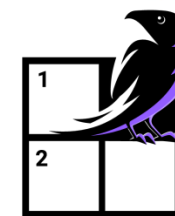




crossword challenge

Can you compete with Webcrow?



WEBCROW 2.0

Prof. Marco Gori
University of Siena

Prof. Marco Maggini
University of Siena

Marco Erlandes
expert.ai
Hybrid Lang Technology Director

Giovanni Angelini
expert.ai
Hybrid Lang Technology R&D

Andrea Zugarini
expert.ai
Hybrid Lang Technology R&D



AI VS HUMAN

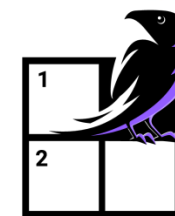
crossword challenge

Can you compete with Webcrow?



24/02/2023

TAILOR WORKSHOP



WEBCROW 2.0

Prof. Marco Gori
University of Siena

Prof. Marco Maggini
University of Siena

Marco Erlandes
expert.ai
Hybrid Lang Technology Director

Giovanni Angelini
expert.ai
Hybrid Lang Technology R&D

Andrea Zugarini
expert.ai
Hybrid Lang Technology R&D



10:00am
10:10pm

Welcome to the WebCrow workshop

10:10am
10:25am

Introducing WebCrow Project

10:25pm
11:00am

Crossword Competition

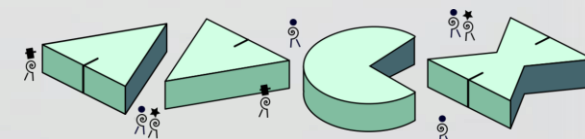
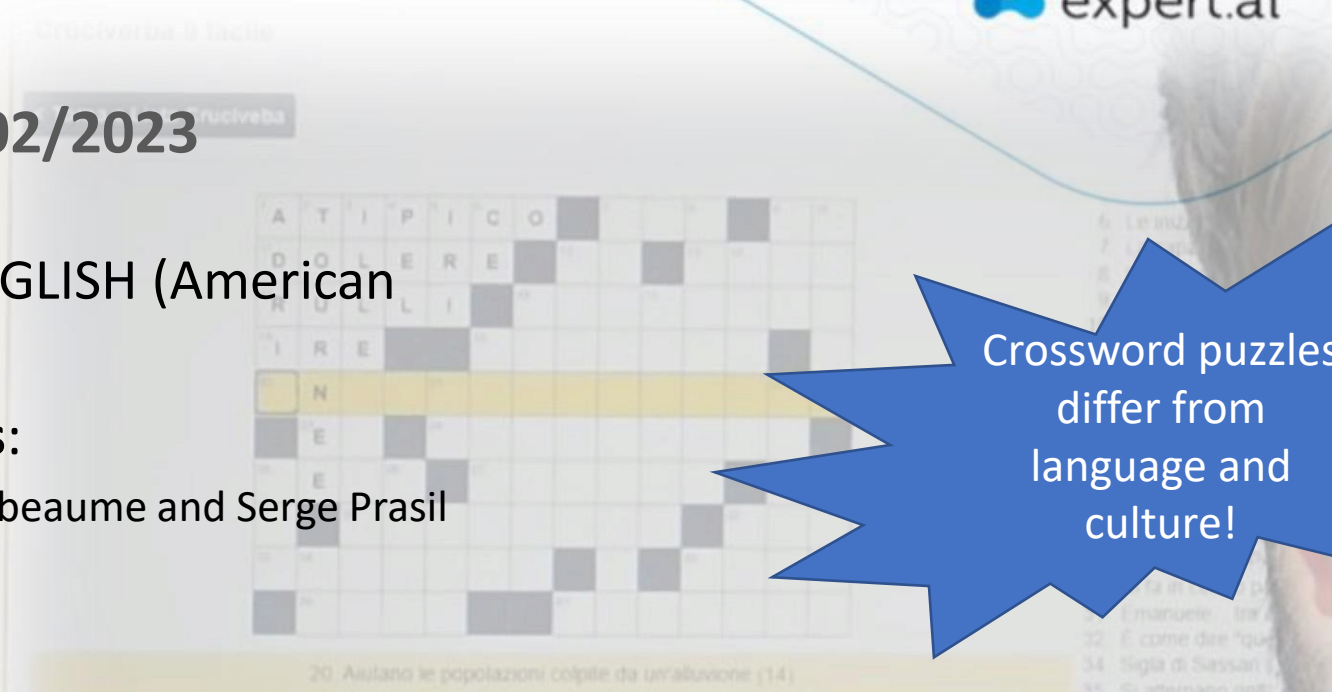
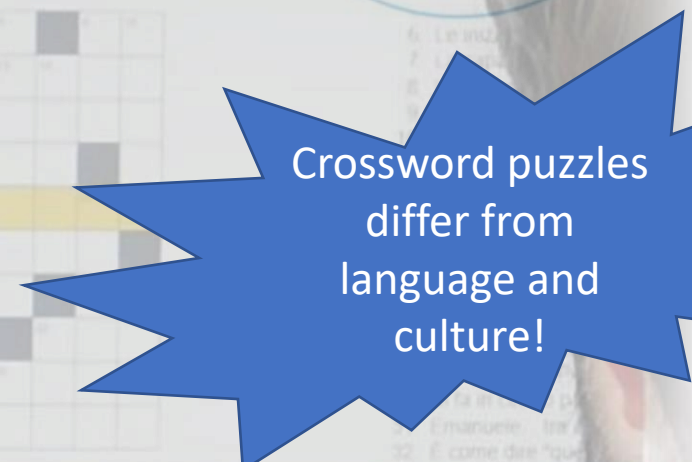
11:00am
12:00am

Technical discussion



The competition | 24/02/2023

- **Three different languages:** FRENCH, ENGLISH (American Crosswords), ITALIAN
- **Three crossword puzzle author/editors:**
 - French puzzles thanks two authors: Michel Labeaume and Serge Prasil
 - AVCX for American Crosswords
 - Giovanni Sordelli (Sudoku Edizioni)
- **Competitors:**
 - All competitors, included WebCrow, will join the challenge using the “Crossword Arena” platform
- **“Crossword Arena” gaming platform**
 - A gaming platform for humans and machines
 - Easily accessible from mobile devices (App and Browser) and PC (Browser)





The competition | 24/02/2023

WEBCROW 2.0

Workshop on "WebCrow Project"

**TAILOR
WEBCROW
WORKSHOP**

First Workshop about WebCrow Benchmark in TAILOR.
Organized by CINI and Expert.ai.
FEB 24th, 2023 – 10:00-12:00

UNIVERSITÀ expert.ai

Tailor Workshop French
24/02/2023 - 10:25
webcrow.diism.unisi.it



**CLIQUEZ ICI POUR LE DÉFI DES
MOT CROISÉS FRANÇAIS**

WEBCROW 2.0

Workshop on "WebCrow Project"

**TAILOR
WEBCROW
WORKSHOP**

First Workshop about WebCrow Benchmark in TAILOR.
Organized by CINI and Expert.ai.
FEB 24th, 2023 – 10:00-12:00

UNIVERSITÀ expert.ai

Tailor Workshop Italian
24/02/2023 - 10:25
webcrow.diism.unisi.it



**PREMERE QUI PER LA SFIDA
SUI CRUCIVERBA ITALIANI**

WEBCROW 2.0

Workshop on "WebCrow Project"

**TAILOR
WEBCROW
WORKSHOP**

First Workshop about WebCrow Benchmark in TAILOR.
Organized by CINI and Expert.ai.
FEB 24th, 2023 – 10:00-12:00

UNIVERSITÀ expert.ai

Tailor Workshop USA
24/02/2023 - 10:25
webcrow.diism.unisi.it



**CLICK HERE FOR AMERICAN
CROSSWORDS CHALLENGE**

<https://webcrow.diism.unisi.it/>



24/02/2023 TAILOR WORKSHOP

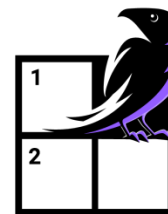
crossword challenge

Can you compete with Webcrow?



10:00 _{am} 10:10 _{pm}	Welcome to the WebCrow workshop
10:10 _{am} 10:25 _{am}	Introducing WebCrow Project
10:25 _{pm} 11:00 _{am}	Crossword Competition
11:00 _{am} 12:00 _{am}	Technical discussion

Good Luck!



WEBCROW 2.0

Prof. Marco Gori
University of Siena

Prof. Marco Maggini
University of Siena

Marco Ernandes
expert.ai
Hybrid Lang Technology Director

Giovanni Angelini
expert.ai
Hybrid Lang Technology R&D

Andrea Zugarini
expert.ai
Hybrid Lang Technology R&D



Can machines solve crossword puzzles?

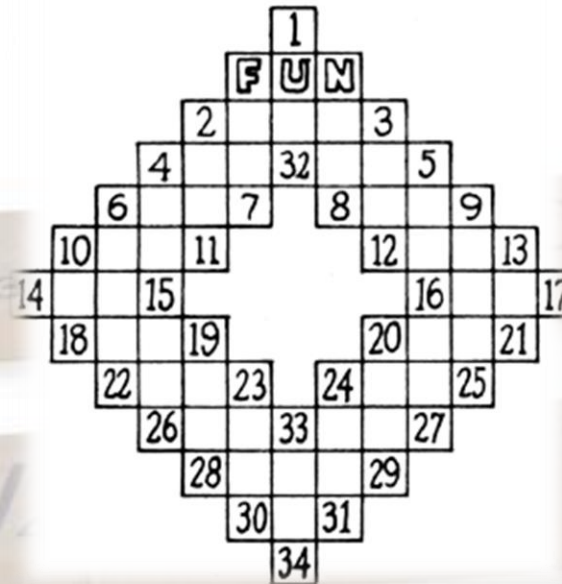
For over a century, crossword puzzles have been an intriguing challenge for humans...

...because of the complexity and nuances of the human language.

*I define language games as games involving natural language in which word meanings play an important role. Because natural language can be used to describe the full range of human experiences, language games are **inconsistent with the closed world assumption**—no fixed set of rules will be sufficient to define game play*

Review: *Computer Language Games* – Littman (CG 2000)

FUN'S Word-Cross Puzzle.



Arthur Wynne's
"first crossword
puzzle" 1913



Orizzontali: 1. Guai se l'onda mi varca o mi spezza
2. In Germania son acqua corrente
3. Ogni dì quando il sole è morente
4. Così soglion le preci finir.

Verticali: 1. Sono un fiore di rara bellezza
2. Il medesimo in lingua latina
3. Quali frutti noi siamo, indovina!
4. Per la messe di là da venir.

Giuseppe Airoldi's
"parole incrociate"
1890



Understanding, Knowledge, Reasoning



Utilizing several logical and language analysis steps focused on **understanding clues**.



Finding the correct meaning requires inferences and reasoning that makes use of **knowledge in various sources**.



Reasoning: a combination of different AI techniques (machine learning, constraint satisfaction, graph processing) is used to select the answer candidates, evaluate them and enter the answers in the puzzle grid.



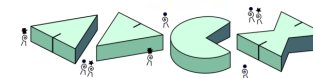
Crosswords' features

• Italian Crosswords



- 2-letter cells are allowed, “blind cells” are allowed
- inflections are very common (verb inflections are common)
- Blank-cell symmetry is not common (especially with word puzzles)
- Multi-word answers are common and follow some conventions
 - Full syntagms without the article (e.g. “PALAZZODELBO” not “ILPALAZZODELBO”) unless they are known named entities (ILFUMATTIAPASCAL)
- A number of clue conventions are used for short answers
- Large areas with no black cells are common

• American Crosswords

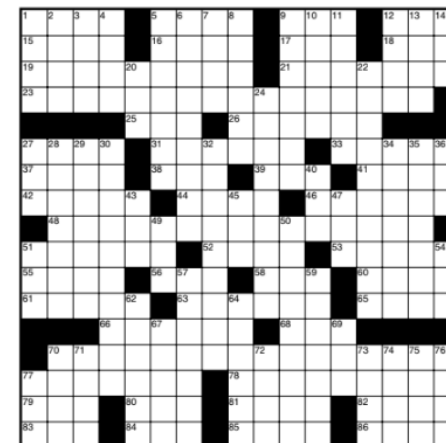
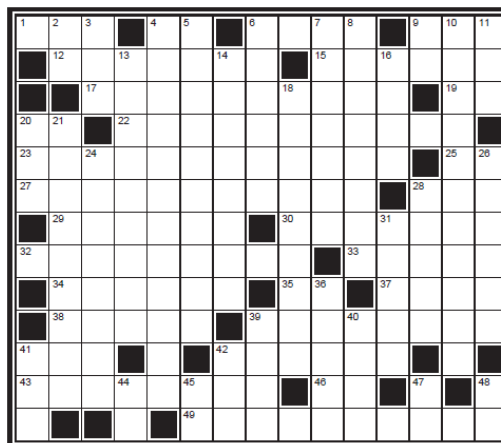


- 2-letter cells are NOT allowed nor “blind cells”
- inflections are present but rare
- Blank-cell symmetry is the standard
- Multi-word answers don't follow conventions (OHNOES, UNFEDORACLE, ITSUS, PLAYEDHARDTOGET)
- A number of clue conventions are used (fill-in-the-blanks, parenthesis, question marks, etc.)

Not cryptic crosswords!

WebCrow currently accepts:

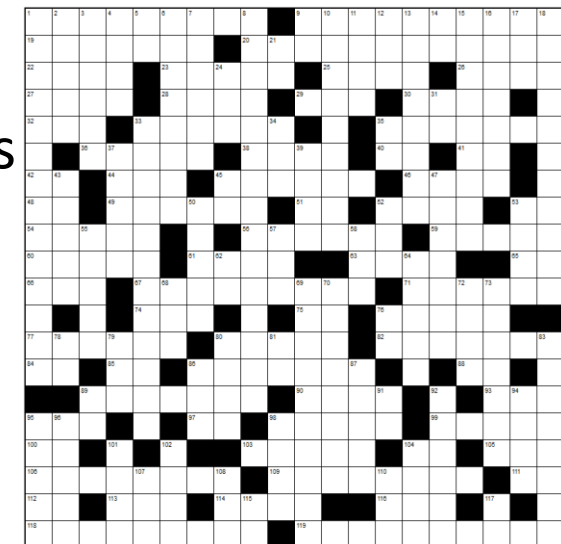
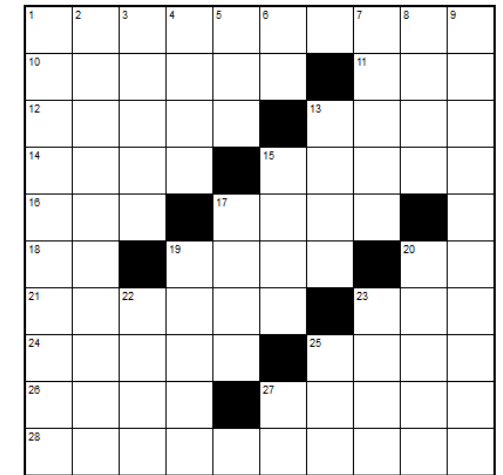
- Crosswords with 26 roman capital letters [A-Z]
- Only one single character per cell





Crosswords' features

- French Crosswords
 - 2-letter cells are allowed, “blind cells” are allowed.
 - Grid dimensions can differ. Sometimes there is symmetry.
 - Inflections are very common, including verb inflections.
 - Common use of specific domain answers
 - Multi-word answers are less common than American crosswords
 - A number of clue conventions are used for short answers
 - Large areas with no black cells are common



WebCrow currently accepts:

- Crosswords with 26 roman capital letters [A-Z]
- Only one single character per cell

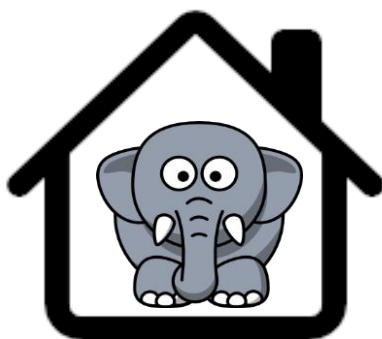


An elephant in the room

- «Automated Crossword Solving» (ACL 2022)
 - Dr Fill/BCS: a joint effort between Matthew Ginsberg and UC Berkeley
 - Strong focus on American themeless crosswords
 - Significant achievement! 99.7% correct letters → Kudos to BCS !!!
 - One unique knowledge base → 6M+ clue-answer pairs
- With WebCrow 2.0 we take into account additional elements:
 - Crosswords is a cross-cultural and cross-language game, widely spread outside the U.S, and each culture and language provide their own challenges!!
 - Human’s experience cannot scale to «hundreds of thousand» of crosswords or millions of clues and this quantity of digital data is not available for any language
 - The large American crossword database is an anomaly, we cannot rely on abundant digital data



- అక్షరం
1. నామ సంఖ్యలకు అంకం నుండి తర్వాతి అక్షరం
 2. ద్వైతం అనే ద్వైతం
 3. దాదాపుగానే తమ అధ్యయనం అంతా కేంద్రం
 4. మూడు ద్వైతం
 5. పాఠశాల
 6. భద్రం
 7. మన దానిని అన్ని అంకం సంఖ్యలకు
 8. అక్షరం అక్షరాలే మేము
- అక్షరం
1. ద్వైతం
 2. అ సంఖ్యలకు అంకం సంఖ్యలకు
 3. ద్వైతం
 4. అక్షరం అంకం
 5. అక్షరం అంకం
 6. అక్షరం అంకం





WebCrow's mission

- «WebCrow 2.0 has a specific mission:
 - It aims at challenging humans in any Western language using a Roman-alphabet
 - We believe that the focusing on one language cannot capture the richness of the game and the full nature of the challenge
- How can we achieve this goal?
 - Adopting a scalable approach:
 - leveraging less on pre-solved puzzles (ideally downsizing to human-compatible databases)
 - exploiting multi-lingual knowledge (expert.ai KGs in 12 different languages)
 - opening the architecture to allow any collaborator to add «experts» in a plug-in approach
 - Transition to a «grid-aware» clue-answering





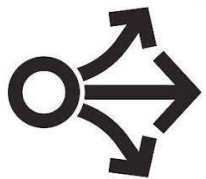
Crossword puzzles solving process

1. “Clue Answering”



- “Clue-answering experts” produces a list of candidate answers for each clue.
- There are NLP “clue-analysis experts”
- Candidate list are evaluated and enriched
- The goal:
 - not missing a correct answer
 - having the correct answer on top of the list with high probability

2. “Belief Propagation”



A process that reinforces the probability of candidate-answers which fit well in the puzzles based on the puzzle constraints.

3. “Grid filling”

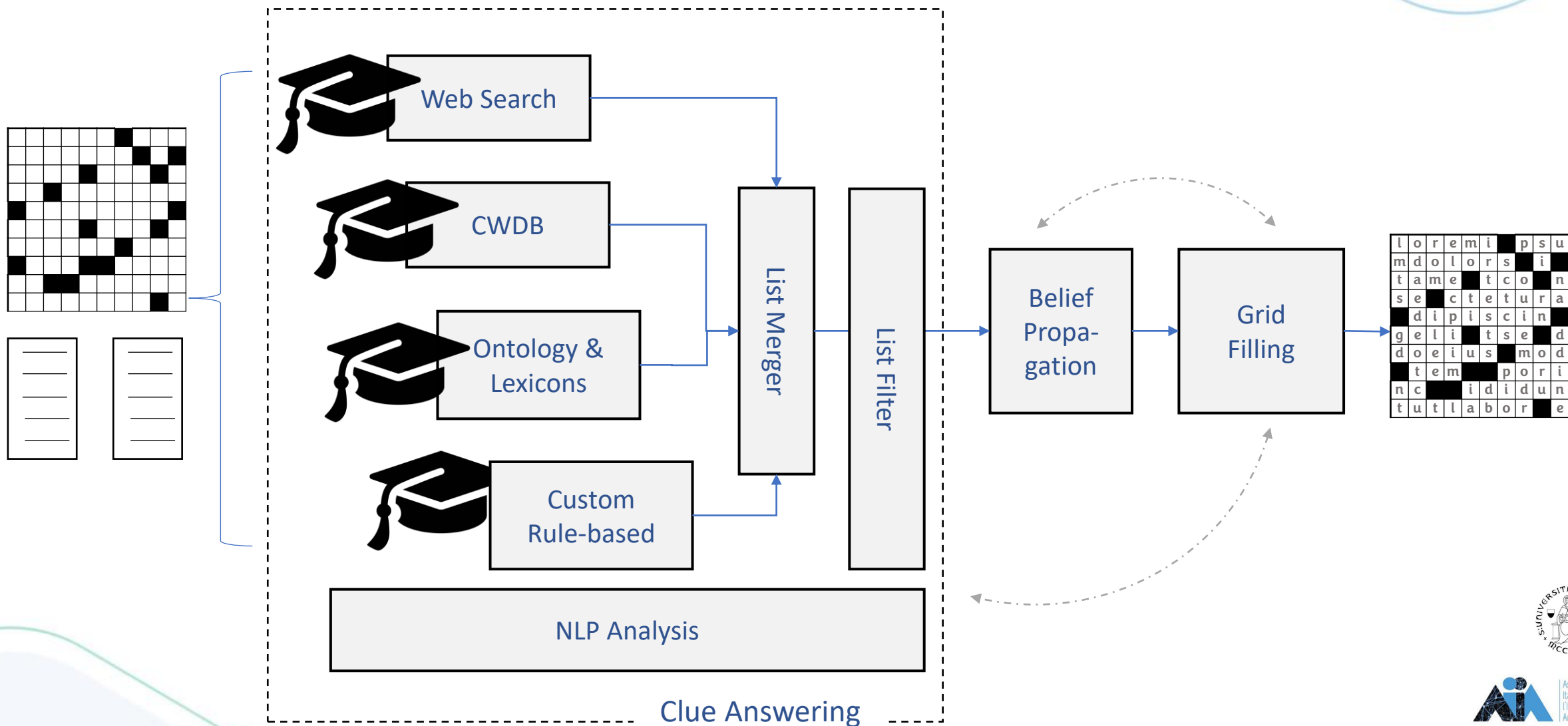
Filling up the puzzle grid using candidate-answers, or also candidate-letters.



Open architecture
WebCrow makes use of external expert systems!



Crossword puzzles solving process





Clue Answering

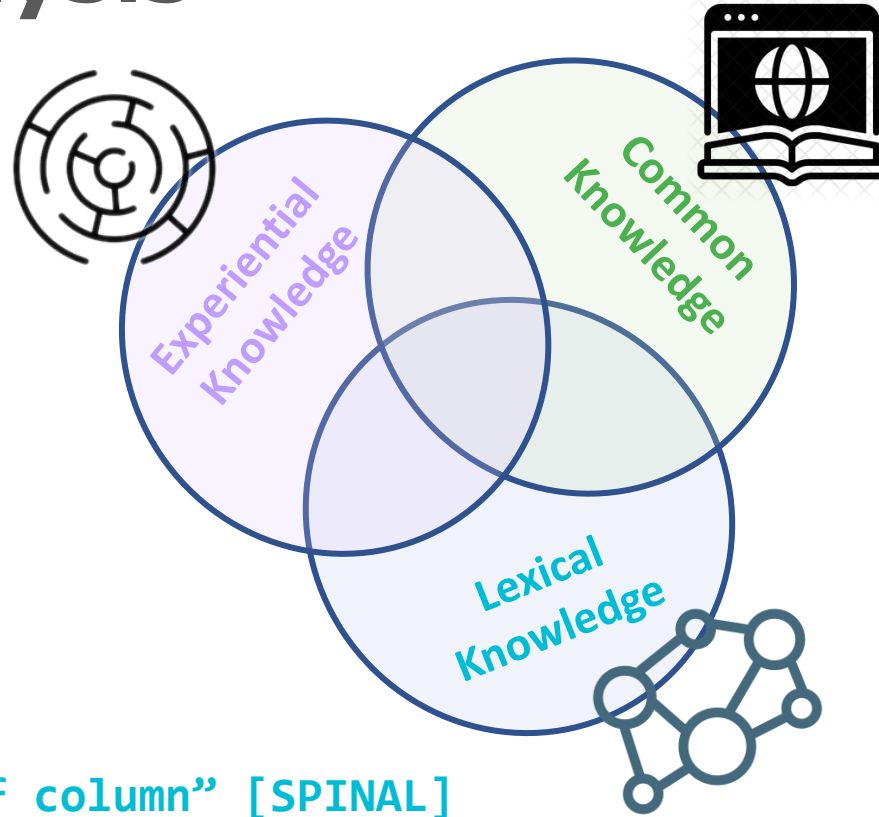
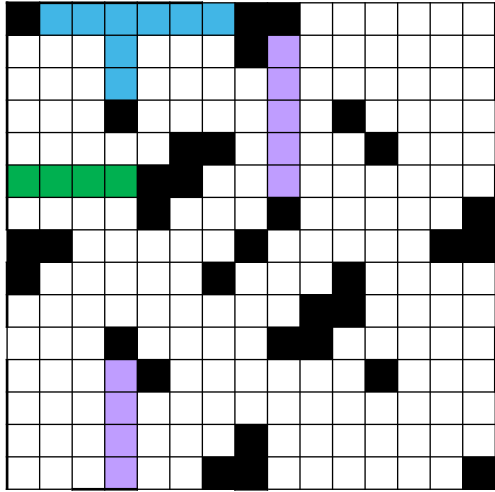
Clue Analysis

- Not all clues are identical: they require different skills and strategies.
- We organize them into 3 macroscopic classes, given 3 types of knowledge involved:
 - **Lexical and Ontological Knowledge:** knowledge about the way we use language to represent the world and organize information
 - **Common Knowledge:** encyclopedic knowledge, common sayings, facts and events of a common cultural background.
 - **Crossword-specific experiential Knowledge:** frequent crossword clues/answers, specific conventions
- But mixtures of these are also possible!
 - Fill-in-the-blank clues (e.g. "It's the end of ___" [ANERA])
 - Famous persons initials (e.g. "Iniziali di Everett" [RE])
 - Cross-Reference clues (e.g. "With 37-Across, co-winner of the 1981 Grammy for Album of the Year" [YOKO])



Clue Answering

Clue Analysis



20A “Inquieti e nervosi” [STRANITI]

27A “Un po' ebbro” [EB]

36D “Vetrinetta da museo” [TECA]

24A “Gregorio, campione di nuoto” [PALTRINIERI]

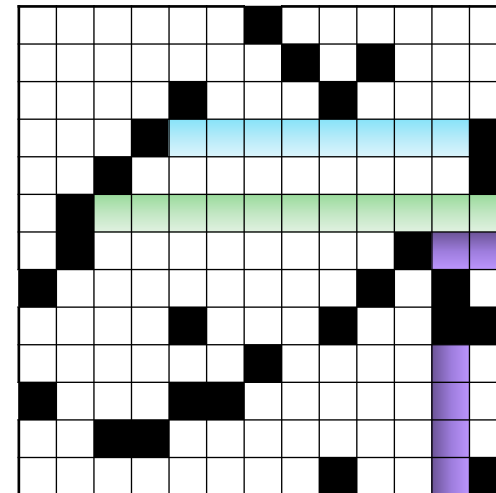
3D “Sick” [ILL]

1A “Supportive kind of column” [SPINAL]

14D “Cylindrical storage vessels” [SILOS]

46D “Drive-___” [THRU]

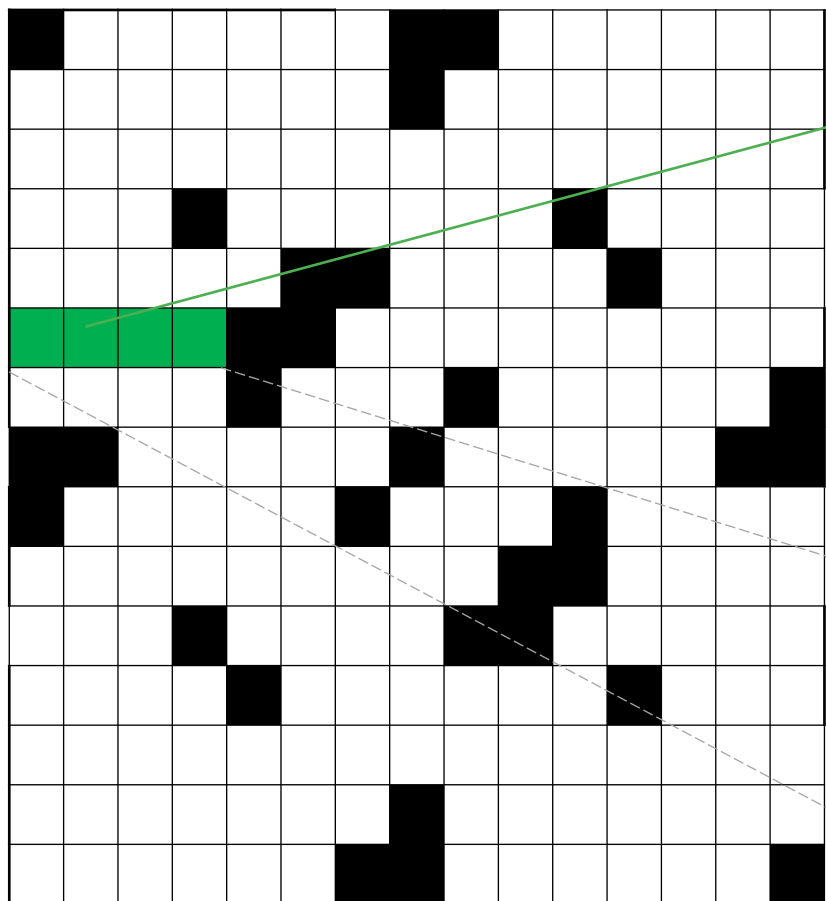
25A “Emilia Jones Best Picture winner” [CODA]





Clue Answering

Common Knowledge



25A “Emilia Jones Best Picture winner”
[CODA]

[Oscars 2022: How 'CODA' won best picture - Los Angeles Times](https://www.latimes.com › awards › story › oscars-2022...)

<https://www.latimes.com › awards › story › oscars-2022...>

Mar 27, 2022 — Emilia Jones and Troy Kotsur in a scene from *best picture* winner, “CODA.” (Apple TV+). By Glenn WhippEntertainment Columnist. March 27, 2022 ...

RANK	CANDIDATE-ANSWER	SCORE
1	CODA	22.6%
2	TROY	6.1%
3	FILM	4.1%
...

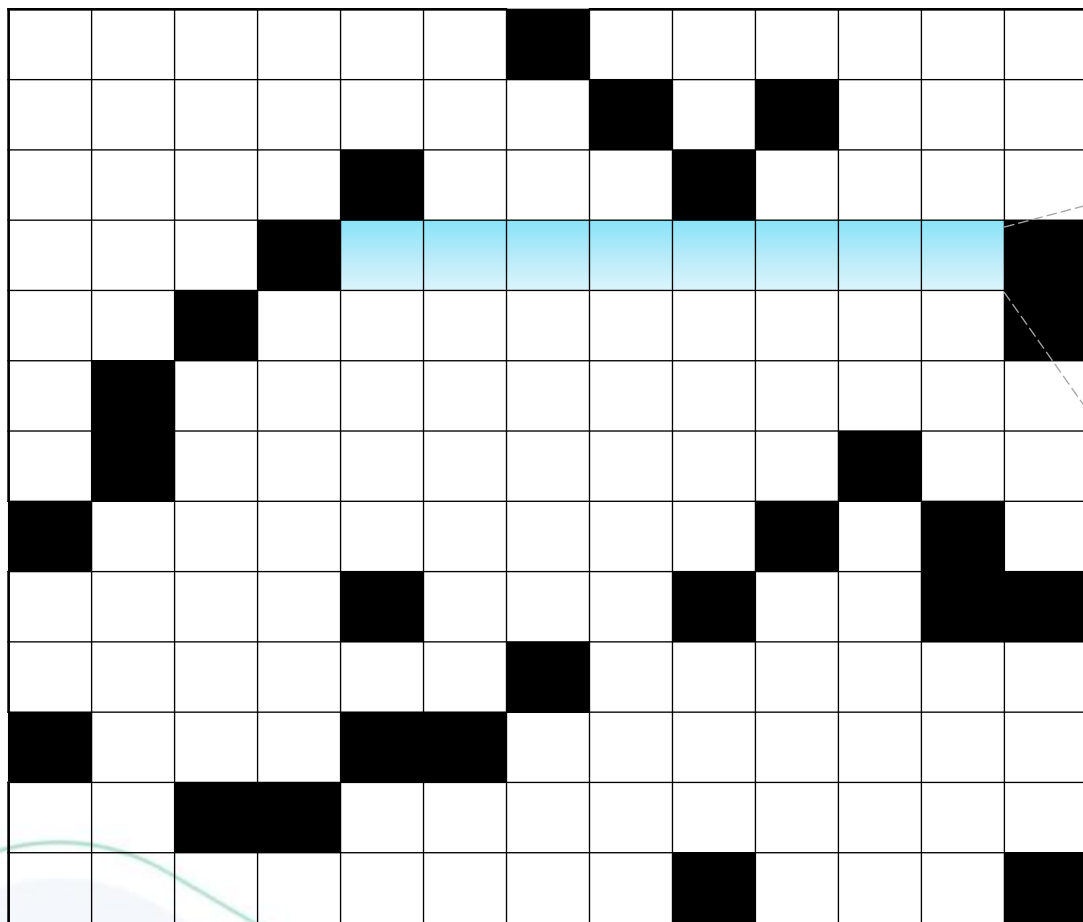
Copyright: AVCX <https://avxwords.com/>





Clue Answering

Lexical and Ontological Knowledge



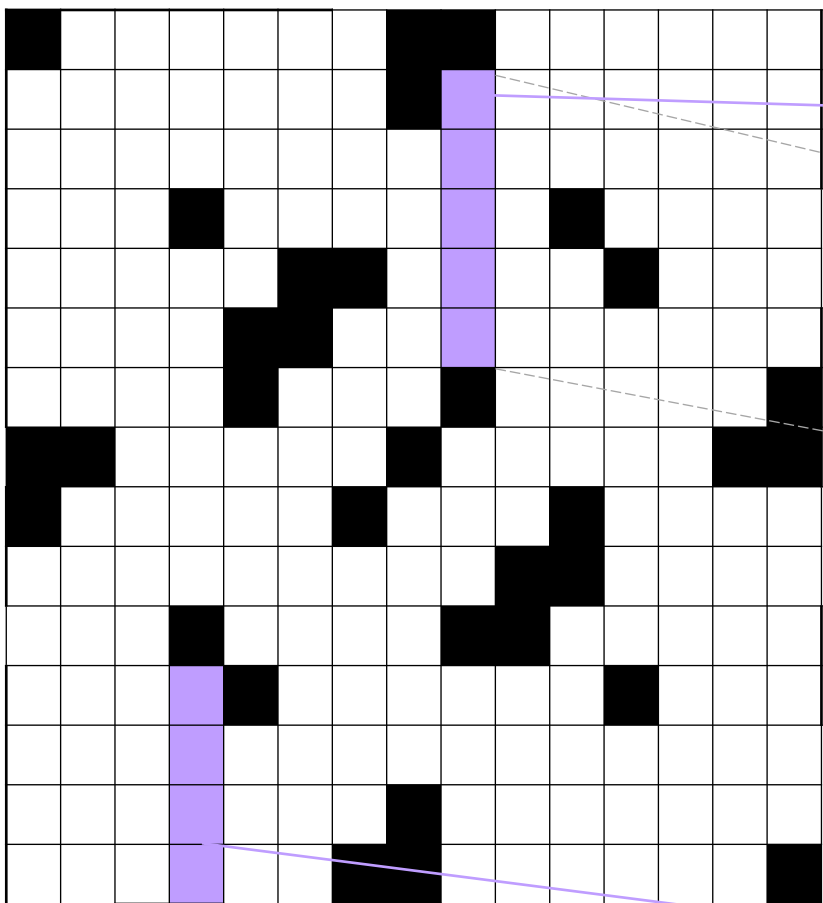
20A “Inquieti e nervosi”
[STRANITI]

RANK	CANDIDATA	SCORE
1	IRRITATI	20,7%
2	STRANITI	14,3%
3	ALTERATI	9,16%
4	SDEGNATI	9,06%
5	ISTERICI	5,22%
...
...
...
24230	GENETICI	0,0000000006%
...
...



Clue Answering

Knowledge from puzzles



14D "Cylindrical storage vessels"
[SILOS]

RANK	CANDIDATE-ANSWER	SCORE
1	SILOS	90.1%
2	TANKS	2.8%
...

Other clues on "SILOS":

- «Towers for storing grain » -> SILOS
- «Where to store grains » -> SILOS
- «Cylindrical building for grain storage» -> SILOS

Copyright: AVCX <https://avxwords.com/>

[46 D.] Drive-___

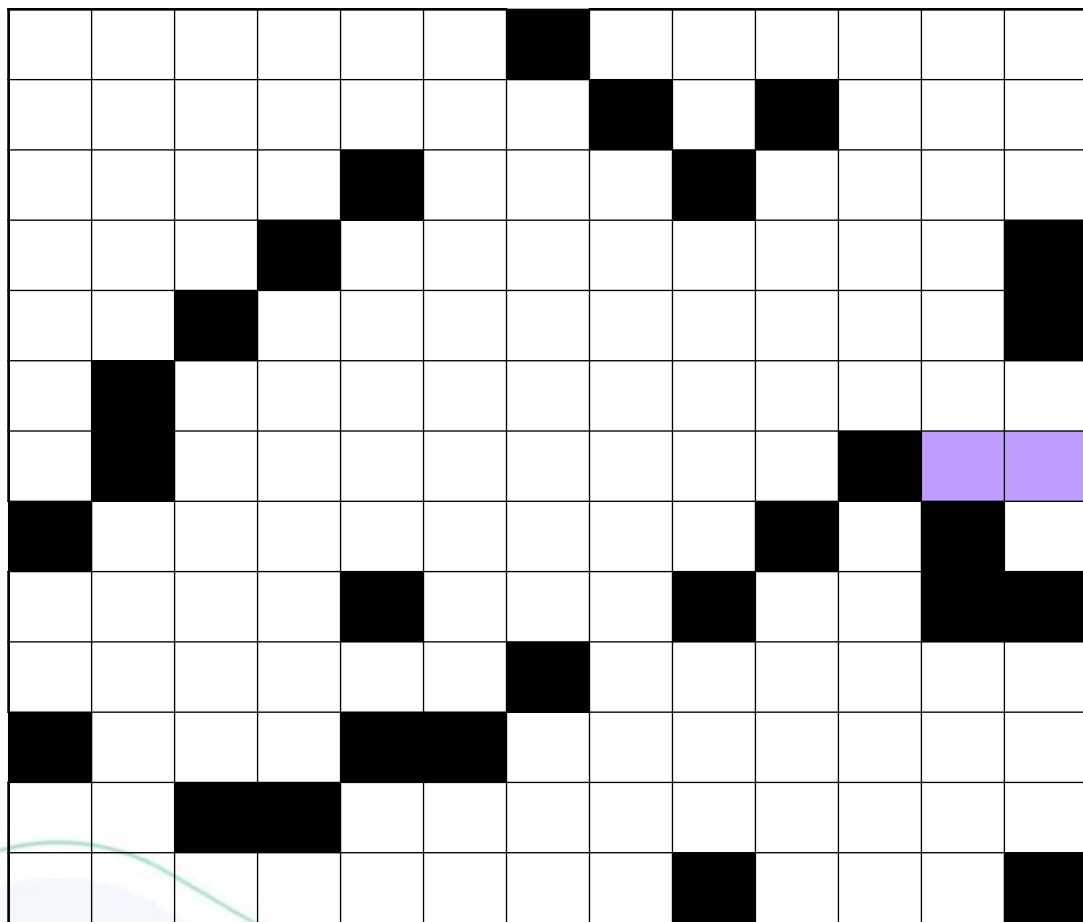


Clue Answering

Knowledge from puzzles

27A “Un po' **ebbro**” [EB]

Ruled based module
for Italian crosswords

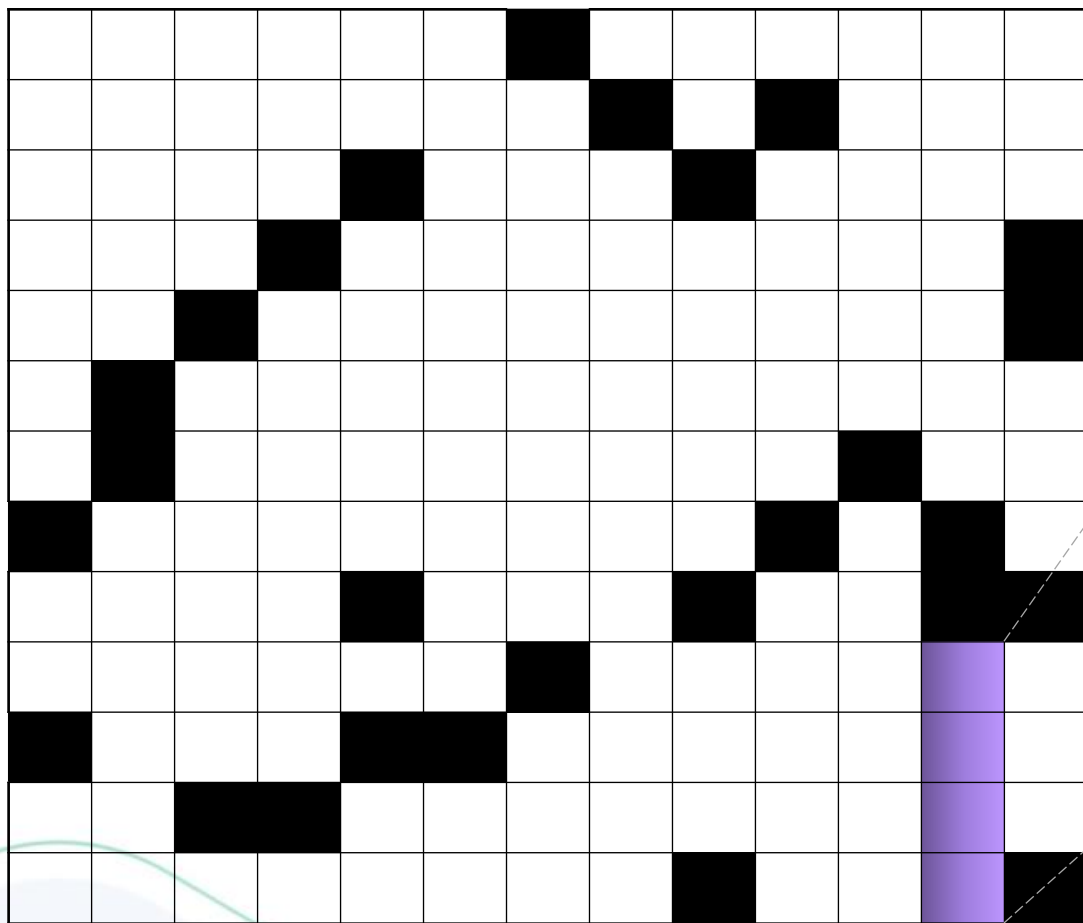


RANK	CANDIDATE-ANSWER	SCORE
1	EB	100%



Clue Answering

Knowledge from puzzles



36V “Vetrinetta da museo” [TECA]

RANK	CANDIDATA	SCORE
1	TECA	89,99%
2	RARA	9,4%
3	ARTE	3,4%
...
2531	VIVA	0,0000000001%
...

- «Una custodia per reliquie» -> TECA
- «Un astuccio da museo» -> TECA
- «Vetrina per reliquie» -> TECA



Clue Answering

Modules overview

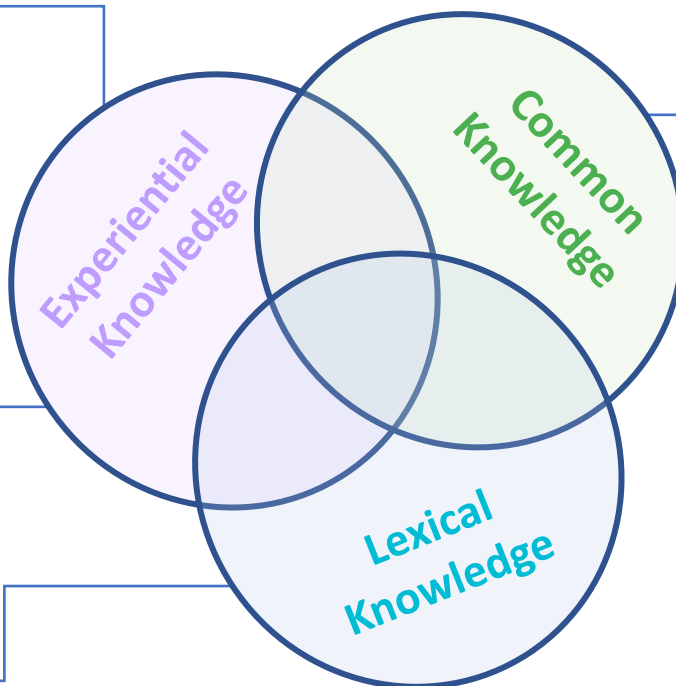
Crossword Database

- American
 - 3.1M <clue,answer> unique pairs
 - 314K unique answers (438K Berkeley CS)
 - approx 50K solved puzzles*
- Italian:
 - 125K unique pairs
 - 41K unique answers
 - approx 2K solved puzzles

Italian Rule-based

expert.ai Knowledge Graph

- | | |
|--|--|
| <ul style="list-style-type: none"> English: <ul style="list-style-type: none"> 400K+ base concepts 4M+ extended wikidata concepts 1M+ relations | <ul style="list-style-type: none"> Italian: <ul style="list-style-type: none"> 300K+ base concepts 3M+ extended wikidata concepts 1M+ relations |
|--|--|



Web Search

- “Full document scan”
- “Snippet scan”

Currently using BING:

- param “country code”:
 - IT / US

No custom logic has been implemented yet!

e.g. No “fill-in-the-blank” strategy.

A specific filter to avoid crossword-related pages has been added for testing on published puzzles

*50K solved puzzles = 3 a day for ~ 50 year





Clue Answering

Clue Answering Technique

- Current solution includes a partial implementation of Neural QA techniques presented in:
 - “A Multi-Strategy Approach to Crossword Clue Answer Retrieval and Ranking” [Zugarini, Ernandes, 2021]
- When using the Crossword Database (CWDB) the Neural QA approach to Clue-answering can follow three approaches:
 - Question-to-Question (QQ): → currently WebCrow only exploits “QQ”
 - Neural encode the CWDB clues
 - neural-encode the un-answered clue
 - rank the most similar clues in the CWDB by encoding similarity
 - use the answers of most similar clues as answer candidates
 - Question-to-Answer (QA):
 - Neural encode the CWDB answers
 - neural-encode the un-answered clue
 - rank the most similar answers in the CWDB by encoding similarity
 - Combine the two approaches
- With small adaptations the same approach can be used also to rank Web-search answers



Clue Answering

List Filtering Technique

1. Clues are classified based on their “Morphological Answer Type” \rightarrow vector **MAT(c)**
 - Currently WebCrow is predicting classes that can have an impact on inflections (which can have a dramatic impact on Italian crosswords):
 - Gender (M,F)
 - Number (S,P)
 - Part-of-speech (Adj, Noun, Verb)
2. We determine the possible morphological classes of each candidate answer, with a confidence score \rightarrow vector Answer Morpho Classes: **AMC(a)**
3. Scores of answers are “clustered” by base-form:
 - different inflections of the same base-form should maintain close probabilities
4. The answer candidates probabilities are then recomputed as:

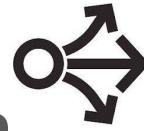
$$p'(a) = p(a) * \text{norm} * \langle \text{MAT}(c), \text{AMC}(a) \rangle$$

- Current implementation is only partial but it allows to reduce single-cell errors by >30% on Italian crosswords



Belief Propagation

Belief Propagation



Input:

- Candidate-answer lists ranked by probability

Constraints:

- Each letter belongs to two different candidate-answer lists

Hypothesis:

- The correct answer should maximize the overall probability, based on each selected candidate-answer probability and the constraints.

Goals:

- Increase the probability (score) of correct answers
- Make the overall process robust to possible missing answers

7A “Corrente artistica francese... pucciniana»
[BOHEME]

C	A	R	N	A	P		B	O	H	E	M	E
R	O	A	D	M	A	P		K		D	I	N
E	S	T	E		R	E	C		A	O	N	I
S	T	O		S	T	R	A	N	I	T	I	
I	A		B	A	I	O	N	E	T	T	A	
M		P	A	L	T	R	I	N	I	E	R	I
A		A	R	I	A	A	R	I	A		E	B
	D	I	E	S	I	R	A	E		R		M
K	R	O	L		V	E	N		A	E		
F	O	L	L	I	A		D	O	S	A	T	A
	G	O	A			M	A	N	A	G	E	R
E	H			G	R	E	G	O	R	A	C	I
M	E	D	I	O	C	R	I		I	N	A	

RANK	CANDIDATA	SCORE
1	ORFICA	15,9%
2	INGRES	11,1%
3	FELICE	7,47%
4	MADAMA	5,75%
...
23	BOHEME	1,62%
24	LIRICA	1,44%
...

RANK	CANDIDATA	SCORE
1	IRRITATI	20,7%
2	STRANITI	14,3%
3	ALTERATI	9,16%
4	SDEGNATI	9,06%
5	ISTERICI	5,22%
...

RANK	CANDIDATA	SCORE
1	CAROVANIERI	4,7%
2	SENZAUSCITA	3,9%
3	ABBANDONATI	2,7%
...

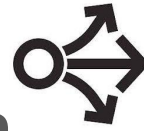
RANK	CANDIDATA	SCORE
1	EDOTTE	15,9%
2	EDOTTI	11,1%
3	TENUTE	7,47%
4	VITALE	5,75%
...	CAPIRE	...
23	CAPITI	1,62%
...

17D “Lo sono i quattrozampe a zozzo per le strade»
[CANIRANDAGI]



Belief Propagation

Belief Propagation



7A "Corrente artistica francese... pucciniana» [BOHEME]

C	A	R	N	A	P		B	O	H	E	M	E
R	O	A	D	M	A	P		K		D	I	N
E	S	T	E		R	E	C		A	O	N	I
S	T	O		S	T	R	A	N	I	T	I	
I	A		B	A	I	O	N	E	T	T	A	
M		P	A	L	T	R	I	N	I	E	R	I
A		A	R	I	A	A	R	I	A		E	B
	D	I	E	S	I	R	A	E		R		M
K	R	O	L		V	E	N		A	E		
F	O	L	L	I	A		D	O	S	A	T	A
	G	O	A			M	A	N	A	G	E	R
E	H			G	R	E	G	O	R	A	C	I
M	E	D	I	O	C	R	I		I	N	A	

Input:

- Candidate-answer lists ranked by probability

Constraints:

- Each letter belongs to two different candidate-answer lists

Hypothesis:

- The correct answer should maximize the overall probability, based on each selected candidate-answer probability and the constraints.

Goals:

- Increase the probability (score) of correct answers
- Make the overall process robust to possible missing answers

RANK	CANDIDATA	SCORE
1	BOHEME	99,1%
2	MONEME	0,05%
3	TOTEME	0,05%
4	DOREME	0,04%
...
23	RICEVE	0,01%
24	POTERE	0,01%
...

RANK	CANDIDATA	SCORE
1	STRANITI	99,7%
2	STRANITE	0,14%
3	STRANITA	0,05%
4	STRANITO	0,04%
5	STRONATI	0,00..%
...

RANK	CANDIDATA	SCORE
1	MAGIRANDAGI	96,2%
2	CANICATTINI	0,08%
3	CANICATTESE	0,0..%
...

RANK	CANDIDATA	SCORE
1	EDOTTE	99,8%
2	EDOTTI	0.008%
3	EDOTTO	0.005%
4	EDOTTA	0.004%
...
23	ESATTE	0.00..%
...

17D "Lo sono i quattrozampe a zozzo per le strade» [CANIRANDAGI]





Belief Propagation

Propagate the probability of each candidate-list

Down answers

TOP 1
TOP 5
TOP 10
TOP 50
TOP 100
TOP 500
TOP 1000
TOP 10000
PRESENT
ABSENT

Based on the puzzle constraints is possible to propagate the probability of each candidate-list to the other candidate-lists.

C	A	R	N	A	P		B	O	H	E	M	E
R	O	A	D	M	A	P		K		D	I	N
E	S	T	E		R	E	C		A	O	N	I
S	T	O		S	T	R	A	N	I	T	I	
I	A		B	A	I	O	N	E	T	T	A	
M		P	A	L	T	R	I	N	I	E	R	I
A		A	R	I	A	A	R	I	A		E	B
	D	I	E	S	I	R	A	E		R		M
K	R	O	L		V	E	N		A	E		
F	O	L	L	I	A		D	O	S	A	T	A
	G	O	A			M	A	N	A	G	E	R
E	H			G	R	E	G	O	R	A	C	I
M	E	D	I	O	C	R	I		I	N	A	

Across answers



Belief Propagation

Propagate the probability of each candidate-list

Down answers

Across answers

TOP 1
TOP 5
TOP 10
TOP 50
TOP 100
TOP 500
TOP 1000
TOP 10000
PRESENT
ABSENT

C	A	R	N	A	P		B	O	H	C	A	R	N	A	P		B	O	H	E	M	E	N	A	P		B	O	H	E	M	E		
R	O	A	D	M	A	P		K		R	O	A	D	M	A	P		K		D	I	N	D	M	A	p		K		D	I	N		
E	S	T	E		R	E	C		A	E	S	T	E		R	E	C		A	O	N	I	E		R	E	C		A	O	N	I		
S	T	O		S	T	R	A	N	I	S	T	O		S	T	R	A	N	I	T	I			S	T	R	A	N	I	T	I			
I	A		B	A	I	O	N	E	T	I	A		B	A	I	O	N	E	T	T	A			B	A	I	O	N	E	T	T	A		
M		P	A	L	T	R	I	N	I	M		P	A	L	T	R	I	N	I	E	R	I		A	L	T	R	I	N	I	E	R	I	
A		A	R	I	A	A	R	I	A	A		A	R	I	A	A	R	I	A		E	B		R	I	A	A	R	I	A		E	B	
	D	I	E	S	I	R	A	E			D	I	E	S	I	R	A	E		R		M		E	S	I	R	A	E		R		M	
K	R	O	L		V	E	N		A	K	R	O	L		V	E	N		A	E				L		V	E	N		A	E			
F	O	L	L	I	A		D	O	S	F	O	L	L	I	A		D	O	S	A	T	A		L	I	A		D	O	S	A	T	A	
	G	O	A			M	A	N	A		G	O	A			M	A	N	A	G	E	R		A			M	A	N	A	G	E	R	
E	H			G	R	E	G	O	R	E	H			G	R	E	G	O	R	A	C	I			G	R	E	G	O	R	A	C	I	
M	E	D	I	O	C	R	I			M	E	D	I	O	C	R	I								I	O	C	R	I					



Belief Propagation

Propagate the probability of each candidate-list

Down answers

Across answers

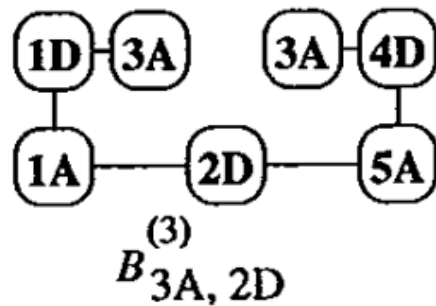
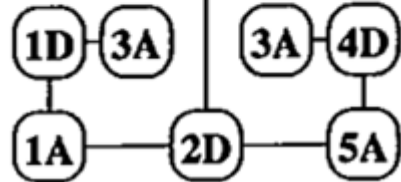
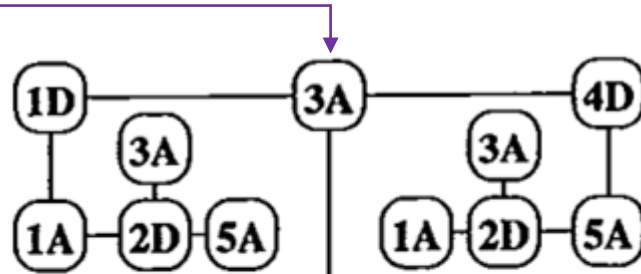
TOP 1
TOP 5
TOP 10
TOP 50
TOP 100
TOP 500
TOP 1000
TOP 10000
PRESENT
ABSENT

C	A	R	N	A	P		B	O	H	E	M	E
R	O	A	D	M	A	P		K		D	I	N
E	S	T	E		R	E	C		A	O	N	I
S	T	O		S	T	R	A	N	I	T	I	
I	A		B	A	I	O	R	E	T	T	A	
M		P	A	L	T	R	I	N	I	E	R	I
A		A	R	I	A	A	R	I	A		E	B
	D	I	E	S	I	R	A	E		R		M
K	R	O	L		V	E	N		A	E		
F	O	L	L	I	A		D	O	S	A	T	A
	G	O	A			M	A	N	A	G	E	R
E	H			G	R	E	G	O	R	A	C	I
M	E	D	I	O	C	R	I		I	N	A	


Belief Propagation

Computing posterior probabilities

1	I	2	N	
3	F	U	4	N
		5	T	O



$U_x^{(d)} \rightarrow$ constraint network with depth d based on slot x

$B_{x,y_i}^{(d)} \rightarrow y_i$ -th branch of $U_x^{(d+1)}$

$b_{x,y_i}^{(d)}(w) \rightarrow$ posterior probability that y_i takes value w in the network B

$q_x^{(d)}(v) \rightarrow$ posterior probability that x takes value v in the network $U^{(d)}$.

$$q_x^{(d)}(v) = k_x^{(d)} p_x(v) \cdot \prod_{i=1}^m \sum_{w | \text{match}_{y_i, x}(w, v)} b_{x, y_i}^{(d-1)}(w)$$

$$b_{y_i, x}^{(d)}(v) = k_{y_i, x}^{(d)} p_x(v) \cdot \prod_{j=1..m, j \neq i} \sum_{w | \text{match}_{y_j, x}(w, v)} b_{x, y_j}^{(d-1)}(w)$$



Grid Filling

Inserting letters by letter probability

C	A	P	N	A	P		B	O	H	E	M	E	
R	O	A	D	M	A	P		K		D	I	N	
E	S	T	E		R	E	C		A	O	N	I	
S	T	O		S	T	R	A	N	I	T	I		
I	A		B	A	I	O	N	E	T	T	A		
M		P	A	L	T	R	I	N	I	E	R	I	
A		A	R	I	A	A	R	I	A		E	B	
	D	I	E	S	I	R	A	E		R		M	
K	R	O	L		V	E	N		A	E			
F	O	L	L	I	A		D	O	S	A	T	A	
	G	O	A				M	A	N	A	G	E	R
E	H			G	R	E	G	O	R	A	C	I	
M	E	D	I	O	C	R	I			I	N	A	

C	A	P	N	A	P		B	O	H	E	M	E	
R	O	A	D	M	A	P		K		D	I	N	
E	S	T	E		R	E	C		A	O	N	I	
S	T	O		S	T	R	A	N	I	T	I		
I	A		B	A	I	O	N	E	T	T	A		
M		P	A	L	T	R	I	N	I	E	R	I	
A		A	R	I	A	A	R	I	A		E	B	
	D	I	E	S	I	R	A	E		R		M	
K	R	O	L		V	E	N		A	E			
F	O	L	L	I	A		D	O	S	A	T	A	
	G	O	A				M	A	N	A	G	E	R
E	H			G	R	E	G	O	R	A	C	I	
M	E	D	I	O	C	R	I			I	N	A	

C	A	R	N	A	P		B	O	H	E	M	E	
R	O	A	D	M	A	P		K		D	I	N	
E	S	T	E		R	E	C		A	O	N	I	
S	T	O		S	T	R	A	N	I	T	I		
I	A		B	A	I	O	N	E	T	T	A		
M		P	A	L	T	R	I	N	I	E	R	I	
A		A	R	I	A	A	R	I	A		E	B	
	D	I	E	S	I	R	A	E		R		M	
K	R	O	L		V	E	N		A	E			
F	O	L	L	I	A		D	O	S	A	T	A	
	G	O	A				M	A	N	A	G	E	R
E	H			G	R	E	G	O	R	A	C	I	
M	E	D	I	O	C	R	I			I	N	A	

Letter probability mass:

- For each slot s we cumulate the probability mass $p_d^s(c)$ of a letter c , in a given direction d , adding all the posterior probabilities of words that contain letter c in the slot s with direction d .
- We compute the probability mass $p^s(c)$ as

$$p^s(c) = k^s p_{ACROSS}^s(c) \cdot p_{DOWN}^s(c)$$
- This can be seen as the probability of the letter c to be correctly inserted in a given cell, considering the constraint network and the answers.
- We can use this to “take the risk” of inserting a letter in order to constraint the following grid filling.
- In our tests we are using 2 criteria:
 - $p^s(c) > 99.99\%$ &
 $best_{ACROSS}^s(c) == best_{DOWN}^s(c)$
 - $p^s(c) > 99\%$ &
 $best_{ACROSS}^s(c) == best_{DOWN}^s(c)$ &
 $p_{ACROSS}^s(c) > 90\%$ &
 $p_{DOWN}^s(c) > 90\%$





Grid Filling

Inserting letters by letter probability

- A “real example” of a partially letter-constrained grid

```

O L D V I C # C R I # # # C # # B R I G A T A
T E O C R A Z I A # I # C A F F E I N A # R #
T O N # A N A # P A N # A N E M O N E # S I M
U N N I # C F P # L I M N O S # L A T T U G A
A B I T O # F L O T T A A E R E A # T R O L L
G A N O # B I A R R I T Z # # # # A O # # I #
E T A # C A R T A I N T E S T A T A # T A E G
N T # S O C I A L M E D I A M A N A G E R # I
A I U O L A # N E E R A # L # R T # R O A N O
R S # F I T T I # N E M B I # # # B I R R A I
I T T I T [A] # # W T # O U T L E T # L I A N E
O A # A E # # G H I A N D O L A # A L A T O #

```

Letter insertion

Advantages of the approach

- Experimental observation have shown that >99.9% of letter insertions are correct, this leads to a correct “constraining” of the grid that can be used for:
 - Grid-filling completion (that currently produces >99% of errors)
 - Detecting which answers should be “re-examined” together with the clue.

65A “Popolazione mesopotamica”
[ITTITI]


Grid Filling
Inserting letters by letter probability

- A “real example” of a partially letter-constrained grid

Letter insertion

I	R	S	#	Y	E	L	L	O	W	#	#	B	O	P
N	A	H	#	A	L	I	A	S	E	S	#	A	N	O
A	N	I	#	S	I	M	I	L	A	C	#	S	E	W
B	O	A	#	I	S	A	D	O	R	A	#	S	T	E
I	N	T	E	R	#	#	#	#	#	R	A	D	A	R
N	T	S	B	#	[B]	[E]	[T]	[H]	[E]	F	I	R	S	T
D	O	U	B	L	[E]	[G]	[H]	[R]	[N]	#	R	[O]	T	I
#	#	#	S	E	[R]	[V]	[E]	[S]	[T]	[O]	#	[S]	E	E
A	P	T	#	[O]	[E]	[R]	[C]	[N]	[R]	A	S	#	#	#
S	H	E	A	#	T	[E]	[A]	[S]	[E]	[R]	V	I	C	E
S	I	X	T	H	[S]	E	N	S	E	#	E	C	O	N
E	L	A	T	E	#	#	#	#	#	K	N	E	E	S
M	L	S	#	A	T	S	T	A	K	E	#	B	T	U
B	I	B	#	T	H	E	I	D	E	A	#	U	Z	I
L	E	B	#	S	A	G	E	H	E	N	#	R	E	N
E	S	Q	#	#	[W]	[O]	[R]	[D]	[L]	[E]	#	N	[E]	G

Advantages of the approach

- Experimental observations have shown that >99.9% of letter insertions are correct, this leads to a correct “constraining” of the grid that can be used for:
 - Grid-filling completion (that currently produces >99% of errors)
 - Detecting which answers should be “re-examined” together with the clue.

28D “The second square in this column should be 4-Across; the last should be green” ” [CEGVTS]



Grid Filling

Filling the puzzle grid

Searching the best combination of clue-answers

- A number of different **options** to choose from
 - Dynamic Backtracking: early Dr Fill
 - Limited Discrepancy Search: Dr Fill
 - A* Search: early WebCrow (maximizing the sum of the log probabilities)
 - Greedy Search: Berkeley BCS (but with an addition game-changer iterative “local search” search)
 - MinConflicts: idea for the future? → maybe combined with Web-Search
- **Current implementation:** Greedy Search, optimizing a local function that takes into account the “prior probability”, the prior probability of crossing words and the letter probability mass.
- Future steps:
 - As for Berkeley BCS, after Greedy Search an iterative “local search” must be carried out
 - We should consider a “**Web-assisted local search**” since at this step the search engine could querying using both the clue and the candidate answer



Experimental Results

ITALIAN (66 crosswords)

Experiment	Puzzles	Correct Words	Correct Letters	Perfect Solutions
La Settimana Enigmistica	10	98.3%	99.5%	40%
Sudoku Edizioni	56	97.4%	99.2%	35.7%
Overall	66	97.55%	99.3%	36.4%

FRENCH (60 crosswords)

Experiment	Puzzles	Correct Words	Correct Letters	Perfect Solutions
Michel Labeaume	10	93.0%	98.0%	
Serge Prasil	10	91.8%	96.1%	
Other	40	65.0%	73.3%	
Overall	60	74.0%	81.2%	

AMERICAN (82 crosswords)

Experiment	Puzzles	Correct Words	Correct Letters	Perfect Solutions
NYT Monday	10	97.9%	99.3%	70%
NYT Tuesday	10	98.7%	99.7%	60%
NYT Wednesday				
NYT Thursday				
NYT Friday				
NYT Saturday				
NYT Sunday				
AVCX	12	88.2%	94.6%	10%
Overall	82	93.8%	97.5%	29.3%

System	Puzzle (%)
BCS QA + BP + LS	81.7
BCS QA + BP	44.3
BCS QA + Dr.Fill Solver	73.7
Dr.Fill QA + Dr.Fill Solver	70.5





Ablation Tests

ITALIAN (20 crosswords)

Experiment	Correct Words	Delta	Correct Letters	Delta
FULL	98,38%	-	99,48%	-
NO WebSearch	88,36%	-10,02%	94,91%	-4,57%
NO Morpho Filter	97,05%	-1,33%	99,15%	-0,33%
NO Lexicon	93,79%	-4,59%	98,18%	-1,30%
NO expert.ai KG	92,91%	-5,47%	97,55%	-1,93%
«NO Lexical Knowledge» setup	88,11%	-9,16%	95,65%	-3,57%
NO RB Module	93,67%	-4,71%	97,99%	-1,49%
NO CWDB-QQ	89,88%	-8,50%	96,21%	-3,27%
«NO CLUE» setup	78,82%	-18,45%	90,28%	-8,94



Next steps and targets

- Approaching 100% on Italian and American crosswords to beat humans
 - Improve Clause-classification and Morphological Filtering:
 - with current implementation >40% of remaining errors on Italian crosswords are due to wrong inflections)
 - plan to include expert.ai morphological analysis and classification technology
 - Introducing a Web-assisted iterative local-search after grid-filling
 - Combining Neural QQ and Neural QA approaches
 - Boost usage of KGs by taking into account: inflections, ontological relations, etc.
 - Change list merging based on the morphological clue-analysis (e.g. on Proper Noun answers Web Search is preferable)
- Extend to other languages:
 - French: improve the current implementation
 - German: we have a collaboration with Univ. of Nuremberg
 - Spanish: ?
 - ... other languages: ?
- Automatically generating puzzles to feed training.



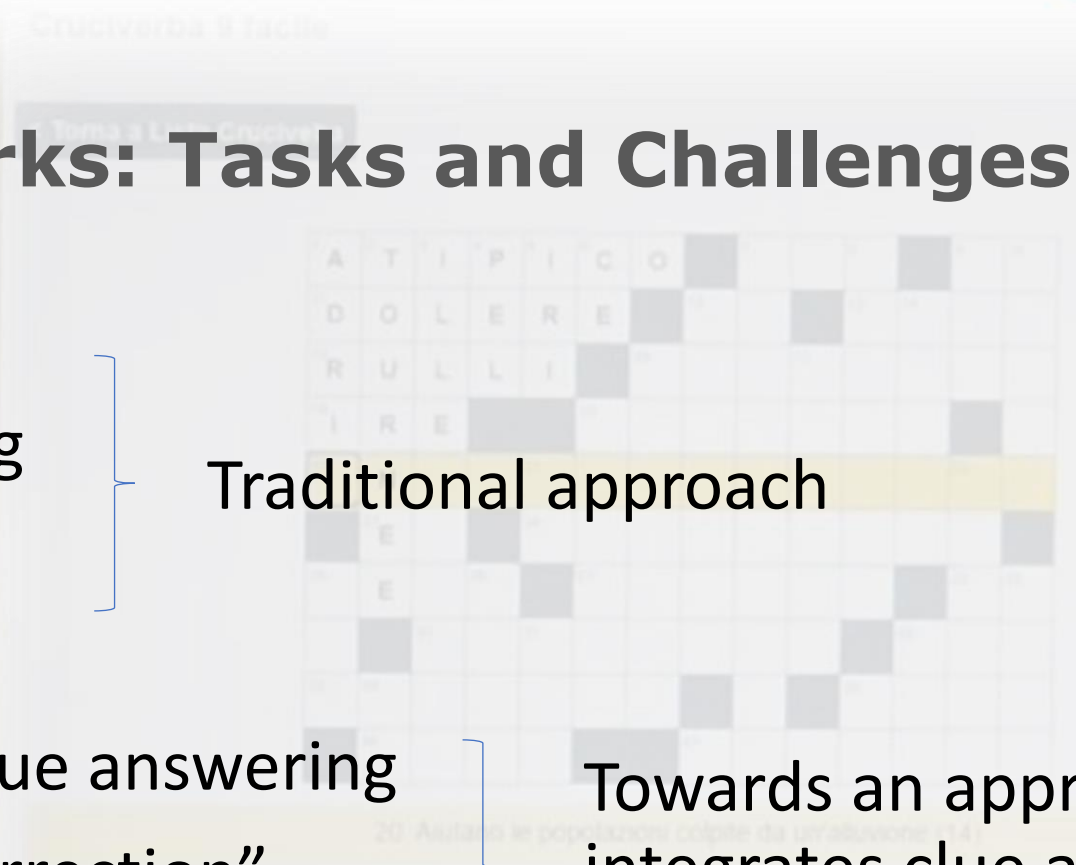
Benchmarks: Tasks and Challenges

- 1. Clue answering
- 2. Grid filling

Traditional approach

- 3. Constrained clue answering
- 4. Crossword “correction”
- 5. end-to-end NeSy model

Towards an approach that integrates clue answering and grid filling





Benchmarks: Tasks and Challenges

> 1. Clue Answering

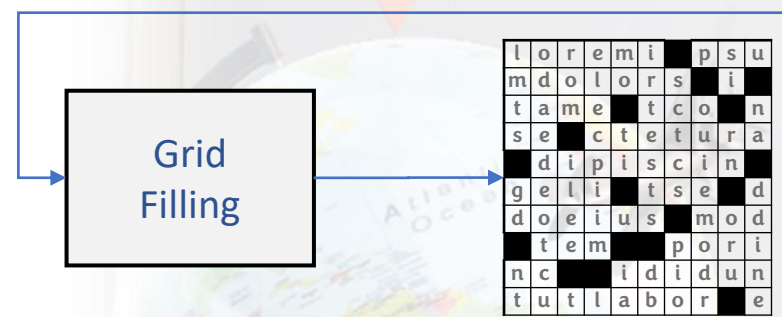
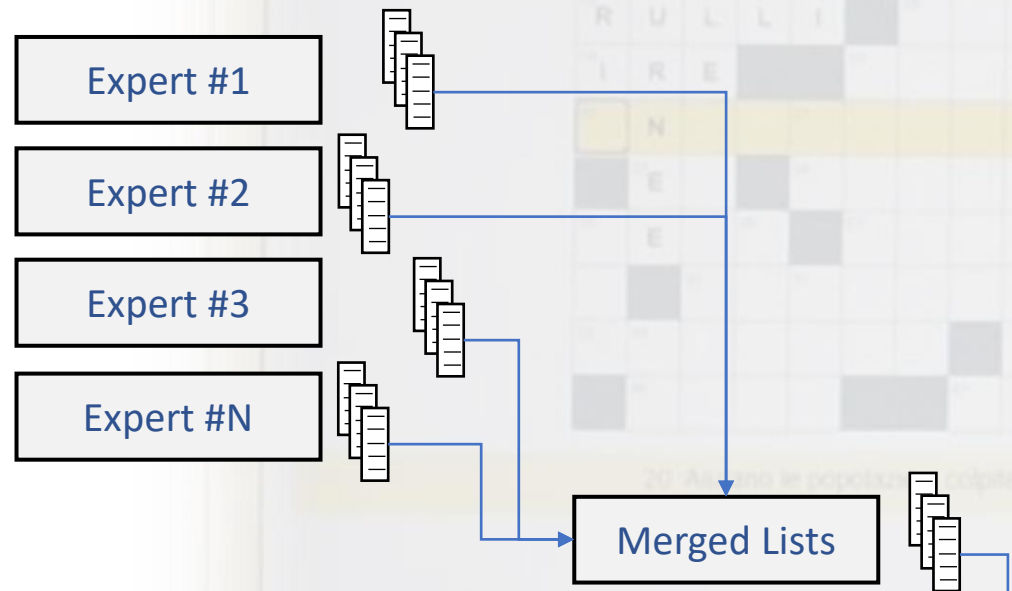
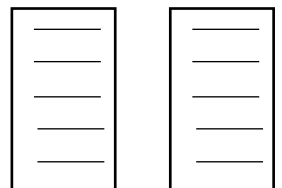
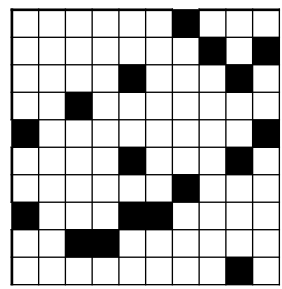
- We currently have the following resources:
 - American Crosswords
 - ~ 6M <clue,answer> pairs
 - ~ 3M unique <clue,answer> pairs
 - ~ 300K unique answers
 - ~ 50K solved puzzles
 - Italian Crosswords
 - ~ 125K <clue,answer> unique pairs
 - ~ 40K unique answers
 - ~ 2K solved puzzles
 - French Crosswords
 - ~ 340K <clue,answer> unique pairs
 - ~ 100K unique answers
 - ~ 7.5K solved puzzles
 - German Crosswords
 - ~ 2M <clue,answer> unique pairs
 - ~ 1.1M unique clues
 - ~ 430K unique answers
 - ~ 2.5K solved puzzles

We could extend the repo to other languages!



Benchmarks: Tasks and Challenges

> 2. Grid Filling



- Starting from the available test puzzles we could generate a great volume of <lists 2 grid> tasks
- Problems involved:
 - Learning2rank → how to re-rank answers
 - Merging → learn how to better merge answers
 - Improve ranking via constraints (e.g. belief propagation)
 - Grid filling (weighted/probabilistic CSP problem)



Benchmarks: Tasks and Challenges

> 3. Constrained clue answering

- Answering a clue with answers that respect a given pattern
 - E.g. “Nota associazione umanitaria”: C????? → **C**ARITAS (not **A**MNESTY, **W**EWORLD, **M**MEDICUS, ...)
 - E.g. “European capital”: D????N → **D**UBLIN (not **L**ONDON, **L**ISBON, **B**ERLIN, **T**IRANA, **W**ARSAW, ...)

- A possible solution: adding constraints to a generative Q/A approach ?
- Starting from available crosswords it is possible to build a specific benchmark of “constrained <clue,answer> pairs”.



ORIZZONTALI:

- Né tuo né suo (3)
- Formano la risma (5)
- Le iniziali del presidente Ciampi (3)**
- Il cantautore di Sapere di sale (5)
- ...

VERTICALI:

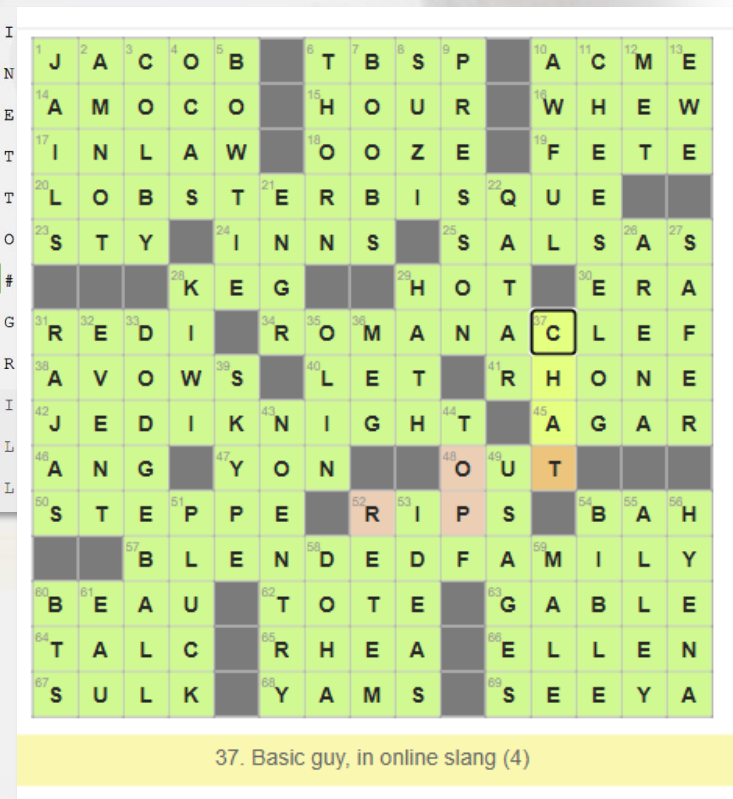
- Adesso... a Napoli (2)
- Coperta di mota (7)
- Acquisizioni di risultati importanti (11)
- Confermata solennemente (7)
- Ridurre in brandelli (8)
- "Incorporated" nelle abbreviazioni (3)
- Nota associazione umanitaria (7)**



Benchmarks: Tasks and Challenges

> 4. Crossword correction

- During grid-filling WebCrow maintains a “confidence” of the assignment of a letter to each cell
- Low-confidence cells tend to be prone to errors
- It is possible to think of a way to locally correct answers going back to the clue
- This challenge is technically very similar to the previous one, but:
 - we should be aiming at 100% accuracy
 - “constraints” could be soft (each cell could come with possible alternative characters and their probabilities)



- <Basic guy, in online slang: “CHAT” → CHAP
- <La consegna un professionista al suo cliente: “CARTAINTERTATA”> → CARTAINTESTATA

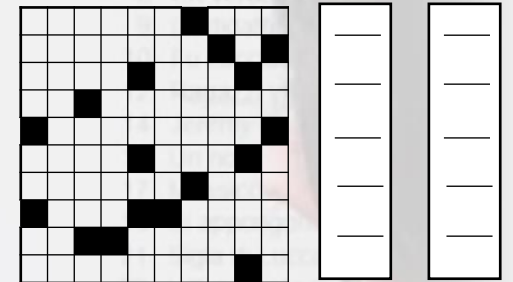


Benchmarks: Tasks and Challenges

> 5. end-to-end (NeSy model) crossword solving

- IF we “generate answers” with good quality and confidence
- IF given a few known letters we “generate answers” with high quality and confidence
- IF given a partially filled grid we estimate which are the most promising next clues to be answered
- IF given a large known portion of an answer we “generate the correct answer” with extreme quality and confidence
- IF given a filled crossword we can review&correct previously given answers

→ THEN we have built an end2end “human like” crossword solver !



Open questions:

- Could we train the entire process end2end?
- Could we make use of external knowledge?



Associazione
Italiana per
l'Intelligenza
Artificiale



WebCrow Team

Prof. Marco Gori

University of Siena

Prof. Marco Maggini

University of Siena

Kamyar Zeinalipour

University of Siena

Tommaso Iaquinta

University of Siena

Szymon Stefanek

University of Siena

Marco Ernandes

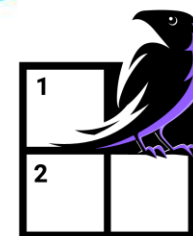
expert.ai

Giovanni Angelini

expert.ai

Andrea Zugarini

expert.ai



WEBCROW 2.0

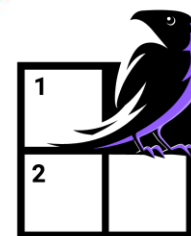
Thank You.



<http://webcrow.diism.unisi.it>



webcrow@unisi.it



WEBCROW 2.0