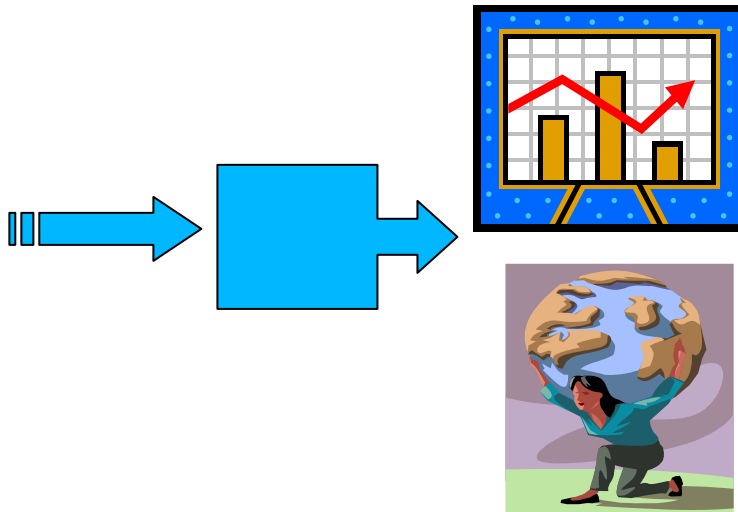


# Toward an Integrated Evaluation Framework

*Bernardo Magnini  
FBK-irst, Trento, Italy*

FlareNet launching event, February 12, 2009

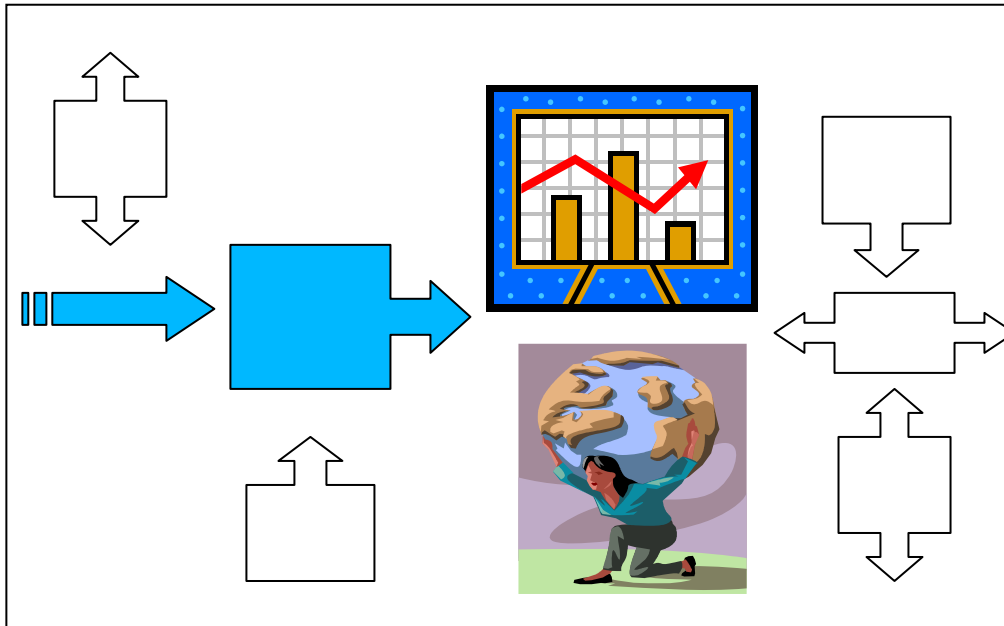


- Word sense disambiguation
- Named entities recognition
- Semantic role labeling
- .....

## PROS

- Tasks are well identified
- Task specific benchmarks have been created
- New tasks have been proposed (e.g. Semeval)

# Task-Oriented Evaluation



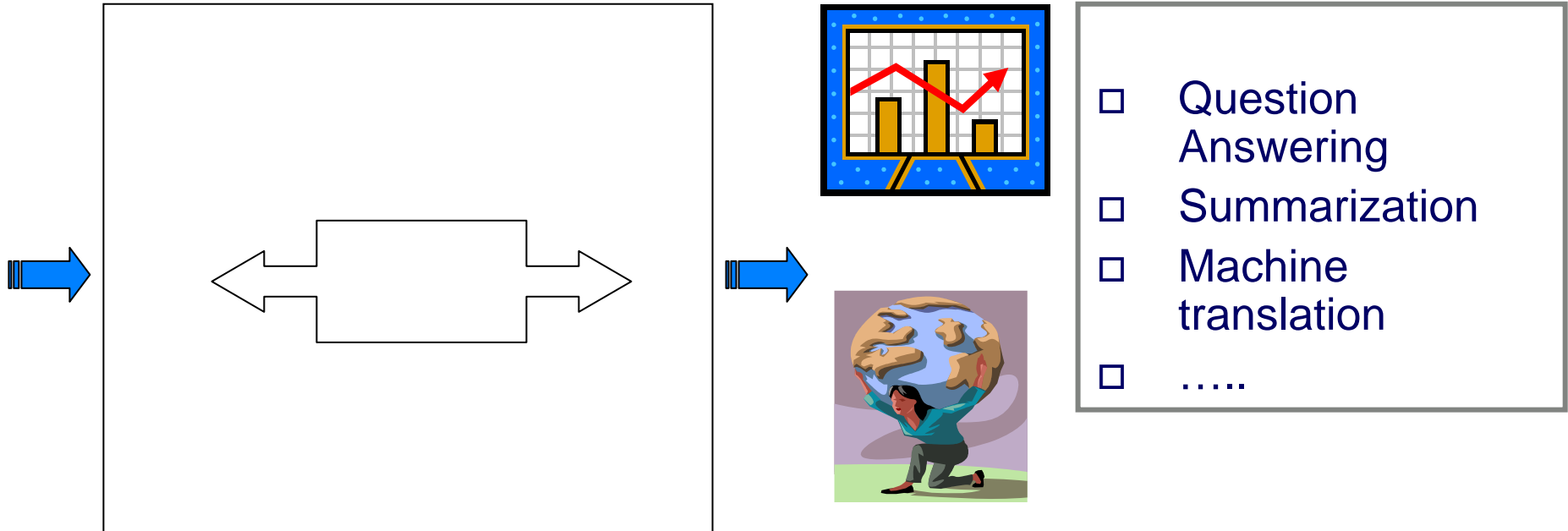
- Word sense disambiguation
- Named entities recognition
- Semantic role labeling
- .....

## PROS

- Tasks are well identified
- Task specific benchmarks have been created
- New tasks have been proposed (e.g. Semeval)

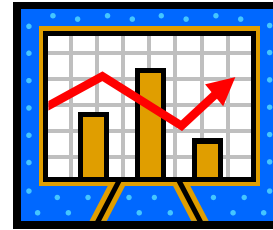
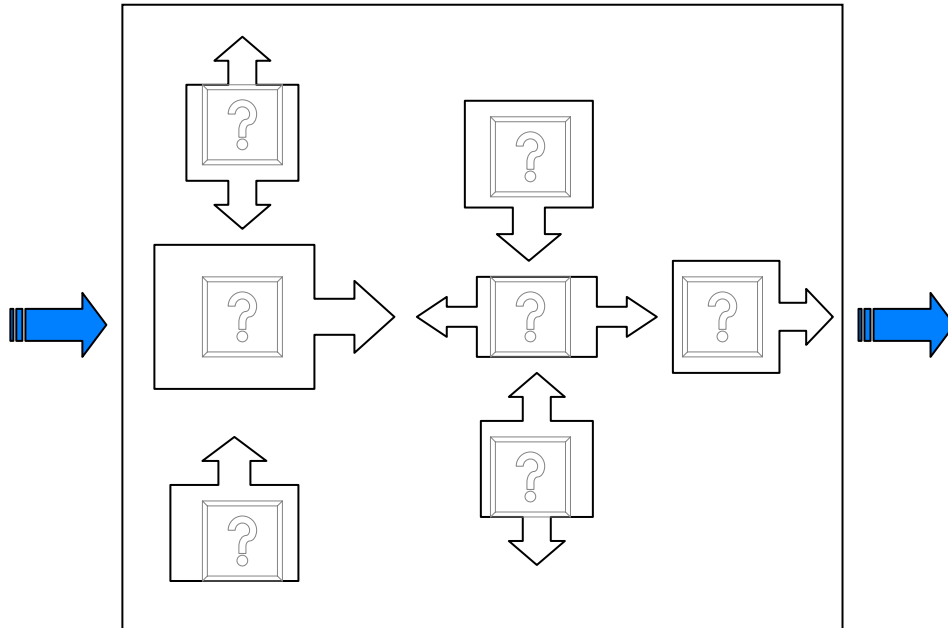
## CONS

- Difficult to evaluate task contribution in applications
- Tasks may appear as “artificial”



## PROS

- Focus on real scenarios
- Users can be involved



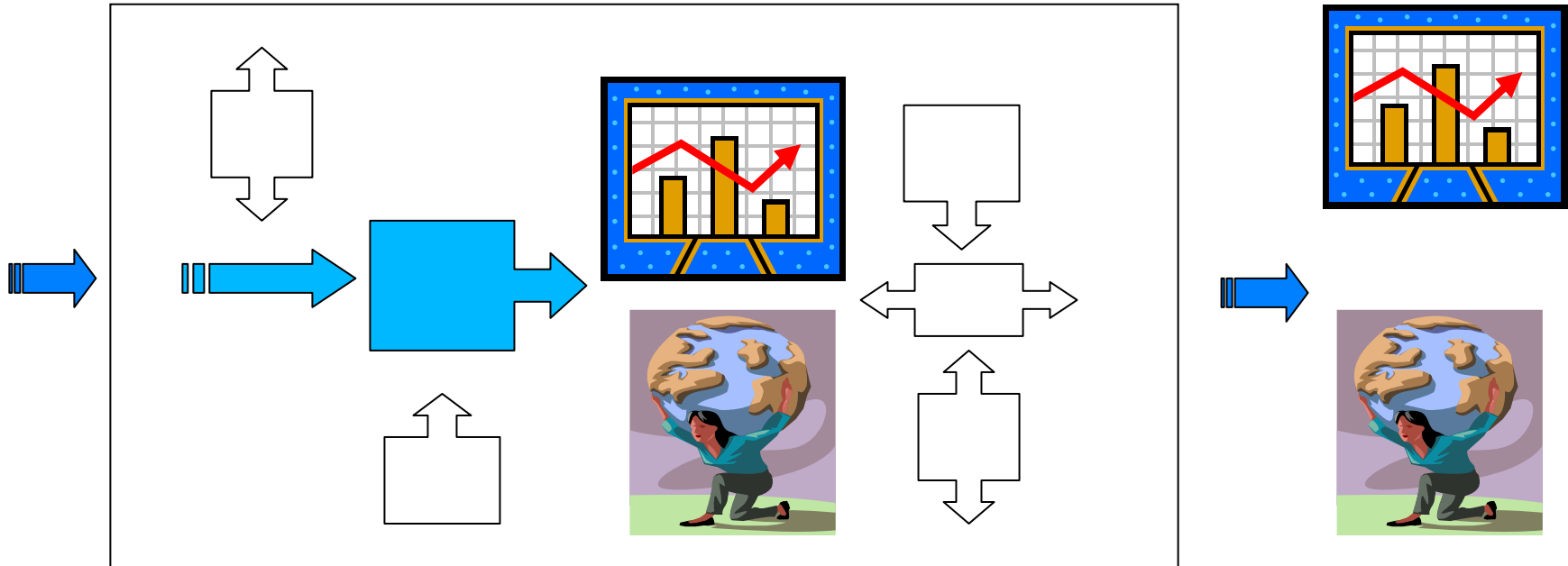
- Question Answering
- Summarization
- Machine translation
- .....

## PROS

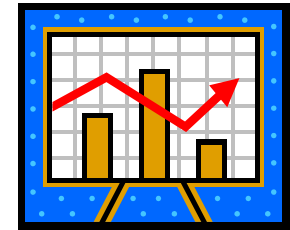
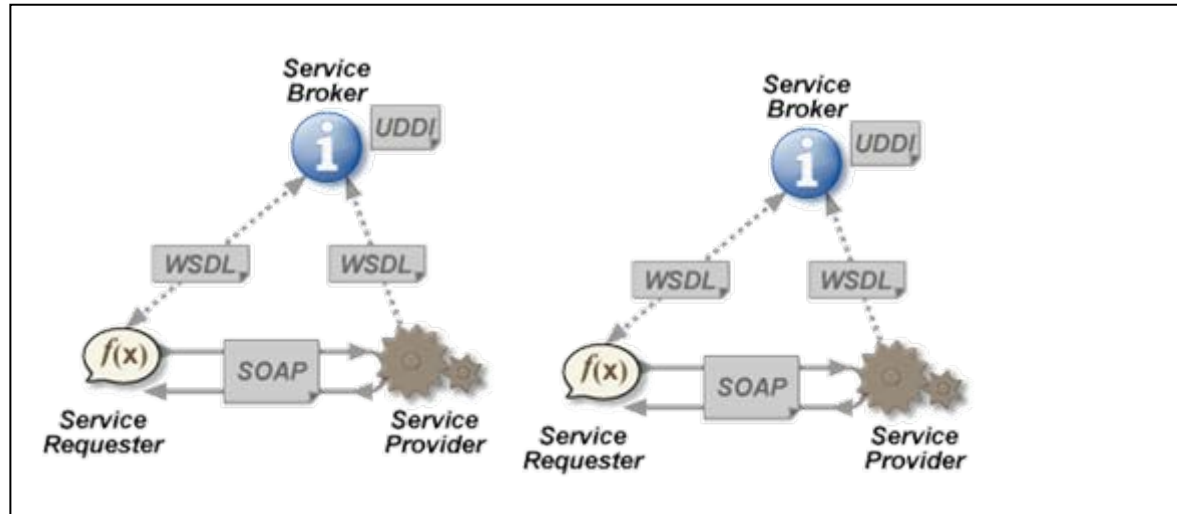
- Focus on real scenarios
- Users can be involved

## CONS

- Difficult to evaluate task contribution in applications
- Slow progresses



- Both single components and the global scenario are evaluated
- Progresses on components have direct impact on applications
- Research is focused on tasks with higher impact



- Define **communication protocols**, flexible enough to accommodate new research challenges, as well as multilingualism
- **Fast prototyping**: newcomers are facilitated to plug-in single components taking advantage of already established services
- **On-line evaluation services**: transparency, evaluation memories, efficiency issues, ...
- **Ongoing experiences**: Qallme for Question Answering, ....