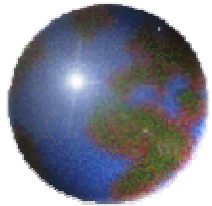




# Model and Counterexample Search: Successes and Challenges

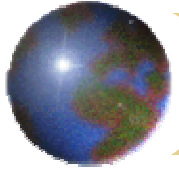


*Jian Zhang*

`zj@ios.ac.cn`

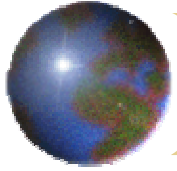
Chinese Academy of Sciences

Aug. 16, 2006



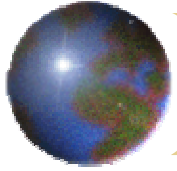
# *Outline*

- ✦ Introduction
- ✦ Some Successes
- ✦ Two Benchmark Problems
- ✦ Issues and Challenges
- ✦ Conclusion



## *Introduction*

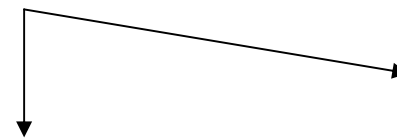
- ⊕ Given  $A$  (Axioms + Premises) and  $C$  (conjecture).  $A$  and  $C$  are logical formulas.  
Question:  $A \rightarrow C$  ?
- ⊕ Proving: Trying to show  $C$  is a theorem.
  - ⊗ Resolution-based:  $A \cup \{\sim C\}$  is unsatisfiable
  - ⊗ Natural Deduction
  - ⊗ Other methods
- ⊕ Disproving: Trying to find a counterexample.



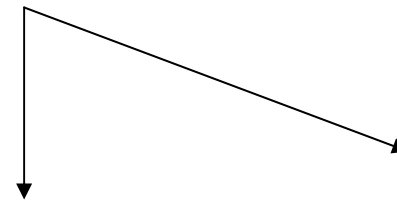
# Model Finding

- ✦ Counterexample:  
model of  $A \cup \{\sim C\}$
- ✦ Propositional Logic --  
SAT
- ✦ First-order Logic --  
*Finite* model finding  
Given  $F$  (set of first-order  
formulas) and  $k$  (*positive  
integer*), find a  $k$ -  
element model of  $F$ .

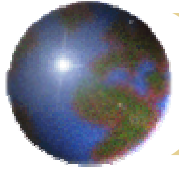
Disproving



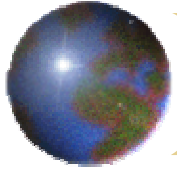
Finding a Counterexample



Finding a Finite Model  
(of some fixed size)



# Some Successes



## Example 1. Combinatory Logic

J. Zhang, *AARN* (1992); *CADE-12* (1994)

A: forall x forall y

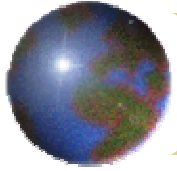
$$\oplus a(a(a(0,x),y),z) = a(x,a(y,z))$$

$$\oplus a(a(a(1,x),y),z) = a(a(a(x,y),y),z)$$

C: forall y exists x

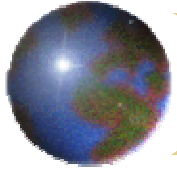
$$\oplus a(y,x) \neq a(x,a(a(y,x)))$$

a	0	1	2	3	4
0	0	0	2	3	4
1	0	0	2	3	4
2	3	3	3	3	3
3	3	3	3	3	3
4	4	4	4	4	4



## *Researchers – a partial list*

- ✦ Gilles Audemard
- ✦ Belaid Benhamou
- ✦ Thierry Boy de la Tour
- ✦ Francois Bry
- ✦ Ricardo Caferra
- ✦ Koen Claessen
- ✦ **Masayuki Fujita**
- ✦ Ryuzo Hasegawa
- ✦ Laurent Henocque
- ✦ Rainer Manthey
- ✦ Bill McCune
- ✦ Nicolas Peltier
- ✦ Niklas Sorensson
- ✦ Mark Stickel
- ✦ John Slaney
- ✦ Tanel Tammet
- ✦ Hantao Zhang
- ✦ Jian Zhang



## Example 2. Quasigroup Identities

M. Fujita *et al.*, *IJCAI* (1993)

forall x forall y

⊕  $f(x,y) \neq f(x,z) \mid$

$y = z$

⊕  $f(x,z) \neq f(y,z) \mid$

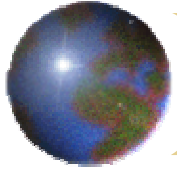
$x = y$

Plus some identities

⊕  $f(x,x) = x$

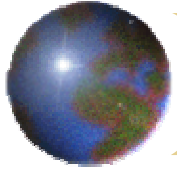
⊕  $f(f(f(y,x),y),y) = x$

f	0	1	2	3	4	5	6
0	0	2	3	1	6	4	5
1	2	1	4	5	0	6	3
2	3	4	2	6	5	1	0
3	1	5	6	3	2	0	4
4	6	0	5	2	4	3	1
5	4	6	1	0	3	5	2
6	5	3	0	4	1	2	6



## *Example 3. Ortholattice Identities*

- ✦ William McCune, Automatic Proofs and Counterexamples for Some Ortholattice Identities. *Information Processing Letters* 65(6): 285-291 (1998)
  - ✦ This note answers questions on whether three identities known to hold for orthomodular lattices are true also for ortholattices. One identity is shown to fail by MACE, a program that searches for counterexamples, ...



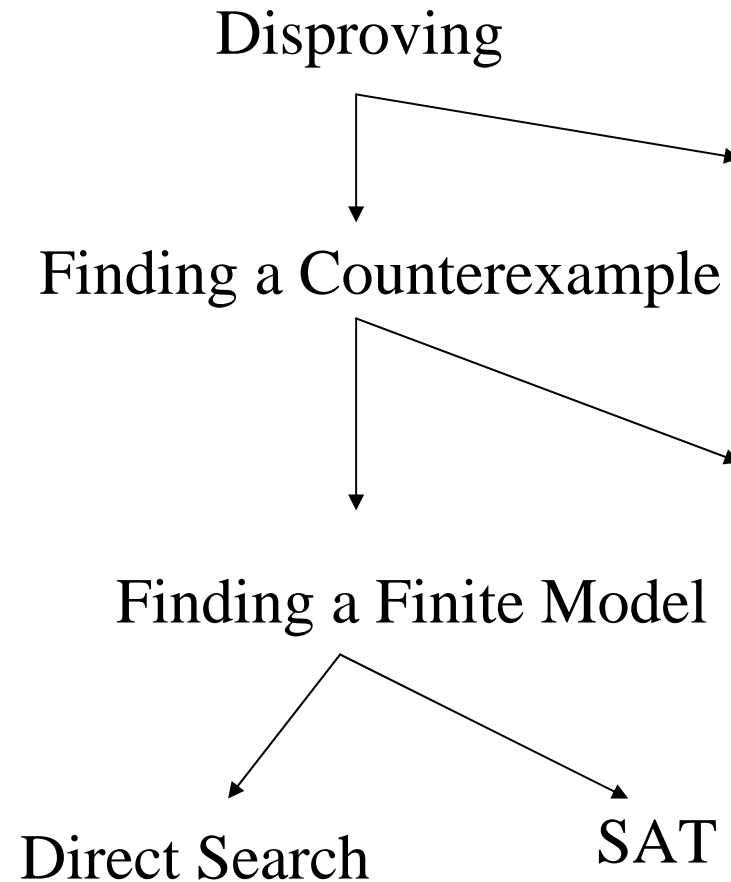
# *Direct Search vs. SAT*

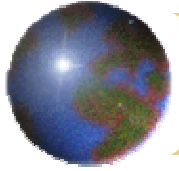
## ✚ Direct Search

Search for the cells' values directly

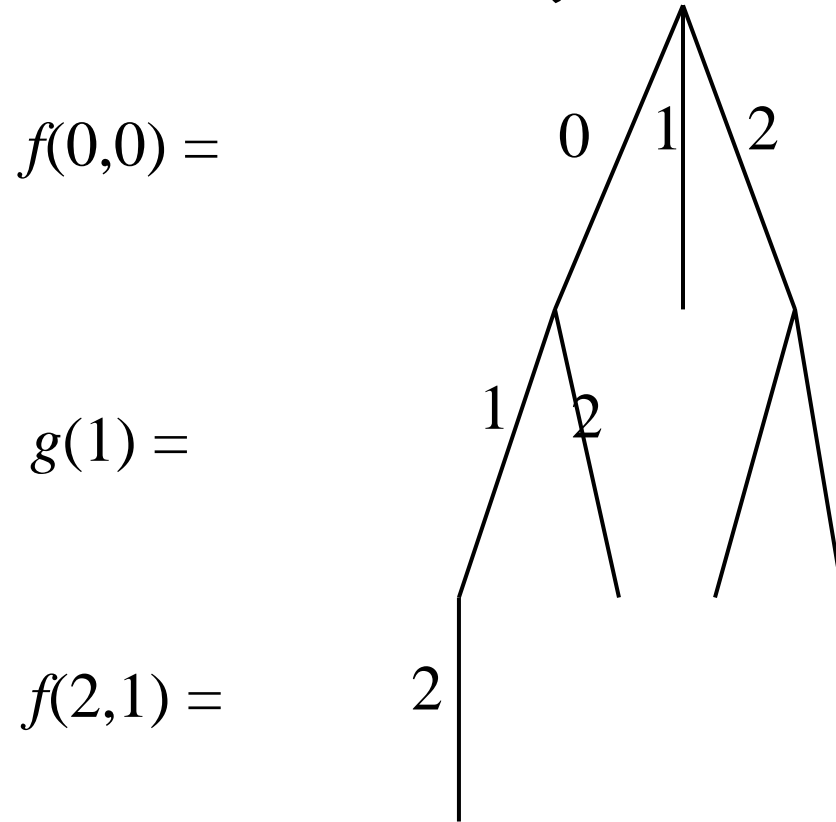
(**FINDER**, **SEM**,  
**MACE4**, ...)

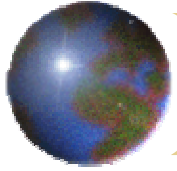
## ✚ Transforming to SAT (**MACE2**, **Paradox**, ...)





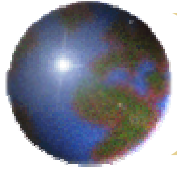
# Direct search (non-binary search tree)





## Comparison

- ✦ The SAT-based method can benefit from efficient SAT solvers.
- ✦ Direct Search methods can exploit structural information and reason in larger steps
  - e.g. the least number heuristic for avoiding symmetric subspaces

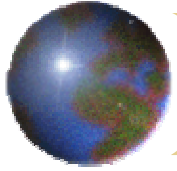


## *Isomorphism / Symmetry*

f	0	1	2	3
0	0	2	3	1
1	3	1	0	2
2	1	3	2	0
3	2	0	1	3

f	0	1	2	3
0	0	3	1	2
1	2	1	3	0
2	3	0	2	1
3	1	2	0	3

$\langle 2, 3 \rangle$



## Constraint propagation in direct search

### ✚ Rewriting

$$\boxtimes f(1, g(2)) = 3 \text{ AND } g(2) = 0$$



$$\boxtimes f(1, 0) = 3$$

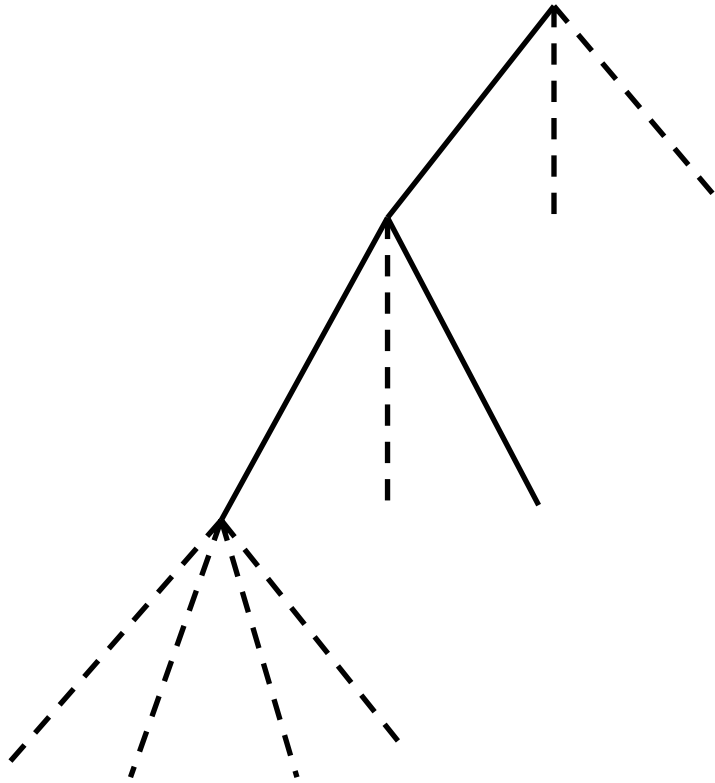
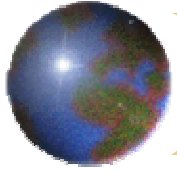
### ✚ Negative propagation

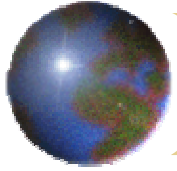
$$\boxtimes h(1, g(f(2, 4)), 3) \neq 5$$

$$\boxtimes h(1, 2, 3) = 5 \text{ AND } g(1) = 2$$



$$\boxtimes f(2, 4) \neq 1$$





## Combination of Model Finding with Others (I)

### ❖ Concurrent “decision procedures”

Given a formula  $H$ ,

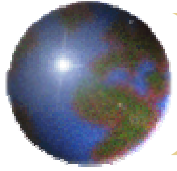
we run two processes concurrently:

$proc_1$ : show  $H$  is unsat. by using resolution

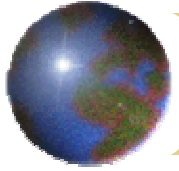
$proc_2$ : find a model of  $H$  (trying size 1, 2, ...)

❖ The above program is a decision procedure for a large subset of first-order formulas:

{  $H$  |  $H$  is either unsat. or sat. in some finite domain }

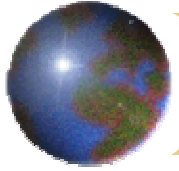


- ❖ Many propositional modal logics can be translated into the classical logic, and the satisfiability prob. can be solved in this way. **reason:** *finite model property*
- ❖ Propositional temporal logic LTL:
  - The smallest model can be found.
  - Example.  $\bigcirc\bigcirc\square p$  has a one-state model, in which  $p$  is true in the state. With the tableau method, we get a 3-state model.
- [Zhang J., Tech.Rep., Dec. 1996]

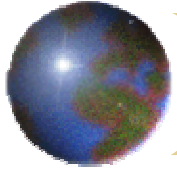


## Combination of Model Finding with Others(II)

- ✦ By looking at the models of a theory, we learn something about the theory.
  - ✦ It might be inconsistent, if you can't find some finite models.
  - ✦ Automatic/semi-automatic generation of simple conjectures/lemmas
  
- [Zhang JAR 1996] [Zhang CADE 1999]



# Two Benchmark Problems



## Example 4. Tarski's High School Problem

J. Zhang, *CADE-12* (1994) *CADE-20* (2005)

Axioms — HSI

$$\oplus x + y = y + x$$

$$\oplus x + (y + z) = (x + y) + z$$

$$\oplus x^* 1 = x$$

$$\oplus x^* y = y^* x$$

$$\oplus x^* (y^* z) = (x^* y)^* z$$

$$\oplus x^* (y + z) = (x^* y) + (x^* z)$$

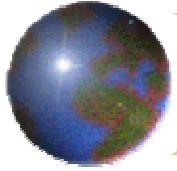
$$\oplus 1^{\wedge} x = 1$$

$$\oplus x^{\wedge} 1 = x$$

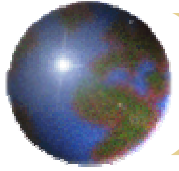
$$\oplus x^{\wedge} (y + z) = x^{\wedge} y^* x^{\wedge} z$$

$$\oplus (x^* y)^{\wedge} z = x^{\wedge} z^* y^{\wedge} z$$

$$\oplus (x^{\wedge} y)^{\wedge} z = x^{\wedge} (y^* z)$$



- ✦ Tarski's *High School Problem*: Does HSI serve as a basis for all the identities of  $\mathcal{N}$ ?
- ✦ Wilkie provided a negative answer to the problem in the early 1980s.
- ✦  $W(x,y)$  holds for all natural numbers, but it does not follow from HSI.



## Wilkie's identity $W(x,y)$

$$(P^x + Q^x)^y * (R^y + S^y)^x = \\ (P^y + Q^y)^x * (R^x + S^x)^y$$

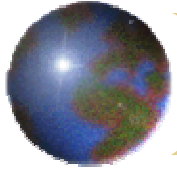
where

$$\oplus P = 1 + x$$

$$\oplus Q = 1 + x + x^*x$$

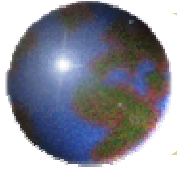
$$\oplus R = 1 + x^*x^*x$$

$$\oplus S = 1 + x^*x + x^*x^*x^*x$$



- ✦ Find a small counterexample/model (by computer programs). [Zhang 1994]  
HSI + {  $\sim W(a,b)$  }
- ✦ Finite models found by Gurevic and by Burris, Jackson, Lee and Yeats:  
of size 59, 33, ..., 15, 14, 12.
- ✦ Is there a model of size 11 ? OPEN
  - ✦ Unlikely to exist -- Stan Burris





## Some Helpful Lemmas

[Burris-Lee 1992] [Jackson 1996]

⊕  $b \neq a. \quad b \neq 1. \quad b \neq 2. \quad // \quad 2 = s(1,1).$

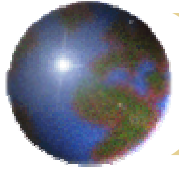
⊕  $b \neq p(a,x). \quad \dots$

⊕  $s(1,a) \neq 1. \quad s(2,a) \neq 1. \quad s(a,a) \neq 1.$

⊕  $p(a,a) \neq 1.$

⊕  $b \neq s(1, p(1,a)). \quad b \neq s(1, p(2,a)).$

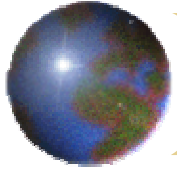
⊕  $\dots$



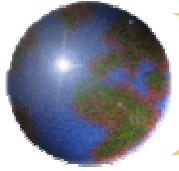
## *Some Experimental Results*

Earlier results:

- ✚ No GBA of size 7. [Zhang (Ph.D.) 1994]
- ✚ No GBA of size 8. [Zhang and Zhang 1995]
  
- ✚ Known lower bounds (by mathematicians):
  - $\geq 7$  [Burris-Lee 1992]
  - $\geq 8$  [Jackson 1995? 1996]

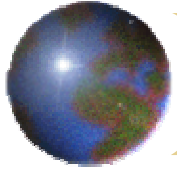


- ✦ New Results in my CADE-20 paper:
  - ▣ No GBA of size 9.
  - ▣ No GBA of size 10.
- ✦ Experiments on a desktop (in 2004) --  
Dell Optiplex GX270: Pentium 4, 2.8 GHz CPU,  
2G memory) running RedHat Linux
- ✦ Programs used: SEM, Mace4 (mace4-2004-C)
  - ▣ hundreds of input files



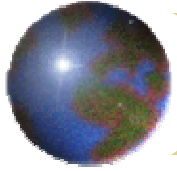
## Some experiences

- ❖ Symmetry elimination is crucial.
- ❖ It is better to turn off the negative propagation rules in SEM and Mace4.
- ❖ The lemmas are very helpful in general  
-- with a few exceptions.
- ❖ Other aspects: human guidance, managing many subsearches, ...



## *More Questions*

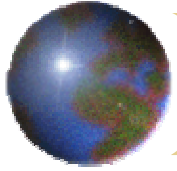
- ✚ Can we find all the 12-element countermodels to Wilkie's identity? (or a different model)
- ✚ Can we find some counterexamples to other identities? e.g. Gurevic (1990):  
$$(P^x + Q^x)^{2^x} * (R^{2^x} + S^{2^x})^x = (P^{2^x} + Q^{2^x})^x * (R^x + S^x)^{2^x}$$
- ✚ Can we disprove Tarski's High School problem without knowing the works of Wilkie and Gurevic?



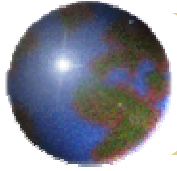
## *Example 5. Large set of idempotent QGs*

L. Zhu 2006

- ✦ A quasigroup  $(Q, f)$  is *idempotent* if the identity  $f(x, x) = x$  holds for all  $x$  in  $Q$ .
- ✦ Two idempotent quasigroups  $(Q, f_1)$  and  $(Q, f_2)$  are said to be *disjoint* if for any  $x, y$  in  $Q$ ,  $f_1(x, y) \neq f_2(x, y)$  whenever  $x \neq y$ .
- ✦ A collection of idempotent quasigroups  $(Q, f_1), (Q, f_2), \dots, (Q, f_n)$ , where the size of  $Q$  is  $(n+2)$ , is called a *large set* if any two of the idempotent quasigroups are disjoint.



- ✚ A quasigroup  $(Q, f)$  is called a *Steiner pentagon quasigroup* if it satisfies the identities:  
$$f(x, x) = x, f(f(y, x), x) = y, f(x, f(y, x)) = f(y, f(x, y)).$$
- ✚ A large set of Steiner pentagon quasigroups of order  $\nu$  is denoted  $\text{LSPQ}(\nu)$ .
- ✚ Problem: does  $\text{LSPQ}(\nu)$  exist, for  $\nu = 11, 21 \dots$  ?
- ✚ Initial Experiments



## SEM input

( elem [5] )

{ f1 : elem elem -> elem  
(QG) }

{ f2 : elem elem -> elem  
(QG) }

{ f3 : elem elem -> elem  
(QG) }

[ x = y | f1(x,y) != f2(x,y) ]

[ x = y | f2(x,y) != f3(x,y) ]

[ x = y | f3(x,y) != f1(x,y) ]

[ f1(x,x) = x ]

[ f1(f1(y,x),x) = y ]

[ f1(x,f1(y,x)) = f1(y,f1(x,y)) ]

[ f2(x,x) = x ]

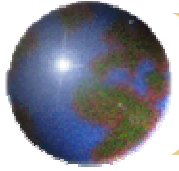
[ f2(f2(y,x),x) = y ]

[ f2(x,f2(y,x)) = f2(y,f2(x,y)) ]

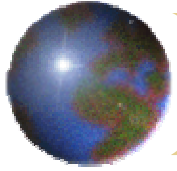
[ f3(x,x) = x ]

[ f3(f3(y,x),x) = y ]

[ f3(x,f3(y,x)) = f3(y,f3(x,y)) ]

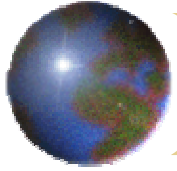


# Issues and Challenges



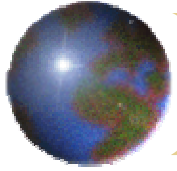
## *Efficiency -- Search Techniques*

- ✦ More search heuristics (for selecting cells/values)
- ✦ Symmetry Elimination
- ✦ Constraint Propagation
  - ▣ Compare with propositional unit propagation
  - ▣ Compare with consistency techniques in CSP
- ✦ Parallel search: MGTP, PSATO
- ✦ ...



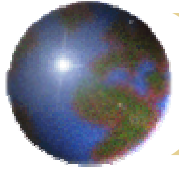
## *Re-examining the "Standard" Approaches*

- ✦ Express the axioms, premises and the conjecture in **the first-order predicate logic**.
- ✦ Transform the formulae  $A + \{ \sim C \}$  into **clauses**.
- ✦ Assume that there is a **finite** model.
- ✦ Obtain a set of **ground (or propositional)** clauses.
- ✦ Perform the search, using various **heuristics**.
- ✦ Accept the **results**.



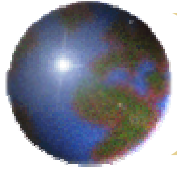
## *New Logic – Description Logics*

- ✚ Semantic Web: HTML → RDF, OWL, ...
- ✚ Basis: description logics
  - ▣ Concepts: Set of individuals, unary preds
  - ▣ Roles: Binary predicates
- ✚ Number restrictions, e.g. ( $\geq 2$  hasChild)
- ✚ Translation into FoL may generate too long formulae (with many variables).



## *New Logic – Higher-order Logics*

- ✦ Berghofer and Nipkow, SEFM'04
  - ✦ When developing non-trivial formalizations in a theorem prover, a considerable amount of time is devoted to “debugging” specifications and conjectures by failed proof attempts.
- ✦ Tjark Weber, Bounded Model Generation for Isabelle/HOL, Disproving 2004.
  - ✦ translation from HOL to propositional logic, s.t. the resulting propositional formula is satisfiable iff the HOL formula has a model of a given finite size.



## *Working with Ground Clauses?*

### ✚ Formulae $\rightarrow$ Clauses ?

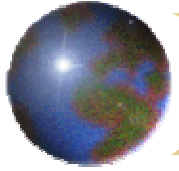
- ✚ Skolem constants introduced.
- ✚ More models generated.

J. Zhang, Extensions to a finite model generator and application to formal methods, Chinese J. of Computers, Feb. 2000 (in Chinese).

### ✚ Clauses $\rightarrow$ Ground Clauses ?

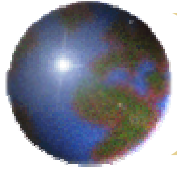
- ✚ More space required. Usually not a serious prob.

### ✚ Representation of ground clauses



## Correctness/Reliability of the search results

- ✚ Checking the result is easy when a model is found.
- ✚ But when the search process terminates without giving a model, shall we believe that the model does not exist?
- ✚ Cross-checking using different model searchers.
- ✚ Verifying the result using theorem provers?



## Non-commutative group of order 3.

### ✚ SEM Input:

3.

$$f(0,x) = x. \quad f(g(x),x) = 0. \quad f(x,f(y,z)) = f(f(x,y),z).$$

$$f(1,2) \neq f(2,1).$$

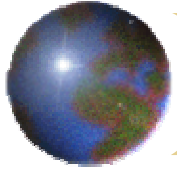
### ✚ Branch 1:

- ✚  $g(0) = 0$

- ✚  $f(1,0) = 0$

- ✚  $f(2,1) = 1$

- ✚  $g(1)$  has no good value.



## OTTER input – for checking the first branch

set(auto).

list(usable).

$x = x.$

$f(E0, x) = x.$

$f(g(x), x) = E0.$

$f(x, f(y, z)) = f(f(x, y), z).$

$f(E1, E2) \neq f(E2, E1).$

$x = E0 \mid x = E1 \mid x = E2.$

$g(E0) = E0.$

$f(E1, E0) = E0.$

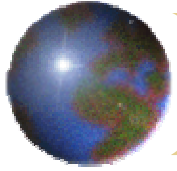
$f(E2, E1) = E1.$

%  $g(E1) = E0.$

%  $g(E1) = E1.$

$g(E1) = E2.$

end\_of\_list.



## *Other Issues*

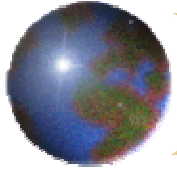
- ✚ Problem formulation: pred, term, lemma
- ✚ Engineering aspects of model searching
- ✚ Infinite models

W. Ahrendt, Deductive search for errors in free data type specifications using model generation, CADE-18, LNCS 2392, 2002.

- ✚ Arithmetic constraints

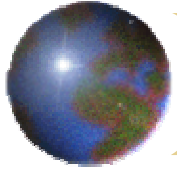
J. Zhang, BoNuS, APAQS 2000; IJSEKE 2001.

- ✚ ...



## *Conclusion*

- ✚ Disproving – an important part of automated reasoning
- ✚ Finite model searching: success/impact
  - ✚ Better if combined with others.  
Mathematical results (domain knowledge) are helpful.
- ✚ Challenges and Research Topics



***THANK YOU !***