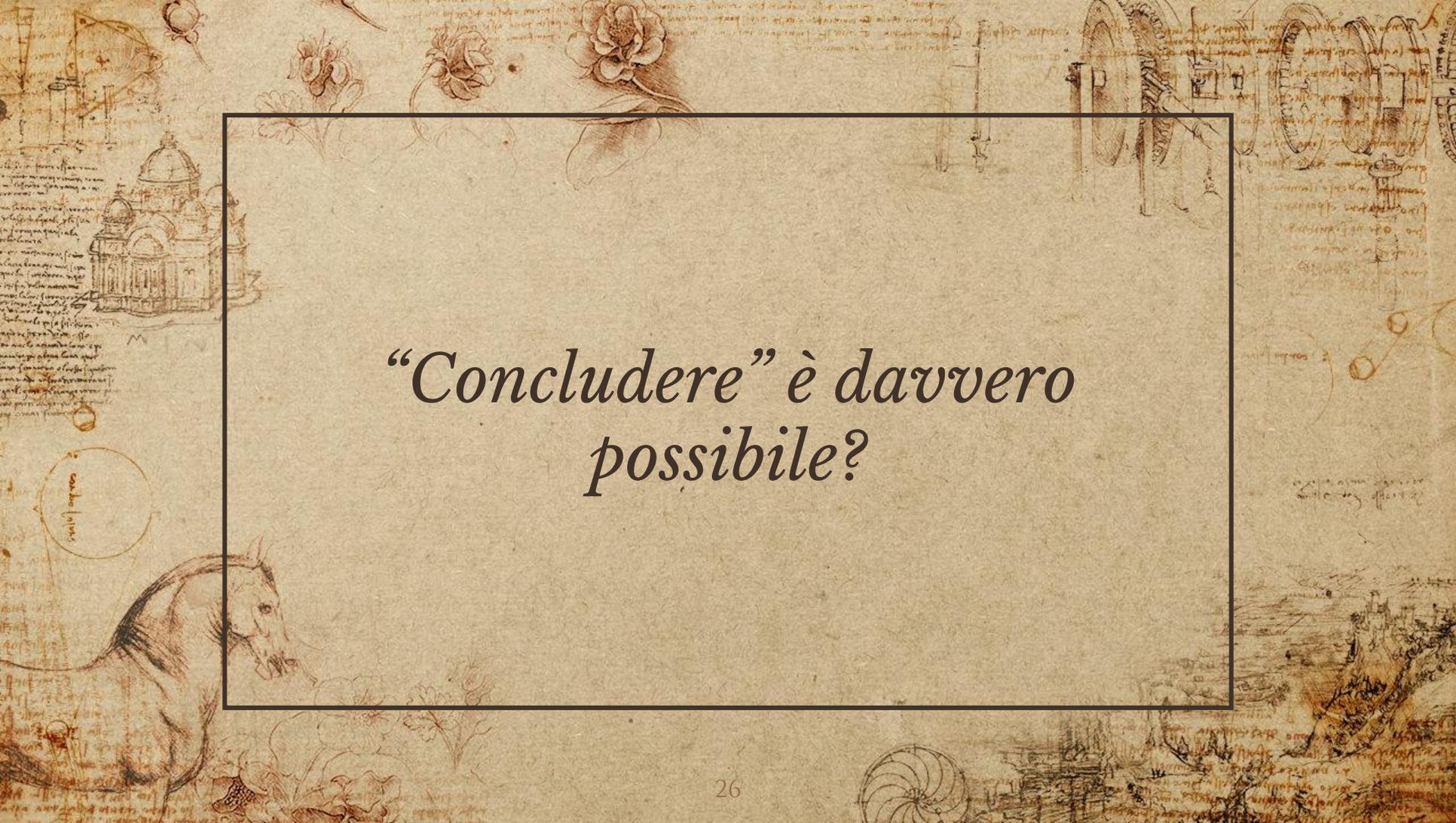
...se 'così fosse!



*Quanto tempo
occorrerebbe per
“concludere”?*

100 ANNI

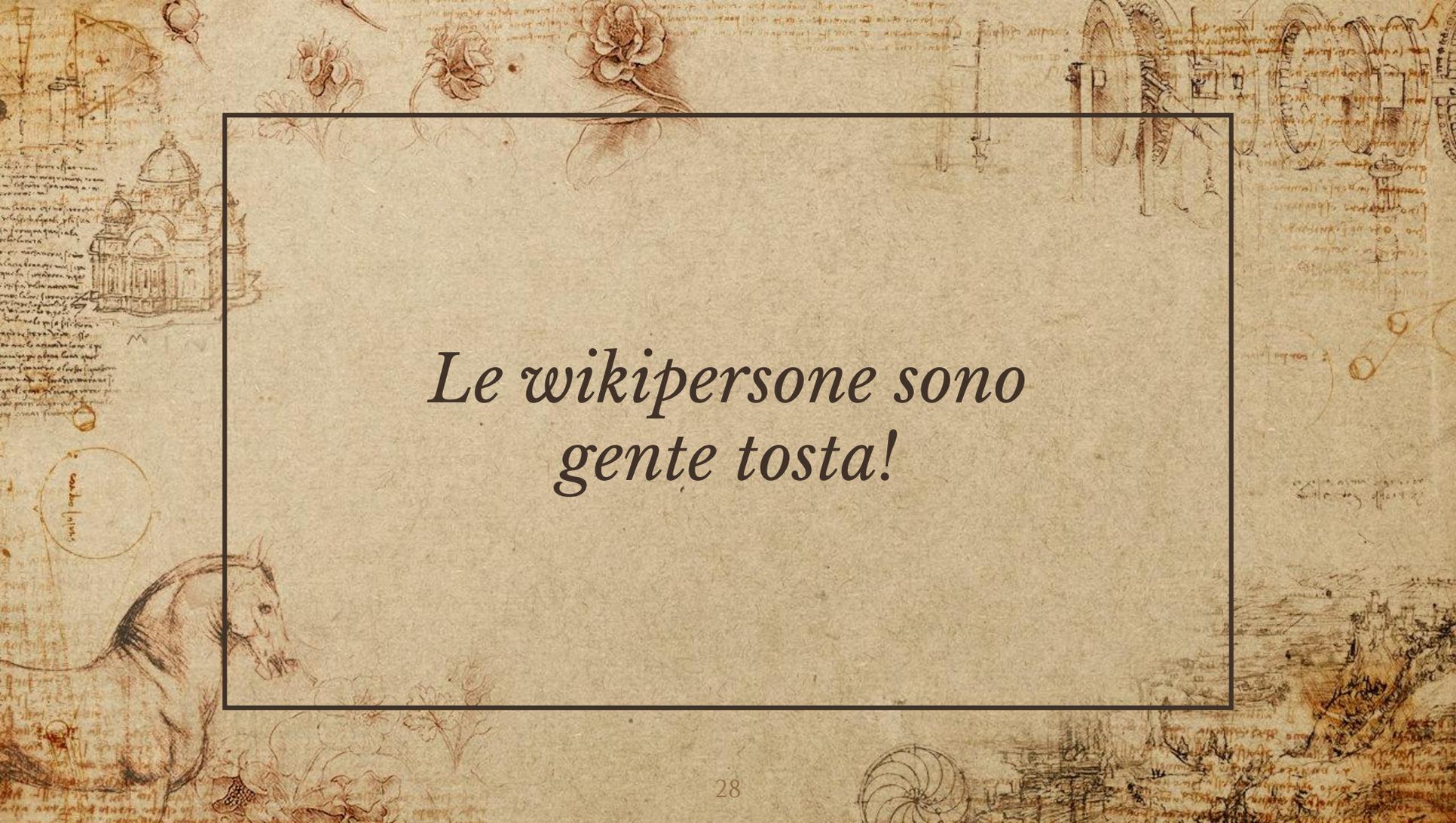
*...se la stima fosse corretta e il genere umano
smettesse di fare ogni cosa, tranne editare!*



*“Concludere” è davvero
possibile?*

PIÙ CONOSCI MENO CONOSCI

- ❖ Continuano a nascere persone interessanti
- ❖ Cose che continuiamo a creare (canzoni, film, opere d'arte, libri, etc.)
- ❖ Cose di cui ci rendiamo conto (teorie, idee, strati della realtà, etc.)
- ❖ Modi di colpire una sfera per divertimento :-)
- ❖ Qual livello di zoom ci soddisfa? (ogni stella, pianeta, asteroide? Ogni filo d'erba?)



*Le wikipersone sono
gente tosta!*

MISSION (IM)POSSIBLE

- ❖ Cose che al momento sono indefinite ma hanno una “fine” (e.g. i monumenti di uno Stato, gli incunaboli e le cinquecentine, etc.)
- ❖ Cose potenzialmente infinite ma Wikipedia è fonte secondaria perciò possiamo stare al passo (e.g. Pianeti “scoperti”, specie “trovate”, etc.)

IDEA

#3



*Le cose da conoscere sono infinite ma ci
“basta” stare al passo dello stato dell’arte*

GRAZIE!

Domande?



Utente: AlessioMela





APPENDICE

QUERY

*Per riprodurre i numeri che hai
visto anche in futuro!*

1.800.000 voci su itwiki

```
SELECT count(*)  
FROM page  
WHERE page_namespace = 0  
AND page_is_redirect = 0
```

81% delle voci esiste anche in un'altra lingua

```
SELECT COUNT(DISTINCT ll_from) /  
(SELECT COUNT(*) from page as p  
WHERE p.page_namespace = 0 AND p.page_is_redirect = 0) * 100 + '%'  
FROM langlinks as ll, page as p  
WHERE p.page_namespace = 0  
AND p.page_is_redirect = 0  
and ll.ll_from = p.page_id
```

13.000.000 di voci distinte sommando le “top 9” Wikipedia

TOP 10 escluso il Cebuano per la troppa presenza di voci da bot.

```
SELECT COUNT(DISTINCT ips_item_id)
FROM wb_items_per_site as w
WHERE w.ips_site_page NOT LIKE ':%%' AND
      w.ips_site_id in ('itwiki', 'enwiki', 'frwiki', 'dewiki',
                       'svwiki', 'nlwiki', 'ruwiki', 'eswiki', 'plwiki')
```

23.000.000 di voci distinte sommando tutte le Wikipedia

```
SELECT COUNT(DISTINCT ips_item_id)
FROM wb_items_per_site as w
WHERE w.ips_site_page NOT LIKE ':%:%'
#per simulare ns0 perdendo pochissime voci
```

```
SELECT COUNT(DISTINCT w1.ips_item_id)
FROM wb_items_per_site as w1, wb_items_per_site as w2
WHERE w1.ips_site_page NOT LIKE '%:%' #per simulare ns0 perdendo pochissime voci
AND w1.ips_site_page = w2.ips_site_page
AND w1.ips_site_id = 'itwiki'
AND w2.ips_site_id = 'enwiki'
```

```
SELECT COUNT(DISTINCT w1.ips_item_id)
FROM wb_items_per_site as w1, wb_items_per_site as w2,
wb_items_per_site as w3, wb_items_per_site as w4, wb_items_per_site as w5
WHERE w1.ips_site_page NOT LIKE '%:%' #per simulare ns0 perdendo pochissime voci
AND w1.ips_site_page = w2.ips_site_page AND w1.ips_site_page = w3.ips_site_page
AND w1.ips_site_page = w4.ips_site_page AND w1.ips_site_page = w5.ips_site_page
AND w1.ips_site_id = 'itwiki' AND w2.ips_site_id = 'enwiki' AND w3.ips_site_id = 'frwiki'
AND w4.ips_site_id = 'dewiki' AND w5.ips_site_id = 'eswiki'
```

CREDITS

Special thanks to all the people who made and released these awesome resources for free:

- ❖ Presentation template by [SlidesCarnival](#)
- ❖ Photographs by [Unsplash](#)
- ❖ Paper texture by [GraphicBurguer](#)