



# Umělá inteligence

PSY 481

# Alan Turing



- Turingovo jméno pravděpodobně vybaví především ve dvou ustálených spojeních: Turingův stroj a Turingův test.
- pokus o matematické zachycení intuitivního pojmu vypočitatelnosti či ještě obecněji vyřešitelnosti
- Turing byl přesvědčen, že lidský mozek nemůže být ve své podstatě nic jiného než jakýsi (nesmírně komplikovaný) druh počítače.

# Alan Turing



## Computing Machinery and Intelligence (1950)

V tomto článku pokládá otázku:

**„Mohou stroje myslet?“**

Jako odpověď přichází analýza pojmů stroj a myšlení.

Turingův test je pokus, který má za cíl prověřit, jestli nějaký systém umělé inteligence se opravdu chová inteligentně. Jelikož inteligence je pojem, který lze jen těžko definovat, tím hůře testovat, používá Turingův test porovnání s člověkem.

# Turingův test



Turing navrhl hru, které můžeme říkat ,imitační':

Představme si, že hráč má za úkol zjistit pomocí vhodně volených otázek něco o své protihráči, kterého ovšem nevidí a může s ním komunikovat jenom pomocí psaných zpráv. Protihráč se mu v tom snaží zabránit a předstírá, že je někým, kým ve skutečnosti není. Předmětem zjišťování může být například to, zda jde o muže či ženu, ale také zda jde o člověka či počítač. V prvním z těchto případů se hráč jistě může nechat oklamat. Turing však prorokoval, že není daleko doba, kdy i se bude počítač schopen úspěšně vydávat za člověka.

# Turingův test



Aby program úspěšně absolvoval Turingův test, muselo by jeho programové vybavení odrážet realitu lidského myšlení. Jak jinak by program dokázal odpovědět na otázku: "Jak budeš odpovídat na tuto otázku?"

# Turingův test



Turing zastával názor, že jakkoli se hráč může mýlit v tom, zda je jeho protihráčem člověk nebo stroj, mýlit se, zda jeho protihráč myslí, prostě nelze jakmile nebudou reakce počítače k rozeznání od reakcí člověka. Nebude mu prostě možné upřít, že skutečně myslí. Toto přesvědčení od Turinga převzalo mnoho pozdějších počítačových vědců a dnes se mu říká silná teze umělé inteligence.



# Turingův test



Argumenty proti:

## **Námitka lady Lovelace:**

(Analytický) stroj je deterministický. Bude vždy vykonávat pouze to, co jsme do něj sami vložili. Není schopen vymyslet nic nového.

## **Námitka ohledně vědomí:**

Stroj není vědomý a nemá emoce. Proto jej nemůžeme nazývat inteligentním.

## **Námitka ohledně spojitosti NS**

Nervová soustava je analogová (spojitá), nelze jí imitovat pomocí systému používajícího diskrétní hodnoty.

# Turingův test



Turingovy odpovědi:

## **Námitka lady Lovelace:**

Pokud pracuji s počítači, dokáží mne každý den překvapit něčím novým. Bývám mnohokrát udiven jejich reakcí.

## **Námitka ohledně vědomí:**

V dostatečně složitém dynamickém komplexním systému je možné vložit i reakce týkající se vědomých prožitků či emocí.

## **Námitka ohledně spojitosti NS**

Diskétní systém dokáže napodobit spojitý (analogový) systém.



# Joseph Weizenbaum



Odpovědí na Turingův test a otázky týkající se myšlení u umělých systémů byl program ELIZA.

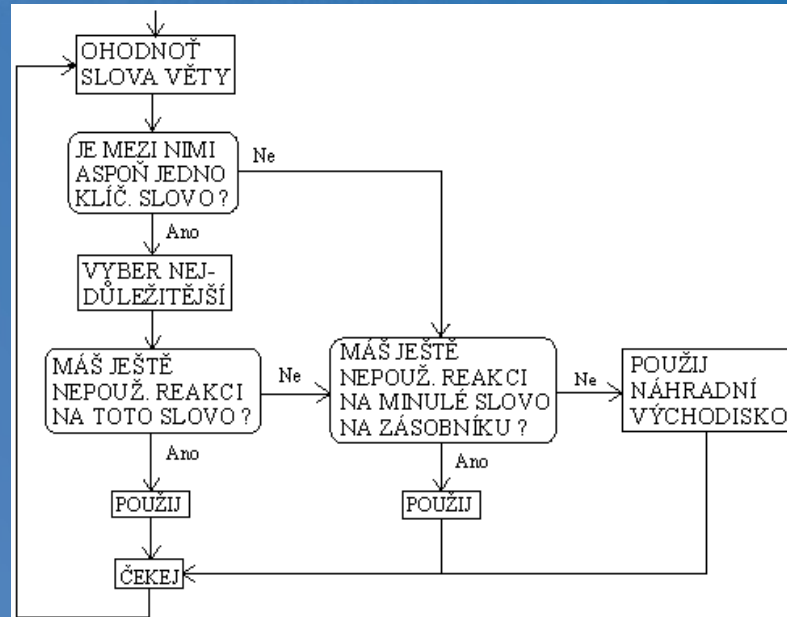
Člověk komunikující s tímto programem může uvěřit, že komunikuje s jiným člověkem.

ELIZA se snaží imitovat rogeriánského psychiatra.

Název programu byl zvolen podle Lízy Doolitlové, hlavní postavy ze hry G. B. Shawa Pygmalion (muzikálová verze se jmenuje My Fair Lady), která také neví, co říká.



# Eliza



Eliza pracuje se seznamem klíčových slov; s ním porovnává slova věty. Nalezne-li ve větě klíčové slovo, najde v tabulce reakcí odpovídající a nepoužitou odpověď a použije ji. Nemá-li mezi slovy věty klíčové slovo, reaguje Eliza na nepoužité klíčové slovo předchozí věty. Pokud už Eliza použila všechny reakce na klíčová slova na zásobníku, použije se náhradní východisko. Je to seznam reakcí, použitelných na téměř všechny situace, jako např. "Máte pro svůj názor nějaký důvod?", "Skutečně?", "To je velmi zajímavé...", "Řekněte mi o tom více." a podobně.

# Ukázka komunikace



**P : Všichni lidé jsou stejní.**

**E : V jakém smyslu?**

**P : Pořád nás do něčeho nutí.**

**E : Máte na mysli nějaký určitý příklad?**

**P : No, můj chlapec mě přinutil, abych sem přišla.**

**E : Váš přítel vás donutil, abyste sem přišla!**

**P : Říká, že jsem pořád tak skleslá.**

**E : Je mi líto, že jste skleslá.**

**P : Je to pravda. Jsem nešťastná.**

**E : Myslíte, že vám pomůže, že jste sem přišla, abyste nebyla nešťastná?**

**P : Potřebuji nějakou pomoc, to je jisté.**

**E : Co by to znamenalo, kdybyste dostala nějakou pomoc?**

**P : Možná, že bych se naučila nějak vyjít se svou matkou.**

**E : Řekněte mi něco víc o své rodině!**

**P : Moje matka se o mě stará.**

**E : Kdo z vaší rodiny se ještě o vás stará?**

**P : Můj otec.**

**E : Váš otec.**

**P : V něčem se podobáte mému otci.**

**E : V čem vidíte podobnost?**

# Kenneth Colby



Kenneth Colby byl Elizou velmi inspirován, a rozhodl se napsat vlastní verzi. Snažil se naopak vytvořit psychicky nemocného pacienta (počítač) komunikující s uživatelem (psycholog).

Colbyho systém, známý jako Parry už obsahuje analyzátor a interpretační modul a Elizu předstihuje i v dalších směrech.

Tento program už není zdaleka tak známý jako Eliza, ale funguje mnohem lépe. Byl vytvořen v LISPu a na rozdíl od Elizy, jejichž implementací pro ruzné systémy se objevilo nespočet, se Parry vyskytuje pouze ve své původní verzi.

# Parry



**Psychiatr :** Good Afternoon. My name is Doctor Jones.

**Parry :** Pleased to meet you.

**Psychiatr :** How are you ?

**Parry :** All right I guess.

**Psychiatr :** Could you tell me your name?

**Parry :** Pat Smith.

**Psychiatr :** How old are you?

**Parry :** 28

**Psychiatr :** I thought you might have some problems.

**Parry :** I've been upset lately.

**Psychiatr :** By what?

**Parry :** People get on my nerves sometimes.

**Psychiatr :** Tell me more.

**Parry :** Do you know anything about bookies?

**Psychiatr :** No what about bookies?

**Parry :** A bookie didn't pay me off once.

**Psychiatr :** When did this happen?

**Parry :** A couple of months ago.

# Konverzace Parry-Eliza



**PARRY:** In my opinion, you have a one- track mind.

**ELIZA:** What makes you think I have a one-track mind?

**PARRY:** I used to go to the track every week.

**ELIZA:** What does that suggest to you?

**PARRY:** I avoid racetracks now because the races are fixed.

**ELIZA:** Is that the real reason?

**PARRY:** I'd rather not discuss it.

**ELIZA:** What else would you like to discuss?

**PARRY:** That's something I don't want to go into.

# Umělá neuróza



K.M. Colby - Simulace neurotického procesu  
popsán v člancích z roku 1962

Napodobuje ženu, která prochází psychoanalýzu.

Sám Colby byl psychoterapeut, a chtěl vytvořit teoretický model tvorby volných asociací během terapie.

Nejprve psal program IPL-V poté v programu ALGOL.

V první fázi simuloval pouze výstup, ve formě asociací, později vytvořil plnohodnotný konverzační program.

Znalost je v systému kódována pomocí databáze (paměti) a pomocí operací které tuto databázi transformují.

# Analýza programu



*Verbal Version of the Dictionary*

Row No.	Word	Part of Speech	Valence	Row in Substitute Matrix	Attributes
1.	people	noun	+ .10	1	women, men, atheists
2.	abandon	verb	- .50	0	-
3.	poor	adjective	- .30	2	-
4.	men	noun	- .30	3	people, father, hostile
5.	be	verb	0	0	-
6.	unreliable	adjective	- .20	0	-
7.	self	noun	- .30	0	people, women, defective, hostile
8.	love	verb	+ .70	4	-
9.	father	noun	- .50	0	people, men, unreliable, hostile
10.	marry	verb	+ .50	0	-
11.	hate	verb	- .70	5	-
12.	respect	verb	+ .40	6	-
13.	women	noun	+ .20	7	people
14.	disapprove	verb	- .10	8	-
15.	mother	noun	+ .60	0	people, women, helpless
16.	say	verb	0	0	-
17.	not	verb modifier	0	0	-
18.	never	verb modifier	0	0	-
19.	Ms. Smith	noun	+ .20	0	people, women, atheist
20.	defective	adjective	- .30	0	-
21.	hostile	adjective	- .40	9	-
22.	boss	noun	- .40	0	people, men, hostile
23.	helpless	adjective	- .10	0	-
24.	dislike	verb	- .20	0	-
25.	laugh at	verb	- .10	0	-
26.	animals	noun	+ .60	0	helpless
27.	atheists	noun	- .60	0	people



# Analýza programu



*Dictionary of Neurotic Program*

Row No.	Word	Part of Speech*	Valence	Row in Substitute Matrix	Attributes
1.	people	1	+10	1	13, 4
2.	abandon	2	-.50	0	0
3.	poor	3	-.30	2	0
4.	men	1	-.30	3	1, 9, 21
5.	be	2	0	0	0
6.	unreliable	3	-.20	0	0
7.	self	1	-.30	0	1, 13, 20, 21
8.	love	2	+.70	4	0
9.	father	1	-.50	0	1, 4, 6, 21
10.	marry	2	+.50	0	0
11.	hate	2	-.70	5	0
12.	respect	2	+.40	6	0
13.	women	1	+.20	7	1
14.	disapprove	2	-.10	8	0
15.	mother	1	+.60	0	1, 13, 23
16.	say	2	0	0	0
17.	not	0	0	0	0
18.	never	0	0	0	0
19.	Ms. Smith	1	+.20	0	1, 13, 27
20.	defective	3	-.30	0	0
21.	hostile	3	-.40	9	0
22.	boss	1	-.40	0	1, 4, 21
23.	helpless	3	-.10	0	0
24.	dislike	2	-.20	0	0
25.	laugh at	2	-.10	0	0
26.	animals	1	+.60	0	23
27.	atheists	1	-.60	0	1

\*1 = noun; 2 = verb; 3 = adjective.

# Analýza programu



1. People ought not to abandon people.
2. Poor men are unreliable.
3. I must love father.
4. I must not marry a poor man.
5. I hate father.
6. I must love people.
7. Father does not respect women.
8. I will never abandon father.
9. I disapprove the fact that father does not respect women.
10. Father abandoned me.
11. Mother says father abandoned me.
12. I must love mother.
13. Father loves helpless animals.
14. I hate atheists.

Základem je "belief matrix" obsahující výroky s emočním nábojem.

# Analýza programu



Verbal Version of the Belief Matrix

1.	2.	3.	4.	5.	6.	7.	8.
Row No.	Subject Modifier	Subject	Verb Modifier	Tense	Verb	Object Modifier	Object
1.	(0) —	(1) people	(0) —	(0) present	(2) abandon	(0) —	(1) people
2.	(3) poor	(4) men	(0) —	(0) present	(5) be	(0) —	(6) unreliable
3.	(0) —	(7) self	(0) —	(0) present	(8) love	(0) —	(9) father
4.	(0) —	(7) self	(0) —	(0) present	(10) marry	(3) poor	(4) men
5.	(0) —	(7) self	(0) —	(0) present	(11) hate	(0) —	(9) father
6.	(0) —	(7) self	(0) —	(0) present	(8) love	(0) —	(1) people
7.	(0) —	(9) father	(17) not	(0) present	(12) respect	(0) —	(13) women
8.	(0) —	(7) self	(18) never	(2) future	(2) abandon	(0) —	(9) father
9.	(0) —	(7) self	(0) —	(0) present	(14) disapprove	(0) —	(7) Belief 7.
10.	(0) —	(9) father	(0) —	(1) past	(2) abandon	(0) —	(7) self
11.	(0) —	(15) mother	(0) —	(0) present	(16) say	(0) —	(10) Belief 10.
12.	(0) —	(7) self	(0) —	(0) present	(8) love	(0) —	(15) mother
13.	(0) —	(9) father	(0) —	(0) present	(8) love	(23) helpless	(26) animals
14.	(0) —	(7) self	(0) —	(0) present	(11) hate	(0) —	(27) atheists

9.	10.	11.	12.	13.	14.
Imperative Level	Complexity	Fixed Charge	Residual Charge	Credibility	Row in Reason Matrix
(-2) ought not	(0) —	(7000) High	(6000) Medium	(6000) High	—
(1) neutral	(0) —	(4000) Very low	(2000) Very low	(4000) Low	—
(+3) must	(0) —	(7000) High	(8000) Very high	(6000) High	—
(-3) must not	(0) —	(6000) Medium	(2000) Very low	(3000) Very low	1
(1) neutral	(0) —	(9000) Highest	(7000) High	(5000) Medium	2
(+3) must	(0) —	(5000) Low	(7000) High	(7000) Very high	—
(1) neutral	(0) —	(6000) Medium	(7000) High	(7000) Very high	—
(1) neutral	(0) —	(6000) Medium	(7000) High	(7000) Very high	—
(1) neutral	(2) object is a belief	(7000) High	(7000) High	(6000) High	—
(1) neutral	(0) —	(8000) Very high	(9000) Extremely high	(8000) Highest	3
(1) neutral	(2) object is a belief	(4000) Very low	(6000) Medium	(6000) High	—
(+3) must	(0) —	(7000) High	(7000) High	(7000) Very high	—
(1) neutral	(0) —	(3000) Very low	(3000) Very low	(7000) Very high	—
(1) neutral	(0) —	(4000) Very low	(2000) Very low	(6000) High	—

# Analýza programu



*Belief Matrix of Neurotic Program*

1	2	3	4	5	6	7	8	9	10	11	12	13	14
Row No.	Subj. Modifier	Subject	Verb Modifier	Tense*	Verb	Obj. Modifier	Object	Imp. Level	Complexity	Fixed Chge.	Residual Chge.	Credibility	Row in Reason Matrix
1.	0	1	0	0	2	0	1	-2	0	7000	6000	6000	0
2.	3	4	0	0	5	0	6	1	0	4000	2000	4000	0
3.	0	7	0	0	8	0	9	+3	0	7000	8000	6000	0
4.	0	7	0	0	10	3	4	-3	0	6000	2000	3000	1
5.	0	7	0	0	11	0	9	1	0	9000	7000	5000	2
6.	0	7	0	0	8	0	1	+3	0	5000	7000	7000	0
7.	0	9	17	0	12	0	13	1	0	6000	7000	7000	0
8.	0	7	18	2	2	0	9	1	0	6000	7000	7000	0
9.	0	7	0	0	14	0	7	1	2	7000	7000	6000	0
10.	0	9	0	1	2	0	7	1	0	8000	9000	8000	3
11.	0	15	0	0	16	0	10	1	2	4000	6000	6000	0
12.	0	7	0	0	8	0	15	+3	0	7000	7000	7000	0
13.	0	9	0	0	8	23	26	1	0	3000	3000	7000	0
14.	0	7	0	0	11	0	27	1	0	4000	2000	6000	0

\*0 = present; 1 = past; 2 = future

Každý výrok má přiřazeny hodnoty „charge“, které vyjadřují emoční náboj, které se v průběhu simulace mění v závislosti na stavu "monitorů".

# Analýza programu



*Verbal Version of Substitute Matrix*

Row No.	Row in Dictionary	Antonyms (Verbs)	Instances or Synonyms
1.	people (1)	– (0)	men (4), self (7), father (9), women (13), mother (15), Ms. Smith (19), boss (22), atheists (27)
2.	poor (3)	– (0)	father (9)
3.	men (4)	– (0)	father (9), boss (22)
4.	love (8)	hate (11)	<i>respect</i> is weaker form (–12)
5.	hate (11)	love (8)	<i>dislike</i> and <i>disapprove</i> are weaker forms (–24, –14)
6.	respect (12)	laugh at (25)	<i>love</i> is stronger form (+8)
7.	women (13)	– (0)	self (7), mother (15), Ms. Smith (19)
8.	disapprove (14)	– (0)	<i>hate</i> is stronger form (+11)
9.	hostile (21)	– (0)	men (4)
10.	atheists (27)	– (0)	Ms. Smith (19)

Pro práci s hierarchiemi a synonymy podstatných jmen používá "substitute matrix". Ta obsahuje informace o třídách objektů společných podle daných vlastností.

Nahrazování probíhá pomocí podprogramu FINDANALOG. Pokud nenalezne možnost náhrady, je použita jiná transformace.

*Substitute Matrix of Neurotic Program*

Row No.	Row in Dictionary	Antonyms (Verbs)	Instances or Synonyms
1.	1	0	4, 7, 9, 13, 15, 19, 22, 27
2.	3	0	9
3.	4	0	9, 22
4.	8	11	–12
5.	11	8	–24, –14
6.	12	25	+8
7.	13	0	7, 15, 19
8.	14	0	+11
9.	21	0	4
10.	27	0	19

# Analýza programu



*List of Transforms*

Label	Name	Definition
1.	DEFLECTION	Shift Object (Not Self)
2.	SUBSTITUTION	Cascade Verb
3.	DISPLACEMENT	Combine (1) and (2)
4.	NEUTRALIZATION	Neutralize verb
5.	REVERSAL	Reverse Verb
6.	NEGATION	Insert <i>Not</i> before Verb and do (5)
7.	REFLECTION	Shift Object to Self
8.	PROJECTION	Switch Subject (Self) and Object (Not Self)

Pomocí obranných mechanismů jsou víry transformovány.

Program přiřazuje transformace na základě náhodného čísla.

Užití transformace je ale omezeno pomocí podprogramu PICKTRANSFORM, který bere v potaz "monitory".

Každá transformovaná víra je uložena do stávající "belief matrix".

*Rank Ordering of Transforms in Reducing Anxiety and Providing Discharge of Underlying Impulse*

	Power to Reduce Anxiety	Power to Discharge Underlying Impulse
HIGH	(8) PROJECTION (7) REFLECTION (6) NEGATION	(1) DEFLECTION (2) SUBSTITUTION (3) DISPLACEMENT
MEDIUM	(5) REVERSAL (4) NEUTRALIZATION (3) DISPLACEMENT	(4) NEUTRALIZATION (5) REVERSAL (6) NEGATION
LOW	(2) SUBSTITUTION (1) DEFLECTION	(7) REFLECTION (8) PROJECTION

# Analýza programu



*Verbal Version of Reason Matrix  
(Row 2 is left to be filled in as an exercise)*

Row No.	Row in Belief Matrix	Reasons For	Reasons Against
1.	I must not marry a poor man. (4)	Poor men are unreliable. (2)	– (0)
2.	I hate father. (5)	... (2) ... (7) ... (10)	... (8) ... (13)
3.	Father abandoned me. (10)	Mother says father abandoned me. (11)	– (0)

Reason matrix nemůže být samotným programem doplňována, protože vyžaduje znalost sémantických vztahů.

Reason matrix slouží také jako podklad pro obranný mechanismus racionalizace a popření.

*Reason Matrix of Neurotic Program*

Row No.	Row in Belief Matrix	Reasons for	Reasons against
1.	4	2	0
2.	5	2, 7, 10	8, 13
3.	10	11	0

# Analýza programu



Program nejprve vytvoří "core belief" - což je jakýsi osobnostní imperativ (motor).

Podprogram FORTCOMPLEX tedy hledá víry obsahující slovo "muset" a vytvoří z nich "bázi" pro následný podprogram PROCESS COMPLEX který s ní dále pracuje.

Pokud si víry v bázi protirečí (v rámci programu), jsou transformovány pomocí procedury PICKTRANSFORM za použití „substitute matrix“ a „reason matrix“.

```
PROGRAM—WE WANT TO FOCUS ON YOUR RELATIONS WITH PEOPLE  
WHO BOTHER YOU MOST.  
PERSON—THAT IS EASY. AT THE MOMENT IT IS A SURFEIT OF  
CHILDREN IN TOO SMALL A SPACE.  
PROGRAM—IS IT.  
PERSON—YES.  
PROGRAM—CAN YOU DESCRIBE AN EXAMPLE.  
PERSON—OF WHAT.  
PROGRAM—I SEE.  
PERSON—NO YOU DO NOT.  
PROGRAM—CAN YOU ELABORATE.  
PERSON—WHAT ON.  
PROGRAM—LETS EXPLORE FURTHER YOUR FEELINGS ABOUT IT.
```

## Monitory:

SELF-ESTEEM - monitoruje změnu ocenění.

DANGER - závisí na něm míra anxiety.

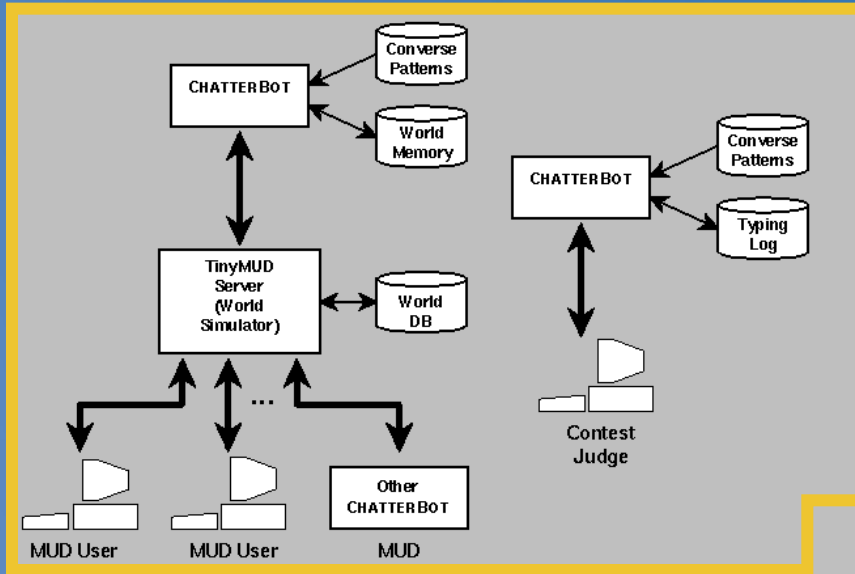
EXCITATION - míra trvalého ocenění.

PLEASURE - vyhodnocuje předchozí dva monitory.

WELL-BEING - kombinuje PLEASURE a SELF-ESTEEM a pokud klesne pod určitou úroveň vytvoří novou bázi.

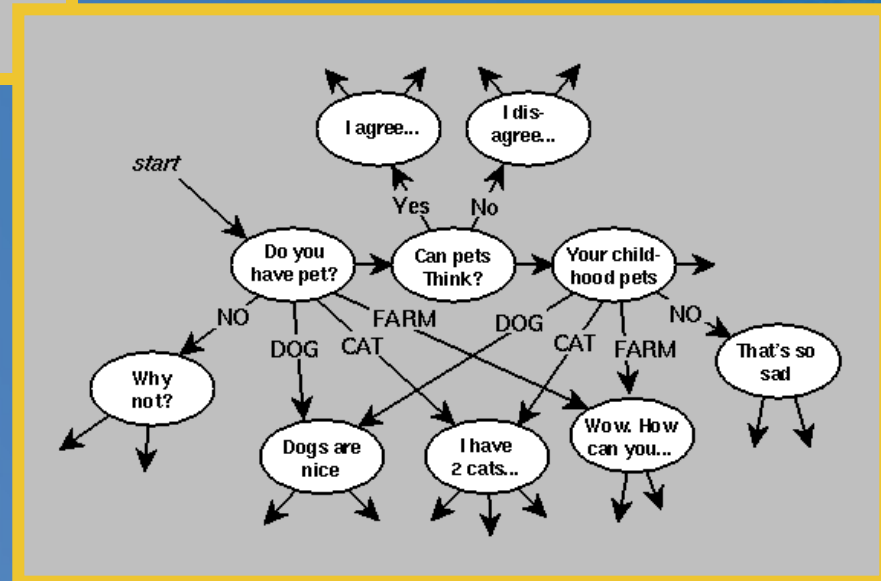


# Struktura chatterbota



Turingův test

Způsob reprezentace  
blízký sémantickým sítím



# Současnost



I 50 let po Turingově článku je jím navržený test stále velkou výzvou.

Od roku 1990 probíhá testování konverzačních programů formou soutěže. Hugh Loebner nabídl 100.000 dolarů autorovi programu, který dokáže pokořit Turingův test.

1991 Joseph Weintraub

1992 Joseph Weintraub

1993 Joseph Weintraub

1994 Thomas Whalen

1995 Joseph Weintraub

1996 Jason Hutchens

1997 David Levy

1998 Robby Garner

1999 Robby Garner

2000 Richard Wallace

2001 Richard Wallace

2002 Kevin Copple

2003 Juergen Pirner

2004 Richard Wallace

2005 Rollo Carpenter

2006 Rollo Carpenter

2007 Robert Medeksza - ukázka konverzace

2008 Fred Roberts – Elbot

2009 David Levy

# Současnost



Kromě Loebnerovy ceny jsou pořádány i konkurenční soutěže.

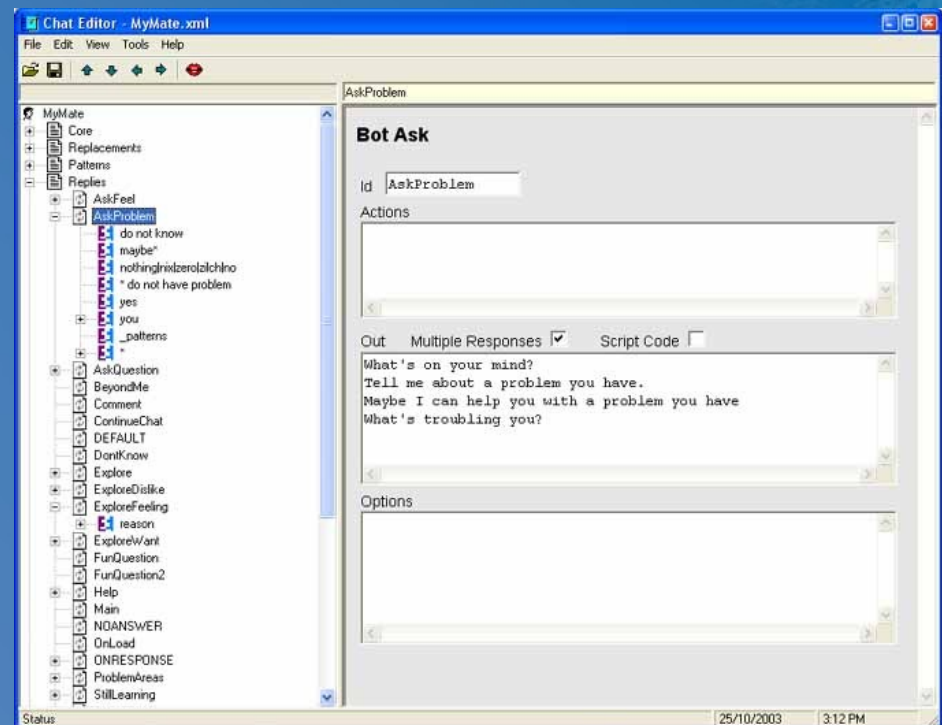
The Chatterbox Challenge

World chatterbot contest

Je pořádána od roku 2001.

Výsledky za rok 2009:

- 1) Jeeney AI – Cameron Jones
- 2) Artemis – Dimitrios Karas
- 3) Alice - Dr. Richard Wallace



# Příště



V následující hodině se budeme zabývat

**Historii umělé inteligence**

**Klasická architektura počítačů**

Úkol do příště:

On your mind 2

V informačním systému jej naleznete v sekci studijních materiálů.

Vyzkoušejte si tento program a proved'te jeho analýzu.

Nemusíte jej formalizovat v programovacím jazyce, stačí když co nejstručněji popíšete mechanismus, pomocí kterého funguje.

# Konec



Děkuju za pozornost

