ICCGI Panel
Challenges in Handling Information Diversity

Moderator: John Terzakis
Intel USA
September 21, 2010
Valencia, Spain

Panelists

- Janet Kourik (Webster University, USA)
- Kyoko Iwasawa (Takushoku University, Japan)
- Antonio Navarro (Universidad Complutense de Madrid, Spain)
- Pierre Robillard (École Polytechnique de Montréal, Canada)

Topics & Discussion

- Janet presented on the volume of information available on the Internet and posed the question of how we determine what to trust.
- Kyoto presented on a compiler design
- Antonio presented on a problem at his university with accessing diverse information spread across three databases.
- Pierre presented the results of a study on how information is communicated (by socialization, by coordination, by cooperation and by collaboration—the highest percentage)

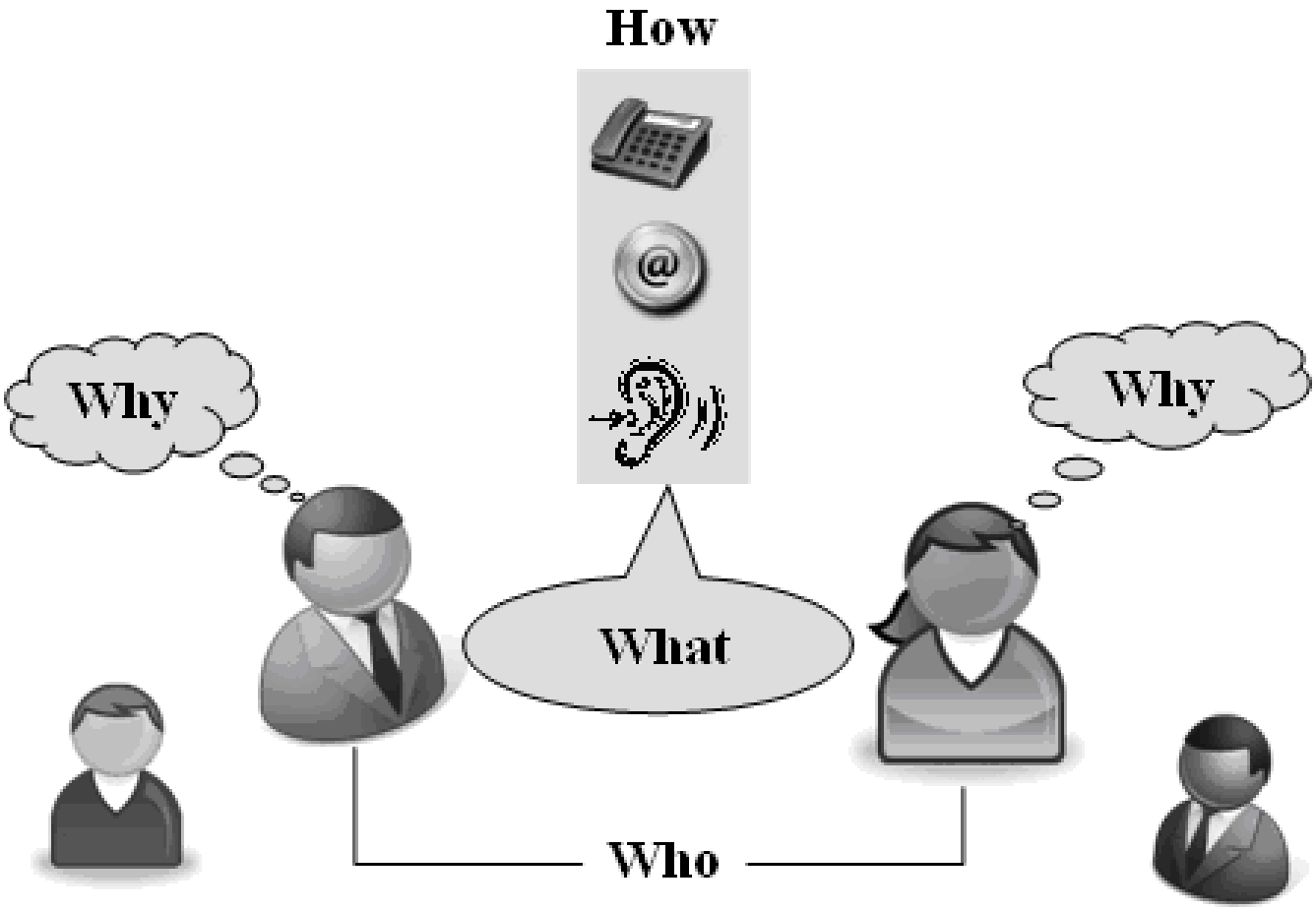
The Fifth International Multi-Conference on Computing in the Global Information Technology
ICCGI 2010

Challenges in Handling Information Diversity

Pierre-N. Robillard
Département de génie informatique et de génie logiciel
École Polytechnique de Montréal
Montréal, Qc. Canada

Surname.Name@polymtl.ca

Ad hoc verbal information
How useful is it?



Interactions Modes for Face-to-Face (FtF) Communications

F0F Instant

12%

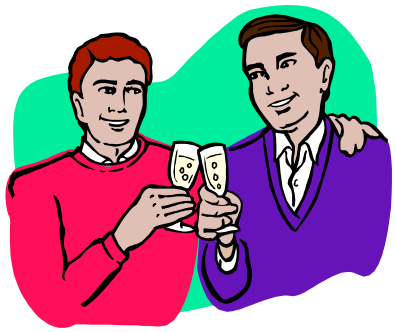
F2F Dyadic

74%

FnF Polyadic

14%

Information Diversity



Socialization

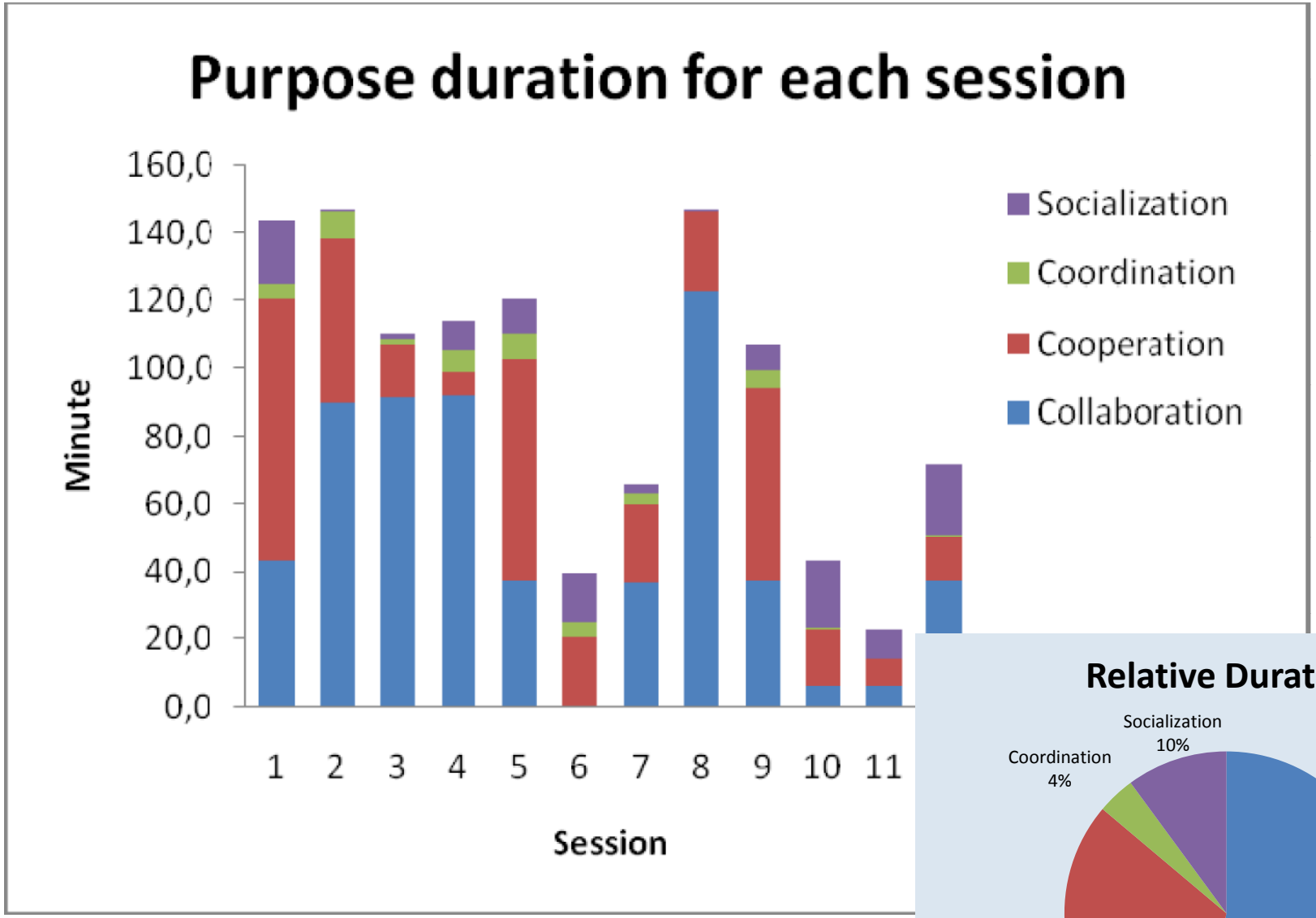
Cooperation

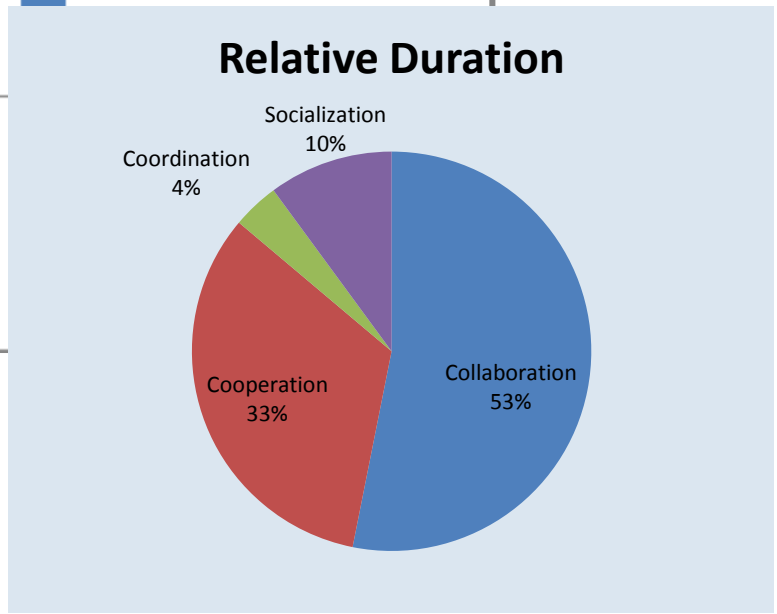
Coordination

Collaboration

Challenging Questions

- **Do we need to go global for all type of communications?**
- **How can we select the appropriate type of communications?**


Purpose duration for each session
160,0
140,0
120,0
100,0
80,0
60,0
40,0
20,0
0,0
Minute
1
2
3
4
5
6
7
8
9
10
11
Session
Socialization
Coordination
Cooperation
Collaboration


Relative Duration
Socialization
10%
Coordination
4%
Cooperation
33%
Collaboration
53%

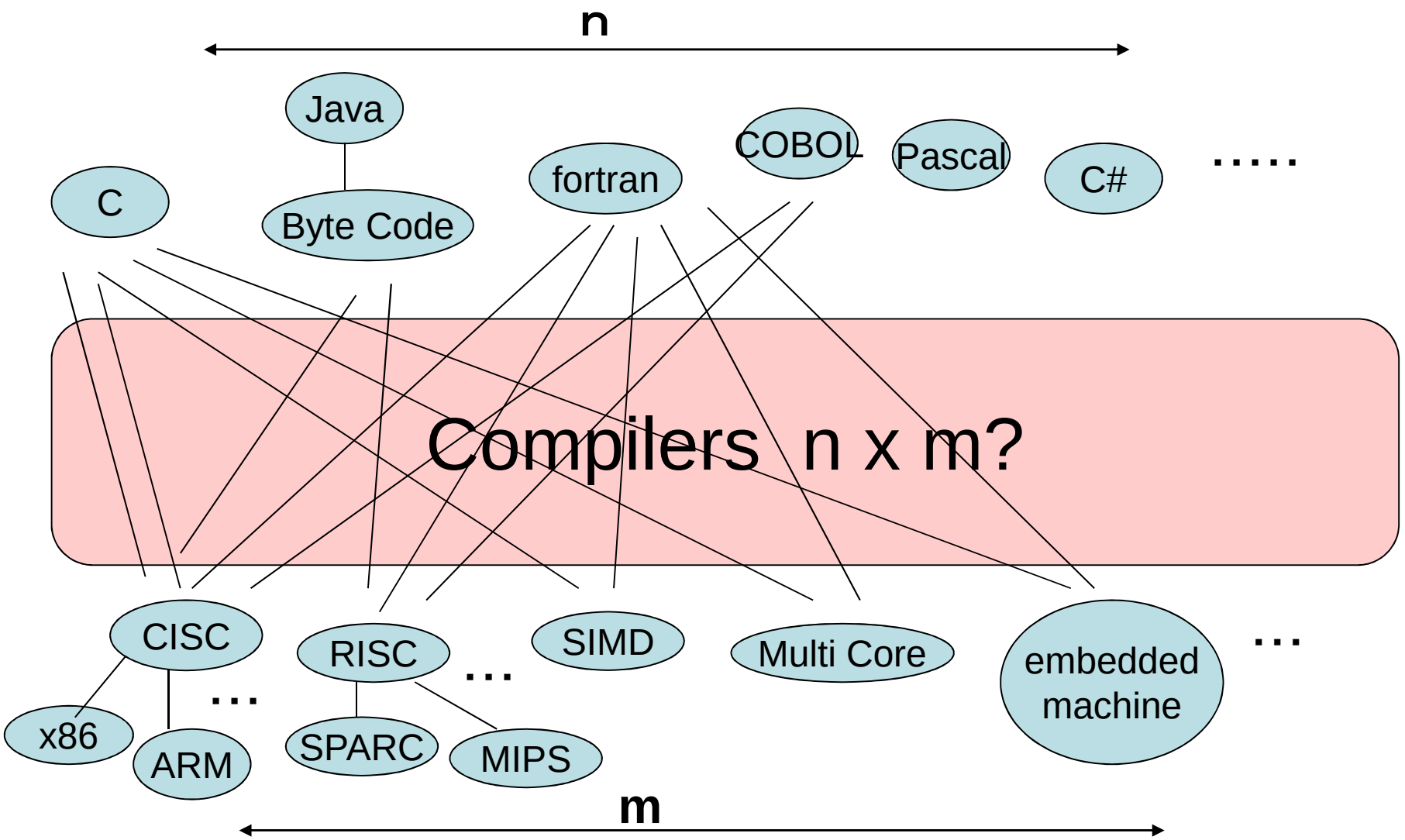
Diversity on compiling Compiler infrastructure

Kyoko Iwasawa

Takushoku University

Tokyo Japan

Diversity on compiler



Diversity of optimization

- Dataflow analysis VS program conversion
  - Loop conversion, code motion etc.
- Data flow of Array elements VS address expression optimization
- Instruction reorder VS register allocation
- Inter-procedural analysis

- Machine independent VS machine dependent

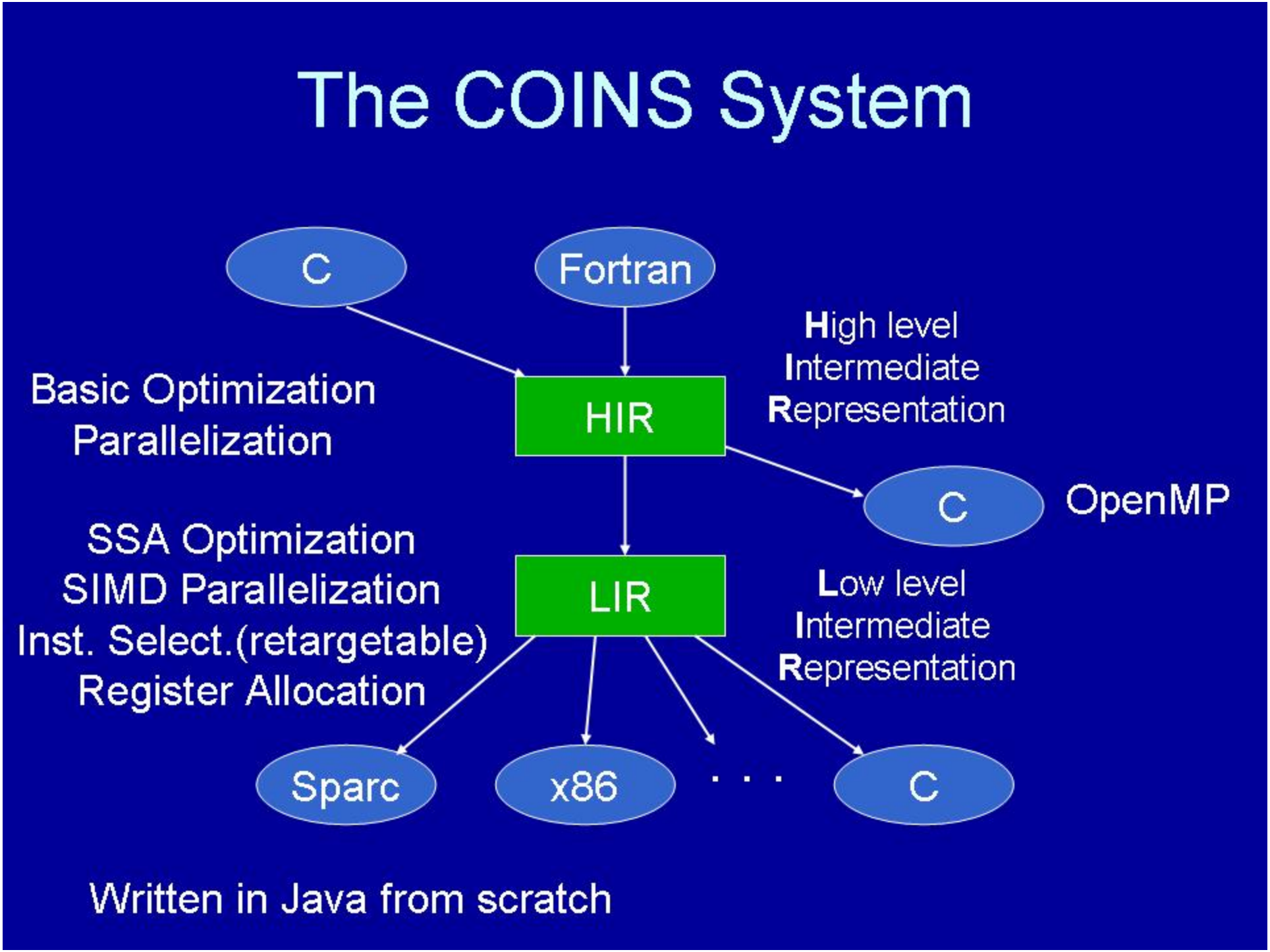
Other problem

- Compiler development is difficult and complex, however it seems that there is a few novel and interesting technique …(really?)
  - it looks traditional and conservative
- Young people tend to hesitate compiler's hard work (at least in Japan).
- Always new system needs their own compiler (especially embedded system)

COINS project (COmpiler INfraStructure)

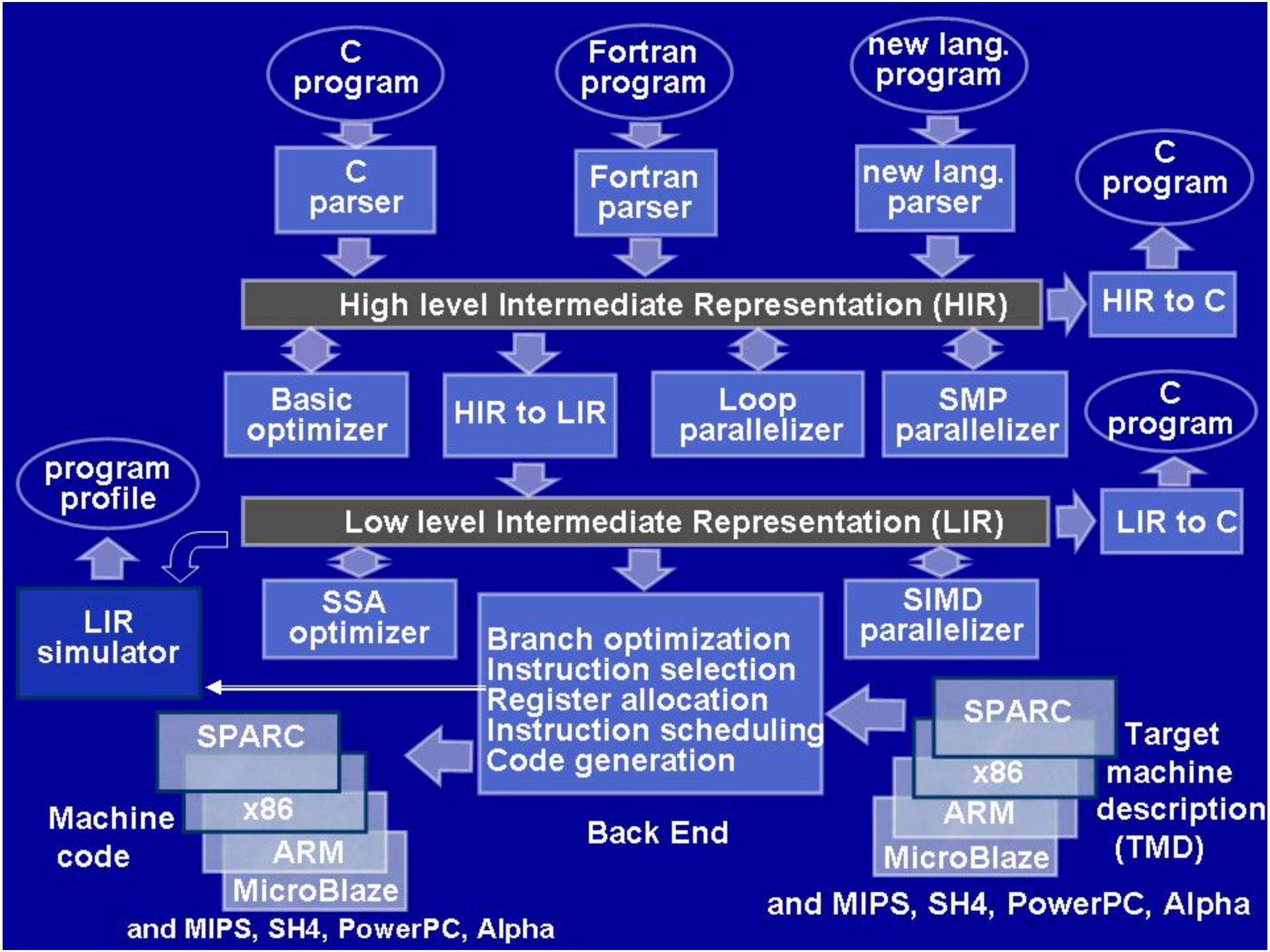
- Every features of compiler are modularized (written by Java)
- Restructure each module (for optimization and parallelization)
- Common intermediate representation
  - Two levels
- Parallelizing features
- Retargetable

The COINS System



COINS's features

- HIR (High level Intermediate Representation)
- LIR (Low level Intermediate Representation)
- Parsers (source program --> HIR)
  - C, Fortran, Java(planed)
- Optimizers for HIR/LIR
  - data flow based (HIR/LIR)
  - SSA based (LIR)
- Parallelizers for HIR/LIR
  - HIR --> OpenMP
  - SIMD parallelization (LIR)
- Code generators (LIR --> machine code)
  - retargetable code generator
  - Sparc, Intel x86


C program
Fortran program
new lang. program
C parser
Fortran parser
new lang. parser
C program
High level Intermediate Representation (HIR)
HIR to C
Basic optimizer
HIR to LIR
Loop parallelizer
SMP parallelizer
C program
program profile
Low level Intermediate Representation (LIR)
LIR to C
LIR simulator
SSA optimizer
Branch optimization
Instruction selection
Register allocation
Instruction scheduling
Code generation
SIMD parallelizer
SPARC
x86
ARM
MicroBlaze
Machine code
and MIPS, SH4, PowerPC, Alpha
Back End
SPARC
x86
ARM
MicroBlaze
Target machine description (TMD)
and MIPS, SH4, PowerPC, Alpha

Machine Description

Register Definition *%i0 - %i5, %o0 - %o5, %l0 - %l7*

```
(def *reg-I32* ( (foreach @io (i o)
                    (foreach @n (0 1 2 3 4 5) (REG I32 "%@io@n")))
          int 32bits  (foreach @n (0 1 2 3 4 5 6 7) (REG I32 "%l@n")) ))
```

Instruction Description *LIR Sparc*

```
(foreach (@op @code) ((ADD add) (SUB sub)
                      (BAND and) (BOR or) (BXOR xor))
  (defcode @code (SET I32 reg (@op I32 reg rc))
  (asm `(@code, $1, $2, $0))     rc: reg or const
  (cost 1)))        cost of this instruction
 (foreach (@n @l) ((2 1) (4 2) (8 3) (16 4) (32 5)) ;; mult by shift
  (defcode mul-sll@l (SET I32 reg (MUL I32 reg (INTCONST I32 @n)))
  (asm `(sll ,$1 (con @l) ,$0))     con @l = 1, 2, 3, 4, or 5
          (cost 1)))
```

Example of Code Generation

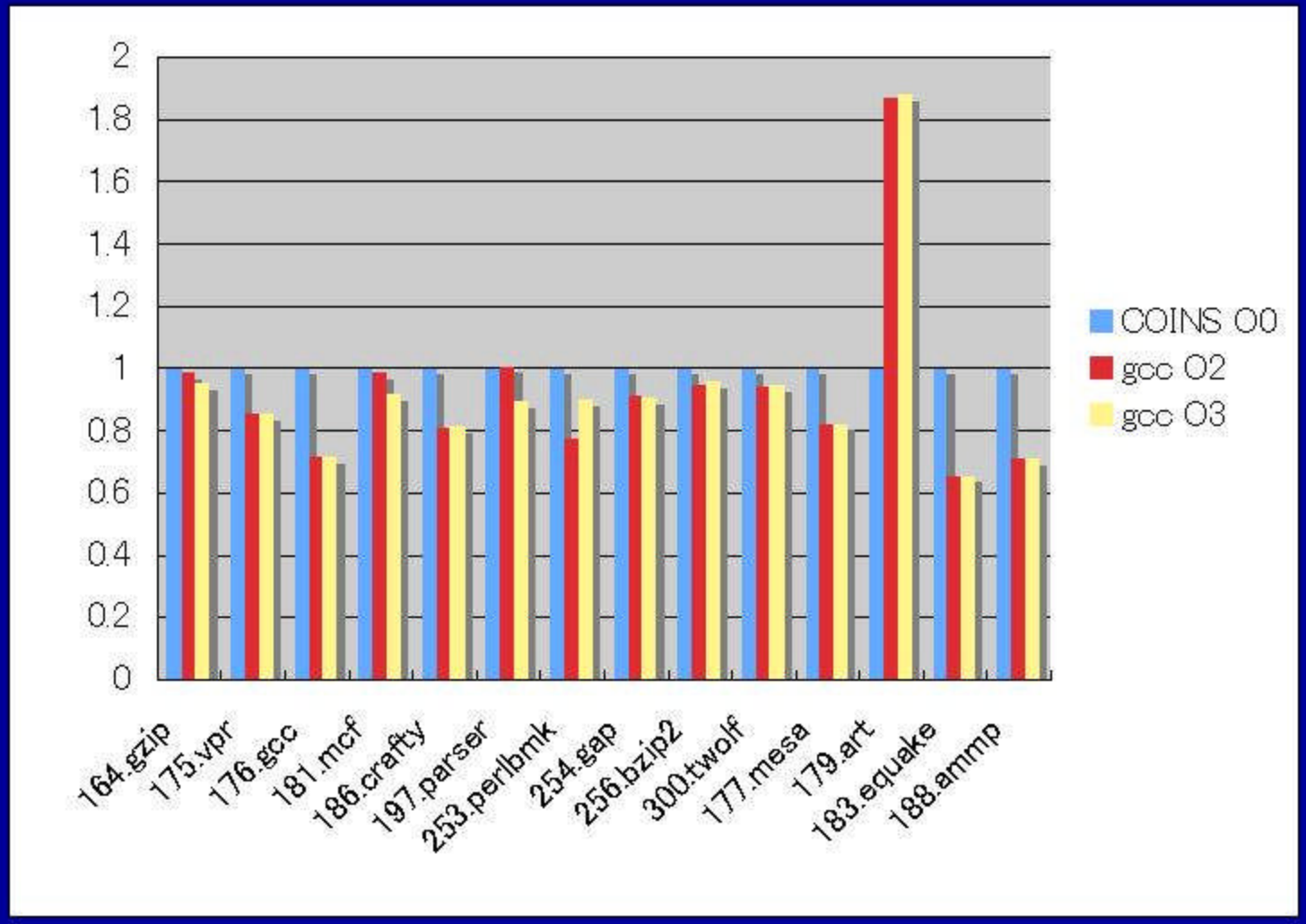
```
(foreach (@n @l) ((2 1) (4 2) (8 3) (16 4) (32 5))
  (defcode mul-sll@l (SET I32 reg (MUL I32 reg (INTCONST I32 @n)))
  (asm `(sll ,$1 (con @l) ,$0))
            (cost 1)))
```

LIR: (SET:I32 %I2 (MUL:I32 %I3 (INTCONST:I32 4)))

Examples of retargeting

Machine	Coded lines	Months	Note
SPARC	1952	-	not available
x86	2533	-	not available
MIPS	2207	3	nonexperienced student
SH4	3596	6	nonexperienced student
ARM	3052	6	nonexperienced
ARM-Thumb	1980	3	nonexperienced
MicroBlaze	1383	2	experienced
Power PC	5018	6	nonexperienced student
Alpha	1216	2	nonexperienced student

Execution time ratio (SPEC2000, x86)



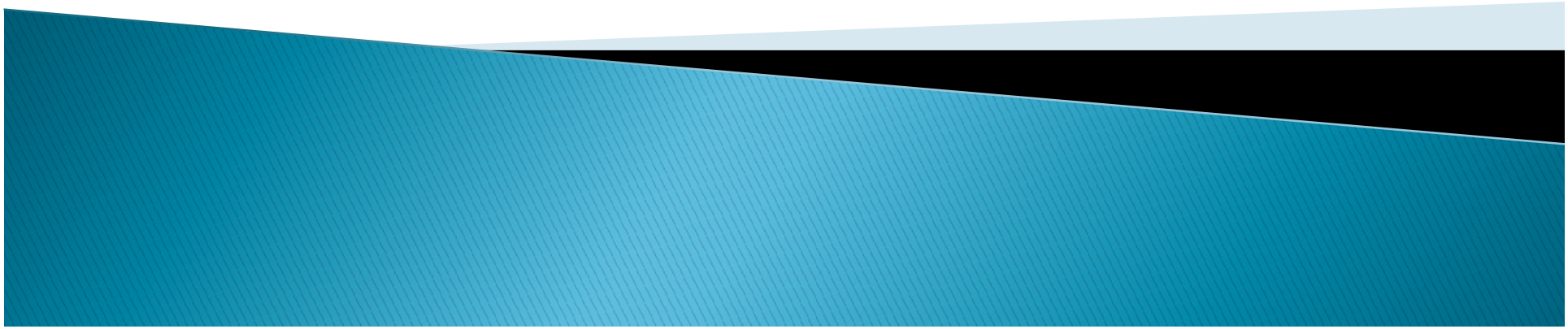
On Epson
Pro-1000
Pentium 4
1.8 GHz
256 MB

Execution time ratio compared to COINS -O0
(no optimization option)

Please see
www.coins-project.org →[English Top]
(www.coins-project.org/international/index.htm)

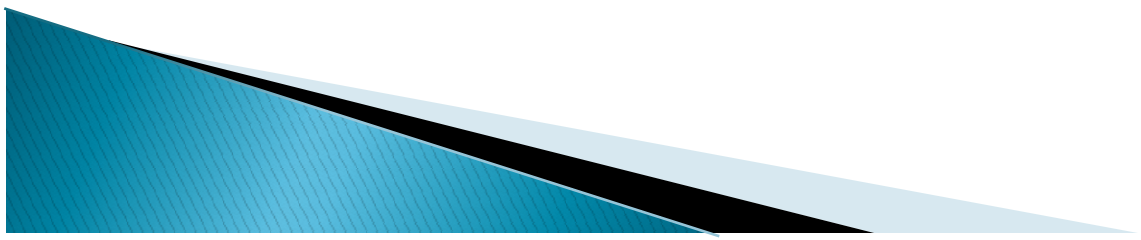
Challenges in Handling Information Diversity

Janet L. Kourik, Ph.D.
Webster University, St. Louis, Missouri



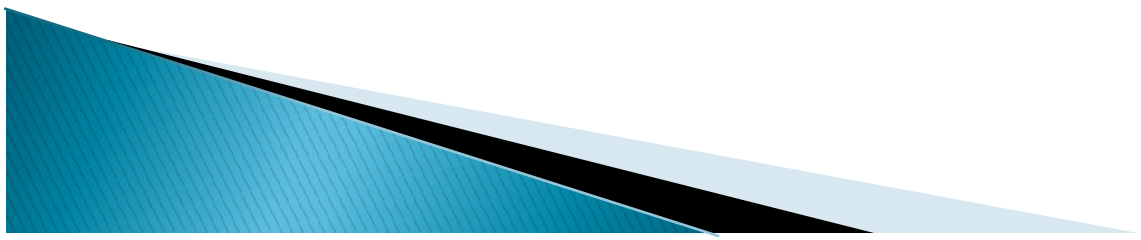
Growth of Data

- Data
  - Data growing rapidly
  - Petabytes and exabytes
- Challenge to find meaning
- What tools do we use?



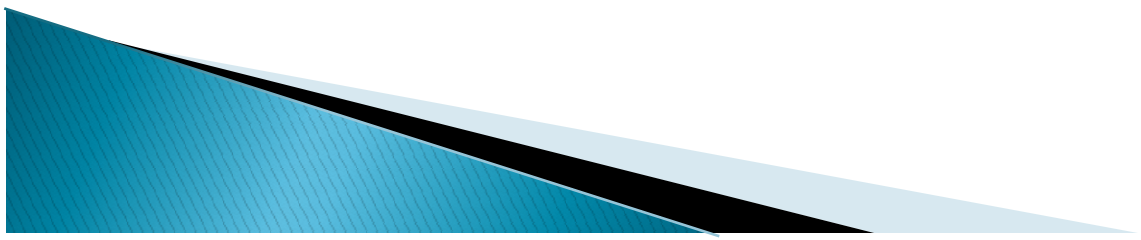
Unstructured Text

- Difficult to find meaning in text
- Techniques emerged – 15 years
- Storage formats, metadata, complex data types., etc.



Questions

- How determine quality of sources/input?
- What tools or concepts can we bring to the task?
- How can we help people make good judgments about the information?



Information Diversity in the UCM Virtual Campus

Antonio Navarro

Dpto . Ingeniería del Software e Inteligencia Artificial

Universidad Complutense de Madrid (UCM)

anavarro@fdi.ucm.es

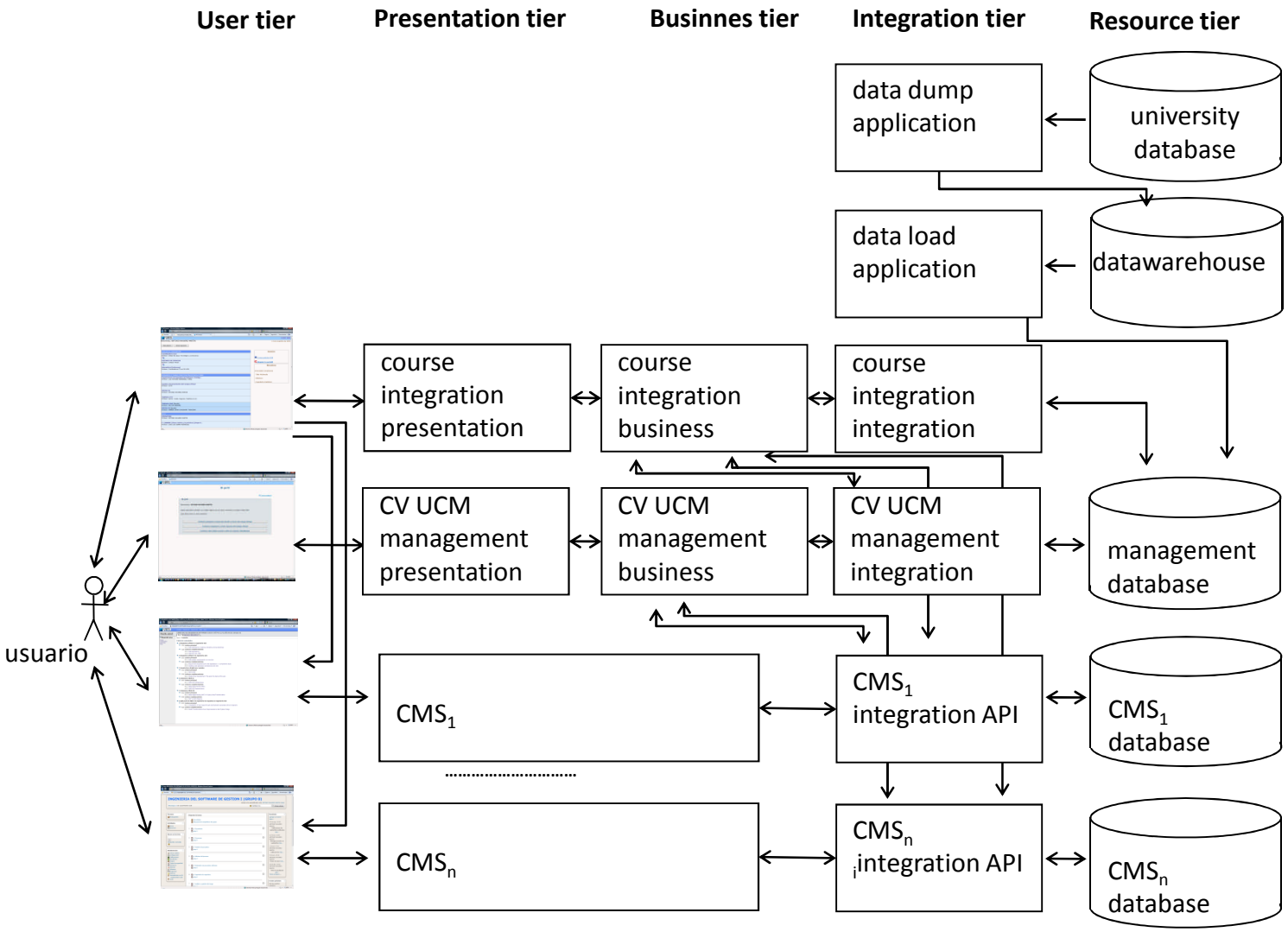
Index

- UCM Virtual Campus
- Information diversity
- Information diversity in the UCM Virtual Campus
- Conclusions

UCM Virtual Campus

- UCM Virtual Campus is a large virtual campus:
  - More than 44,000 students registered (51%)
  - More than 3,500 lecturers registered (56%)
- Three Course Management Systems (CMSs) are available and integrated:
  - WebCT 4.1
  - Moodle 1.9.2
  - Sakai 2.4.0

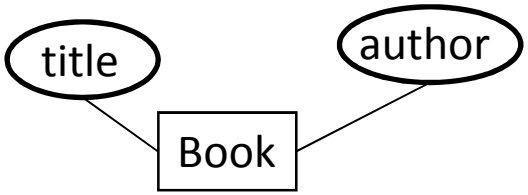
UCM Virtual Campus



UCM CV software archnitecture

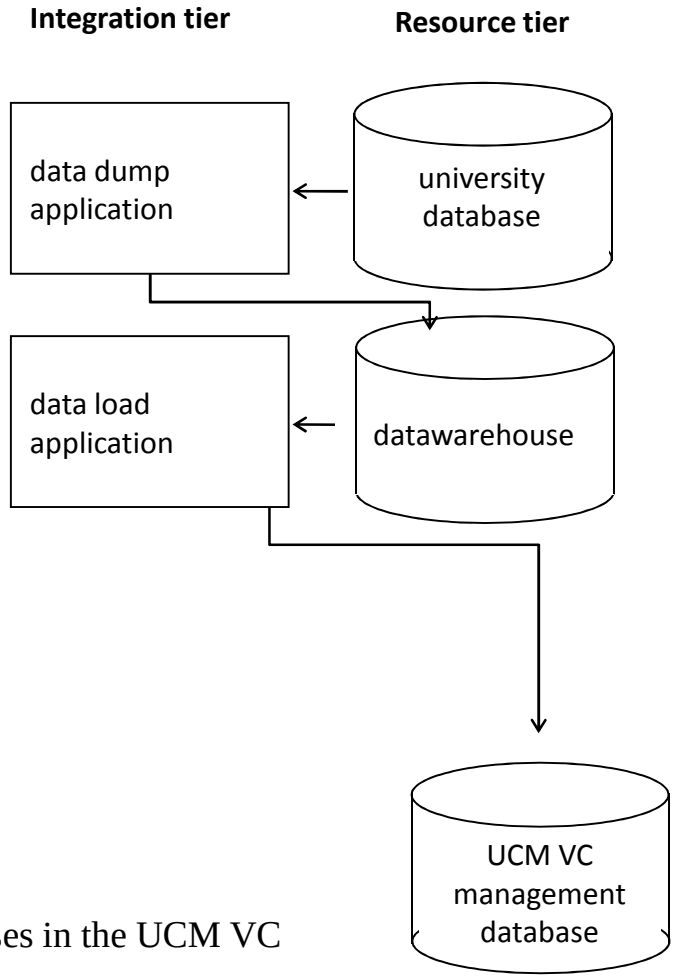
Information diversity

- At least, there are two types of information diversity:
  - Structural (abstract syntax)
    - eg: <!ELEMENT book (title, author, description)>
      <!ELEMENT book (title, author)>
  - Syntactical (concrete syntax)
    - eg: <!ELEMENT book (title, author)>



Information diversity in UCM VC

- Structural diversity



Three of the databases in the UCM VC

Conclusions

- Information diversity mean problems
  - Syntactical diversity mean *medium* problems
  - Structural diversity mean *big* problems
- Structural diversity in the UCM VC is one of the biggest problems
- CMS integration is the other big problem

Information Diversity in the UCM Virtual Campus

Antonio Navarro

Dpto . Ingeniería del Software e Inteligencia Artificial

Universidad Complutense de Madrid

anavarro@fdi.ucm.es