



Artificial Intelligence in Industry: The good, the bad and the ugly

Gregg Vesonder
AT&T Labs - Research
Shannon Labs
gtv@research.att.com



Roadmap

- About me
- Early History
- "Golden Age of AI"
- More Advanced AI
- OO and .com eras
- Current Work
 - Mise en place
 - Preservation
 - Case Studies - Some ESs, KDD Paper
 - Data Mining Gastronomique



Vesonder's Relevant Bio

- Software for 30+ years
- PhD in Cognitive Psychology – Computer modeling of learning and memory
- [Bell|AT&T] labs for 25 years
- Dozens of projects, many AI but not all
- OO and .com (Online Platforms)
- Stevens: Software Engineering, Software Architecture and Design



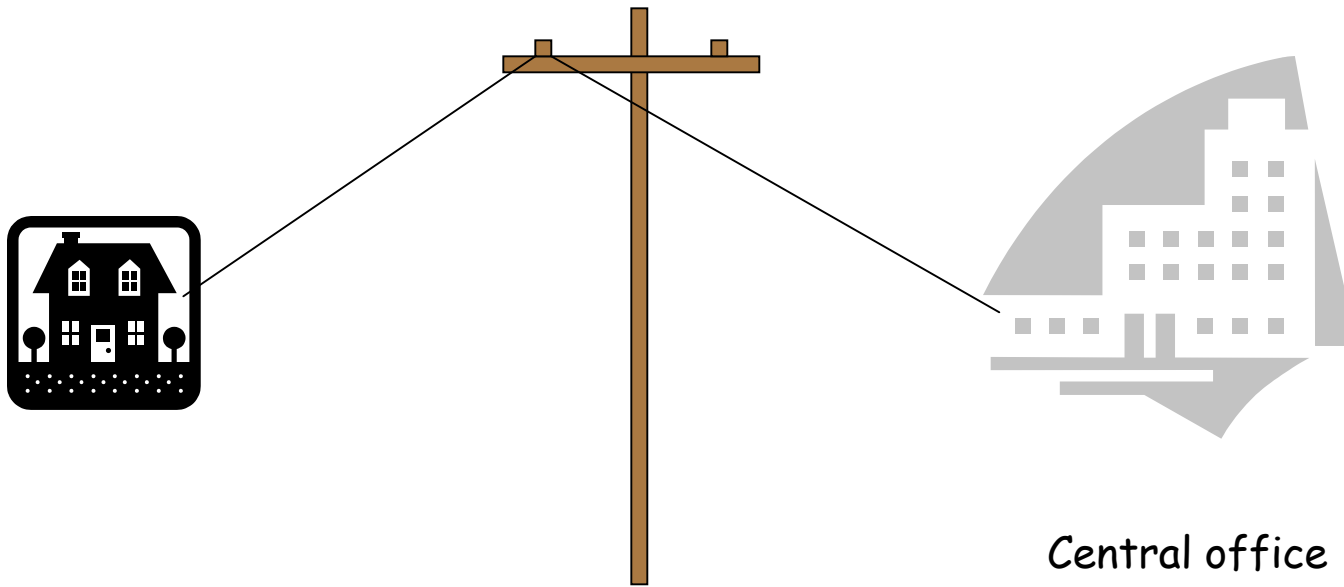
My First AI Project

(in industry)

- Serendipity - in the field with another system (CRAS)
- Noticed analysts struggling with reports -- doing correlations
- Met after work
- Generated a proposal -- this was in 1979
 - DEC had R1 - rule based system for configuring vaxes
 - Schlumberger had the Dipmeter Advisor
 - Stanford had MYCIN



ACE





Knowledge Engineering Technique

- Select a specific topic
- Gather “book knowledge” of topic and construct rules
- Incorporate rules in a computer program
- Run program on real situations
- Discover where program differs from what the expert suggests
- Repeat last three steps until program agrees with expert (satisficing)



ACE Rule

IF a range of pairs within a cable has generated a large number of customer trouble reports

ANDIF a majority of work on these pairs was done in the terminal block

THEN look for a common address for these pairs



Some other rules

MYCIN:

IF the infection type is primary-bacterima

ANDIF the suspected entry point is the gastrointestinal tract

ANDIF the site of the culture is one of the sterile sites

THEN there is evidence that the organism is bacteriodes

DIPMETER ADVISOR:

IF there is a red pattern over a fault

ANDIF the direction is perpendicular to the fault

ANDIF the length is > 200 feet

THEN the fault is a growth fault



Lessons Learned

- Autocratic vs. Democratic
- Flavors of expertise: medical diagnosis versus cable analysis
- Comfortable user interface
- Software interface:
 - UNIX + AI
 - ~30% of work is AI related
 - Quick code development
- Rules and procedures
- Skilled Knowledge engineers - know the domain
- Added value, part of the work flow
- Personal - role of luck and the thrill of seeing your efforts used.



Digression

- A2b music: November 1997
 - Really seeing folks use stuff
- iTunes today
- Back to AI





Led to Major Initiatives In AI: The “golden” age

- ALVEY - Great Britain
- ESPIRIT - European Common Market
- FIFTH Generation - JAPAN
- Strategic Computing - United States
 - Autonomous land vehicle
 - Pilot's associate
 - Fleet manager
- And most corporations



More Expert Systems

- OKIES
- MIDS
- SARTS AUTOTEST
- Other AI LODE – kernel knowledge bases from databases
- Led to Environment tuned to our needs – C5 and ES technology embedded in larger systems



In Telecom

- Maintenance
- Provisioning (forecasting, planning & designing)
- Network Administration (routing, billing, facility assignment)



Expert Systems in Telecom

	'83	'84	'85	'86	'87	'88	'89
Maintenance	1	1	2	5	3	9	4
Provisioning	-	-	-	1	1	3	1
Network Administration	-	-	2	1	-	3	1
TOTAL	1	1	4	7	4	15	6



More Advanced AI: PROSE

- Knowledge based configurator for switch adjuncts
- Used C-Classic – pure declarative form, provided quick and consistent knowledge editing
 - Forces consistency by:
 - Straightforward representation of product's compatibility rules
 - Enforcement of consistency and integrity of the compatibility rules and the configurations they generate
 - Generates configurations from minimal number of inputs



C-Classic Inference

- Classification - find all descriptions applicable to an object: find all descriptions more general and specific than a proposed object description
- Completion or propagation of logical consequences, including, but not limited to inheritance
- Contradiction detection
- Simple forward chaining rules or triggers
- Dependency maintenance for retraction and recovery



AI Winter - OO and .com

- Many of the initiatives did not meet expectations - drop in attendance
- Most went to OO (cfront) and then on to web (a2b, cospace-community in general)
- Signs of spring



Web- Cospace





Newest Stuff: AI and Datamining

- Systems Approach (work with Wright and Dasu)
- Unique issues and combinations of issues
 - Mise en place
 - [most|all] runs are unique
 - Data Quality is crucial
 - Granularity
 - Downstream systems
- Process issues
 - Knowledge engineering throughout
 - Verification and validation issues



Bouillabaisse Data Mining

- Data exists in some repository/corpus
- Know the fields and relationships
- At least familiar with some domain
- Others have mined the data - community
- Reference efforts -- helps Verification (built system right) and Validation (built right system)
- ...
- World Wide Telescope - Jim Gray



Stone Soup Data Mining

- A Fable in many parts
- The data is not in one place, in fact it is in many places
 - Don't know the quality
 - Don't know what it means and there is no one source to discover it (multiple, conflicting experts - Brooks "never go to sea with two chronometers, go with one or three")
- Data does not remain there - have to capture it
 - usually on arcane systems



Stone Soup -2

- Once you get it - more experts, pilot runs (very much like Knowledge Engineering technique)
 - BTW it is in EBCDIC, described by COBOL copybooks, you're running UNIX...
- Discover you need other data to interpret it - back to previous page
- At this point it has been months - if lucky
- Time to formalize the collection process
- Did I mention the data is huge!
- Time to do some "data mining" - knowledge and quality
- Archiving issues - reproduction (depends on what is available and who contributes)



Knowledge Engineering Technique

- (So old that it needs to be reprised)
- Knowledge Engineer becomes familiar with domain, architecture and operation
- KE meets with experts to understand operations and issues
- Team uses knowledge to create first (and subsequent) passes at working system
- Experts critique results, provide new knowledge and iterate on previous step until a satisfactory (or best possible) conclusion is achieved

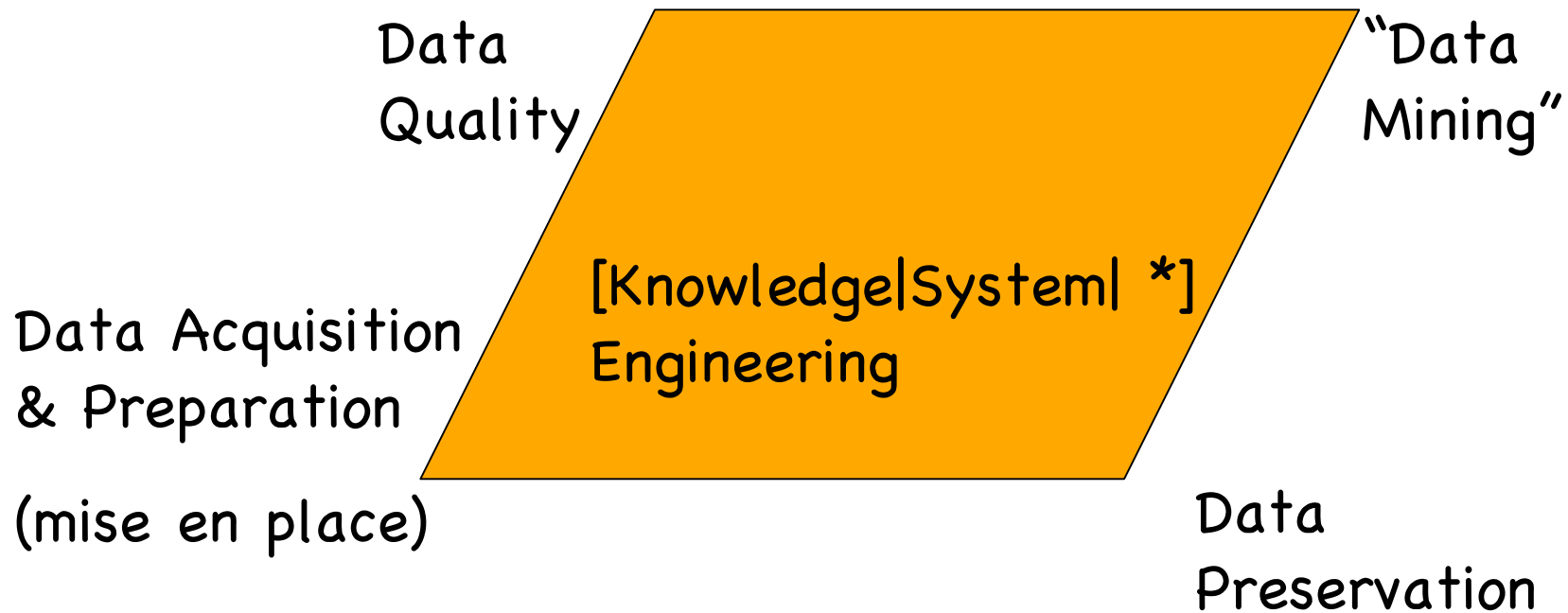


Stone Soup-3

- About this time one of your feeds changes - actually it was several months ago
- Verification and validation throughout
- Preservation of data, summarized data, interim reports and techniques - really time "capsules"



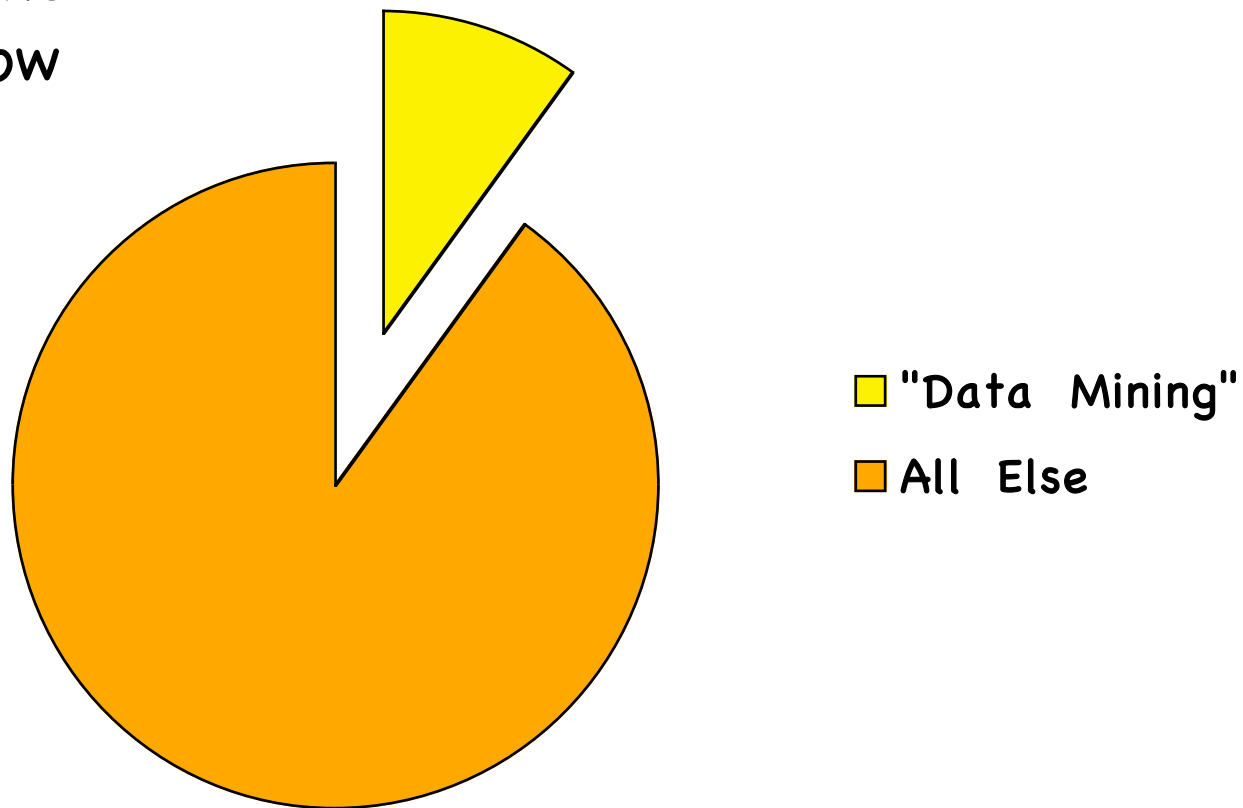
A View of the Space





A Rough Estimate of the Effort

Of course the 10% can grow over time, but...





The Life Cycle

- Discover data needed - KE
- Get data/Establish Feed
 - Discover and perhaps get additional data to interpret data - KE
 - Verify & Validate feed
 - Assess data quality
- Discover Reference results for V & V (may be earlier)
- Prepare environment and Run Data
- V &V - KE (iterate - may take you to top again)
- Preserve environment and archive
- Continuously check "upstream" issues - improve data quality
- Usually there is increased level of understanding



Knowledge Engineering (KE)

- Book Knowledge on topic sparse
- Parni on calls for months - patience to find knowledge nuggets
 - Finding appropriate expert but:
 - Current project ~50% of time on calls with Subject Matter Experts
 - Experts Disagree - more conference calls
- Initial run - bridge knowledge gap other way
- Prep/Run time measured in large units



Preservation

- No ready made archives
- Preserve data, software and comparisons
 - Data and meta data synchronized (e.g. time dependent)
 - Redundancy, security, ..
 - Recoverability



The Data Attributes

(APOLOGIES - COULD NOT FIND PREDEFINED TAXONOMY)

- Single vs multiple streams
- Self contained -several ways
- Temporally based - several ways
- Accessible repository
- Reference implementation - testing, V&V
- Size
- Complexity
- (a work in progress, more to come)



Mise en place

- “put in place” chopping, mincing, measurement, peeling, washing
- Significant planning activity to start a run
 - Data ready - off tape and accessible - could be N different feeds
 - Data verified
 - Sufficient system resources (disk, memory, ...)
 - Consistent software builds
- Candidate for AI planning techniques, ES for monitoring run (insuring available disk resources, trapping failures, ...)



KDD Paper Example

- Case study from KDD
- AI techniques addressing quality issues of the data
- Instance of our general methodology that can be used at every stage of the lifecycle
 - Knowledge Engineering based
- Spent a lifetime in multi hour conference calls



Data Quality

Dasu, Vesonder, Wright

- Common for operations databases to have 60–90% bad data
- Audits are used to detect errors for later correction
- Enlightened approach is to proactively prevent errors before they occur BUT the business operations rules for these databases are inaccurate and incomplete and acquiring it has challenges.
- The solution we presented was using Knowledge Engineering and Rule Based programming to capture and represent the data.



Typical Project Characteristics

- Knowledge is available in a fragmentary way, often out of logical or operational sequence
- Expertise is split across organizations – little incentive to cooperate
- Business rules change frequently
- Experts do not agree – inconsistent rules
- Project personnel change frequently
- Little project accountability in matrixed organizations



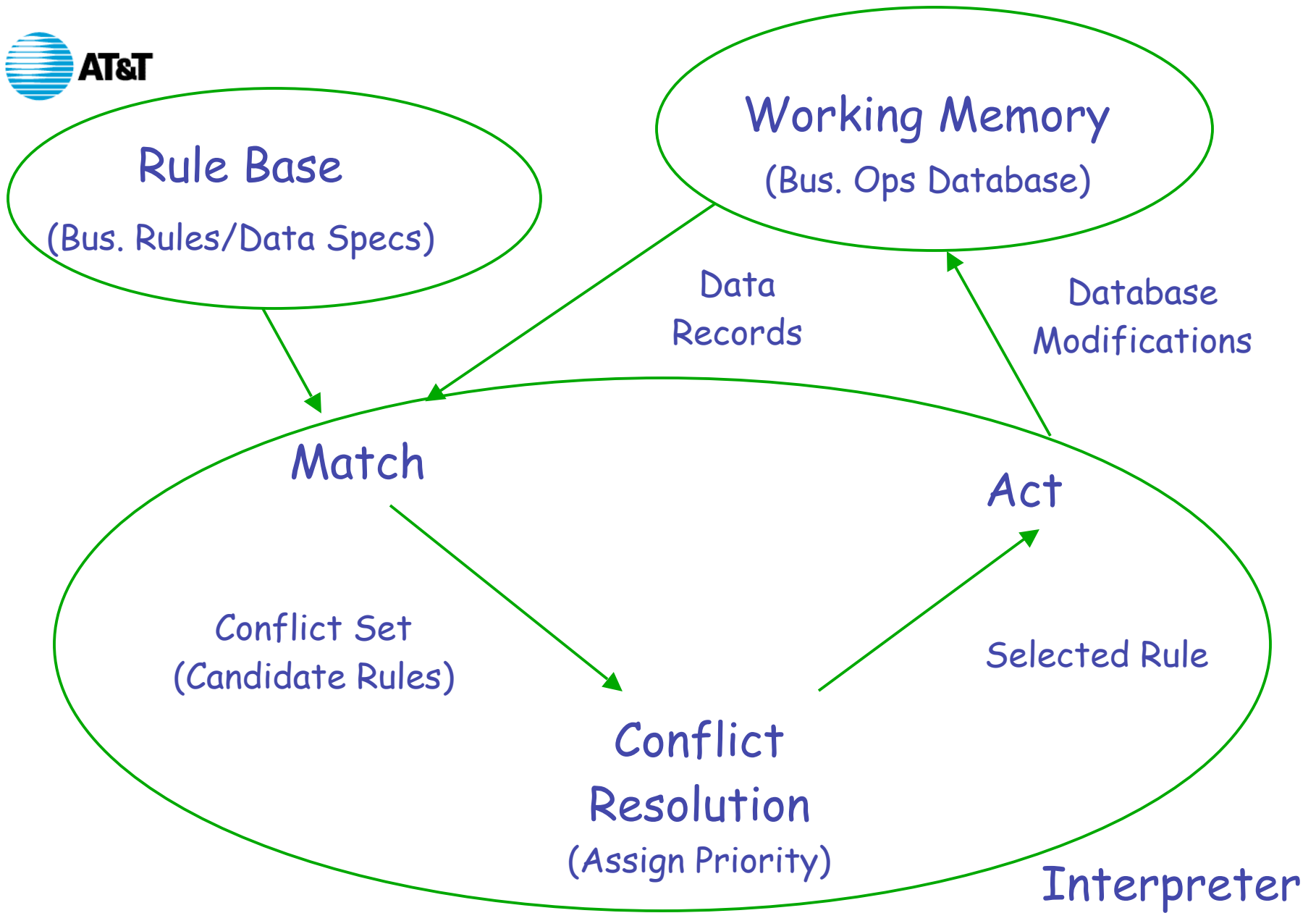
Knowledge Engineering

- Knowledge Engineer becomes familiar with domain, architecture and operation
- KE meets with experts to understand operations and issues
- Team uses knowledge to create first (and subsequent) passes at rules
- Experts critique results, provide new knowledge and iterate on previous step until a satisfactory (or best possible) conclusion is achieved



Quality Case Study

- 20 experts - a challenge
- Original in SAS
- Rule conversion focused knowledge in meaningful, manipulatable chunks
- Data quality engineer of present and future will need techniques to capture, vet and deploy knowledge of the data, process and necessary continuous audits and do this at scale.





Mise en place and Planning

- Planning algorithms, means-ends analysis to do cutting and chopping
 - Check for and Secure resources
 - Assemble data
 - Schedule jobs
 - Monitor run
 - Assemble output -- distributed computing
 - Flag results



Data Mining Gastronomique

- Data Quality - see Parni & Ted book reference
- AI Techniques:
 - Planning - especially for Mise en place
 - Expert Systems - Rule base/Agent systems for monitoring/quality
 - Also use Ganglia and other tools
 - KE at most points



Datamining Conclusions

- Provide a broader view of what constitutes data mining
- Process orientation - addresses complete system development
 - Sometimes the data isn't on the web, in a corpus or on a CD
 - Quality issues
- Mise en place a big issue, since each run is special
- AI as one approach to the issues
- Much more coming



General Conclusions

- The Good: new technology, high productivity, satisfied users, sense of accomplishment, span of effect, resources as a function of success. Not a paper chase, chance to work on significant problems
- The Bad: many jump on bandwagon, signal to noise ratio is high, many misconceptions of AI depending on experience, management buy in
- And The Ugly - the real world is, it is not all AI, really need to have a systems approach and there is beauty in the ugly



Shameless Plug

- Check out my blog, the Software Universe,
<http://vesonder.typepad.com/universe/>
- Or go through my very uncreative homepage:
<http://homepage.mac.com/vesonder>



References

- Tamraparni Dasu, Gregg Vesonder, Jon R. Wright: Data quality through knowledge engineering. KDD-2003: 705-710
- Jon R. Wright, Elia Weixelbaum, Gregg T. Vesonder, Karen Brown, Stephen Palmer, Jay Berman, Harry Moore: A Knowledge-Based Configurator that Supports Sales, Engineering, and Manufacturing at AT&T Network Systems. AI Magazine 14(3): 69-80 (1993)
- Tamraparni Dasu, Theodore Johnson: Exploratory Data Mining and Data Cleaning. John Wiley 2003.