

Pesquisa em Ciência da Computação

Ivandr  Paraboni

USP / EACH

Programa de P s-gradua o em Sistemas de
Informa o (PPgSI)

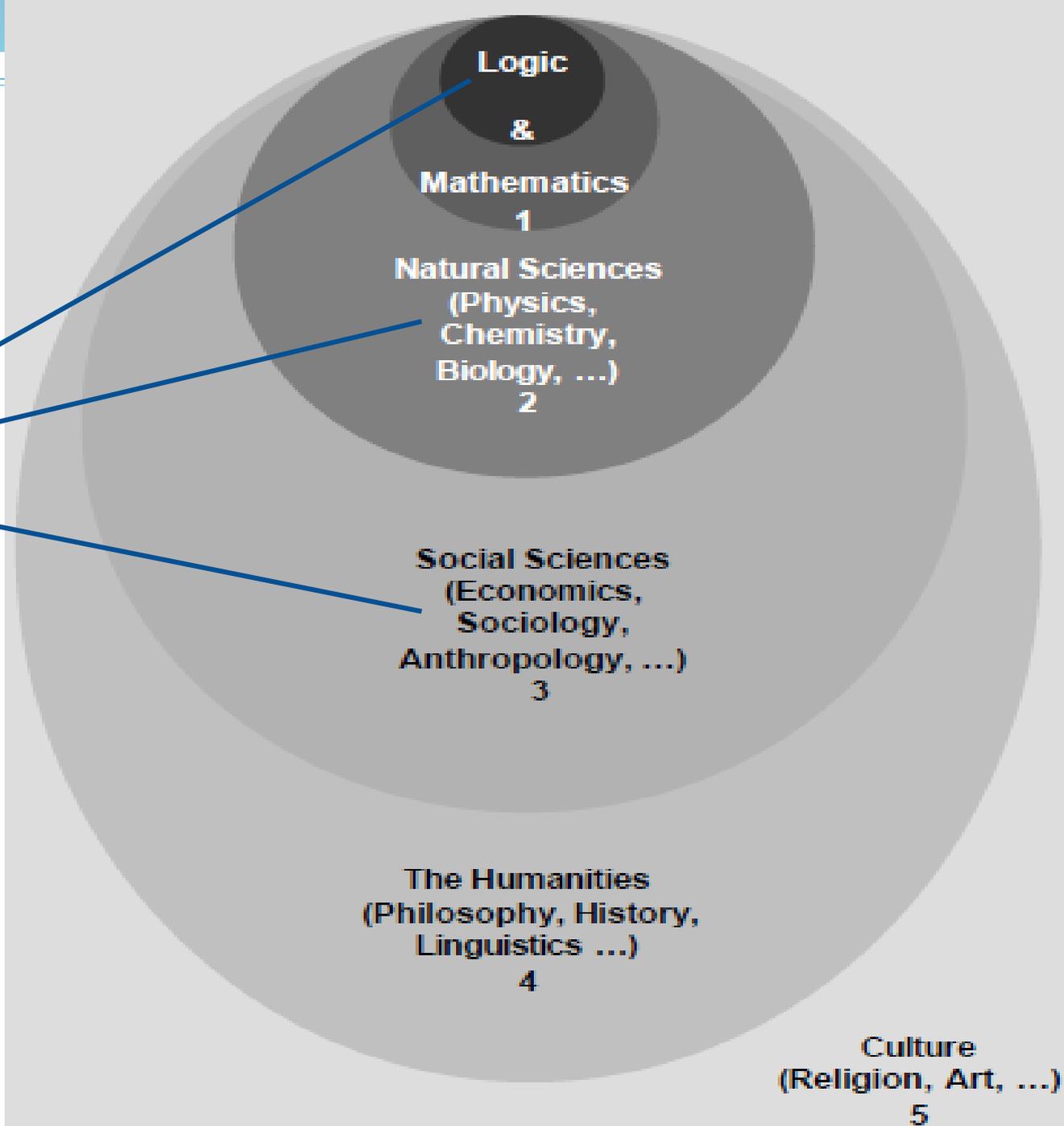


Ciência da computação

Ciência da Computação (CC)

- Surgida na década de 40.
- Baseada em uma série de disciplinas preexistentes.
- Como todas ciências modernas, é decorrente da maior facilidade de comunicação entre campos distintos do conhecimento.
 - Teoria da Computação / Bioinformática / Redes...?
- Combina teoria e prática, abstração (geral) e projeto (específico).

Ciência da
Computação



Computação: Ciência ou Tecnologia?

- Ciência e Tecnologia são diferentes em objeto de estudo, princípio de mudança, fins e atividades.
 - mas distinção não é perfeita.
- Conceitos tradicionais de ciência são obsoletos e não explicam a variedade de ciências modernas.
- Conceito moderno de ciência inclui CC e muitas outras áreas.

“Leis” produzidas em CC

- Em muitos sentidos, análoga à pesquisa em ciências tradicionais:
 - A coleta de dados sobre um sistema pode levar à criação de leis (mais) gerais sobre sistemas.
 - A descoberta de um fenômeno que não pode ser explicado pelas teorias vigentes leva à substituição da teoria.
- Mas as “leis” da CC possuem características únicas:
 - possuem um caráter *estatístico*, válidas para grandes quantidades de exemplos mas não necessariamente para um exemplo individual.
 - são “rasas” e mais próximas de descobertas do que de leis gerais das quais podemos tirar muitas conclusões.
 - possuem validade por um determinado período.

Um exemplo de “lei” em Engenharia de Software

- A manutenção de software consome pelo menos 50% do custo total do software (Huff, 1990).
 - é uma lei estatística; um projeto mais barato não a invalida.
 - não permite derivar muitas conclusões além do que ela diz.
 - Reflete um momento ou situação específicos.



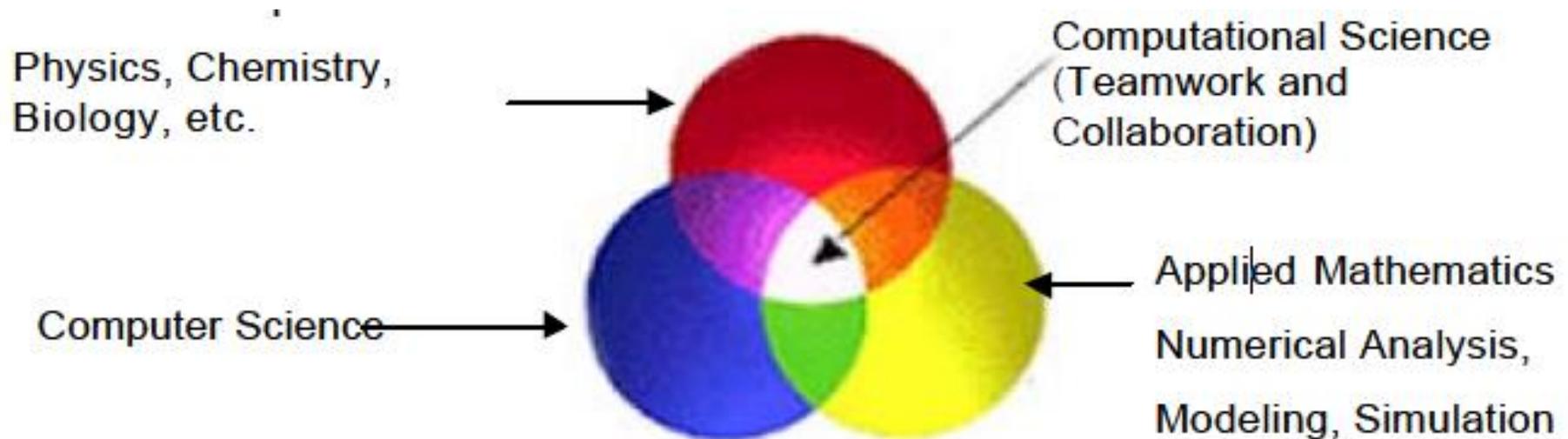
As 3 grandes áreas da Ciência da Computação

(1) Computação teórica

- Fortemente ligada à lógica e matemática.
 - metodologia clássica de construção de teorias como sistemas lógicos de objetos (axiomas) e operações (regras) para derivação e prova de teoremas.
- Novas teorias não substituem as anteriores, nem descrevem fenômenos novos.
 - base do conhecimento não é questionada por novas teorias.
- Forte ênfase na definição dos limites da computação.
 - $P = NP$?
- Projeto e análise de algoritmos e suas complexidades.
 - experimentos podem ser usados para avaliar o desempenho “prático” de algoritmos.
- Métodos típicos: iteração, recursão e indução.

(2) Simulação computacional

- Em anos recentes, a fusão de CC com outras ciências deu origem a uma nova metodologia, com foco na modelagem e simulação.



(2) Simulação computacional (cont.)

- Domínio de ferramentas computacionais avançadas são agora exigidos de várias áreas, e não apenas CC.
 - visualização 3D.
 - técnicas de simulação.
 - manipulação de grandes volumes de dados.
 - habilidade de acesso a dados distribuídos.
- Estas ferramentas possibilitam hoje avanços em todas áreas do conhecimento de uma forma que seria impensável algumas décadas atrás.
 - Computational Science
- Saiba mais:
 - Humphreys (2004) “Extending ourselves: Computational Science, Empiricism and Scientific Method”.

(3) Computação experimental

- O objeto de pesquisa em CC é a informação.
 - mesmos métodos de pesquisa da matéria, energia etc.
- Experimentos são usados para testar teorias e também para fins exploratórios.
- A comunidade gradualmente aceita uma teoria se
 - todos os fatos conhecidos sobre seu domínio podem ser deduzidos a partir dela.
 - a teoria resiste a uma quantidade considerável de testes.
 - a teoria faz previsões corretas.
- Experimentos devem ser repetidos muitas vezes, e seus resultados devem ser reproduzidos (replicação).
- Mas experimentos demonstram apenas a *presença* de falhas (ou *bugs*) no sistema, não a sua *ausência* (Dijkstra).

Experimentos em CC

- Experimentos objetivam avaliar a teoria proposta.
- No caso de proposta de novos algoritmos ou sistemas os experimentos correspondem à avaliação dos mesmos.
 - apresentando garantias formais de funcionamento.
 - demonstrando não apenas que é superior ao trabalho prévio, mas o quanto é superior.
- Forte ênfase em **métodos quantitativos**

O que avaliar?

- O modelo proposto precisa ser avaliado por meio de medidas ou métricas
- Refletem as propriedades importantes do modelo
- Só há sentido em medir aquilo que corrobora para a teoria testada
- Exemplo: foi desenvolvido um sistema X para melhorar a taxa de acertos em um determinado problema
 - Avaliação: medida de acurácia e tempo de execução
- Identificar métricas adequadas pode ser difícil
 - Especialmente em projetos mal formulados

Exercício

1. Imagine o que você pretende desenvolver no seu mestrado como sendo um produto (científico)
 - Exemplos: um sistema, modelo, artefato, algoritmo etc.

Exercício

1. Imagine o que você pretende desenvolver no seu mestrado como sendo um produto (científico)
 - Exemplos: um sistema, modelo, artefato, algoritmo etc.
2. Pense no objetivo deste produto.
Qual a contribuição prevista?
Para que ele serve?
Em que o produto final da pesquisa é melhor do que as alternativas existentes, se houver?

Exercício

1. Imagine o que você pretende desenvolver no seu mestrado como sendo um produto (científico)
 - Exemplos: um sistema, modelo, artefato, algoritmo etc.
2. Pense no objetivo deste produto.
Qual a contribuição prevista?
Para que ele serve?
Em que o produto final da pesquisa é melhor do que as alternativas existentes, se houver?
3. As vantagens ou benefícios esperados precisam ser **quantificados**.
Quais métricas seriam aplicáveis neste caso?

Avaliar não é apenas medir

- A medição das propriedades do modelo desenvolvido é um aspecto fundamental da produção científica, mas não é tudo.
- É preciso um **termo de comparação (baseline)** para dar sentido ao que foi medido.
e.g., o sistema que representa o **estado da arte** ou, na falta deste, um baseline próprio que seja tão **robusto** quanto possível.
- A diferença entre o resultado do modelo proposto e baseline precisa ser **significativa**, com uso de um **teste estatístico** adequado para o tipo de medida empregada.

A importância da avaliação

- Uma avaliação completa e correta, com uso de um ou mais baseline robusto e significância estatística, é um indicador seguro da qualidade da pesquisa.
- Este assunto será tratado em maior detalhe em outro momento da disciplina e, de modo especial, na atividade de elaboração do projeto.
- Por enquanto, comece a se habituar ao tipo de avaliação feita na sua área:
 - O que é medido (métricas)?
 - Quais os termos de comparação tipicamente empregados?
 - Quais os testes estatísticos utilizados?



Avaliação intrínseca versus extrínseca

Propriedades do modelo

- Intrínsecas
 - Inerentes ao modelo
 - e.g., acurácia, tempo de execução, número de falsos positivos etc.
- Extrínsecas
 - Externas ao modelo
 - e.g., os efeitos do modelo sobre outro modelo / sistema etc., avaliação de usuários etc.

Avaliação intrínseca

- Avaliação de propriedades do sistema independente do uso.
 - Cada área possui suas próprias métricas estabelecidas.
- São medidas propriedades intrínsecas do modelo (como acurácia, tempo de execução etc.)

Um exemplo de avaliação intrínseca

- Problema: classificação de documentos (e.g., e-mails) em categorias (esportes, ciências, etc.)
- Proposta: um modelo do tipo X que seja melhor do que os existentes para esta tarefa
 - Melhor COMO?

Um exemplo de avaliação intrínseca

- Problema: classificação de documentos (e.g., e-mails) em categorias (esportes, ciências, etc.)
- Proposta: um modelo do tipo X que seja melhor do que os existentes para esta tarefa
 - Melhor COMO?
- Avaliação intrínseca: medida de acurácia para a tarefa de classificação de e-mails do conjunto *20-NewsGroup*.
 - 20 classes de documentos manualmente anotados e revisados.
 - conjuntos balanceados.
 - grande volume de dados.
 - acurácia de outros sistemas é amplamente conhecida.

Avaliação extrínseca

- Em certos casos pode ser importante questionar se, além de formalmente correto, o trabalho atende objetivos práticos.
 - É útil? É bom? Resolve o problema?
- Medidas intrínsecas avaliam aspectos isolados da solução.
 - podem apresentar baixa correlação com o impacto real.
- Medidas extrínsecas avaliam o propósito do sistema e sua capacidade global de desempenhar a tarefa em questão.
 - podem ser mais gerais do que as medidas intrínsecas.

De volta à classificação de e-mails

- Suponha que o modelo desenvolvido não apresenta resultados muito convincentes
- Ou que na verdade meu interesse não era obter máxima acurácia, e sim “simplificar” a vida do usuário
- Simplificar COMO?
 - Por exemplo, reduzindo o tempo médio diário gasto na organização da caixa de entrada.
- Objetivo do modelo pode ser obter um determinado efeito sobre algo externo ao modelo
 - redução de tempo do usuário
- A medição de propriedades externas ao modelo é uma avaliação extrínseca



Dois exemplos

Exemplo 1

- Problema: geração de expressões espaciais em mundos virtuais tridimensionais (Silva & Paraboni, 2015)
- Objetivo do modelo era produzir instruções de navegação
 - Pressione o botão verde
 - Abra a porta da direita
 - Caminhe até o fim do corredor

Sistema: “Pressione o **botão verde à direita da planta**”



Como avaliar isso?

- É preciso definir claramente quais os objetivos da solução proposta
 - Poderíamos pensar em um modelo que é rápido?
 - Que gera instruções gramaticalmente corretas?
- Foco do trabalho foi a naturalidade (*humanlikeness*) das instruções geradas de forma automática
 - Objetivo era gerar instruções que fossem o mais semelhante possíveis às instruções que um instrutor humano teria produzido naquela mesma situação
- Então como avaliar?



Abordagem adotada

- Técnica de avaliação: apenas intrínseca
 - comparando os resultados do modelo proposto com um conjunto de dados (um cópús de instruções de navegação) preexistente.
 - Expressões gerada pelo sistema X expressões do conjunto de referência
- Poderia ter sido feita também de forma extrínseca?
 - Como?



Exemplos de métricas extrínsecas possíveis

- Tempo de navegação do usuário
- Número de erros cometidos
- Quantidade, tempo ou distância de caminhos refeitos
- Satisfação global

Não esqueça

- A avaliação extrínseca pode substituir ou complementar a avaliação intrínseca
- Quanto mais avaliação, melhor!

Exemplo 2

- STOP (Reiter et. al., 2003) é um sistema de geração automática de cartas personalizadas para incentivar fumantes a abandonar o cigarro.
- Cartas personalizadas eram produzidas a partir de um questionário de 4 páginas respondido pelo destinatário.
- Avaliação extrínseca: testes clínicos com 2553 fumantes recebendo cartas personalizadas ou um texto fixo.
 - objetivo: verificar se o emprego de cartas personalizadas pelo sistema, enfatizando as circunstâncias de cada fumante, **acarretaria maiores índices de abandono do cigarro** do que o texto padronizado.

O sistema STOP

- Resultados revelaram que a probabilidade de abandono do cigarro *era a mesma*, tanto no caso de cartas personalizadas quanto no caso do texto padrão.
 - não há vantagem significativa em usar o sistema STOP.
- Considerações:
 - Autores não sabem exatamente o porquê do fracasso.
 - Possíveis causas: avaliação muito rigorosa, aplicação impossível, falta de conhecimento sobre o problema, erros técnicos etc.
 - Resultado negativo ainda pode ser relevante (e citável!).
- Mais detalhes:
 - Reiter et. al. (2003) “Lessons from a failure: Generating tailored smoking cessation letters”.

Experimentação e o limite do observável

- Abordagens empíricas tradicionalmente fazem a distinção entre fenômenos observáveis e não observáveis.
 - visão antropocêntrica
 - limitada aos 5 sentidos.
- Avanços da computação no entanto permitem a expansão de nossas habilidades, frequentemente com desempenho superior ao humano.
 - telescópios automáticos.
 - mapeamento genético.
 - simulação e previsão de condições climáticas.

Intrínseca ou extrínseca?

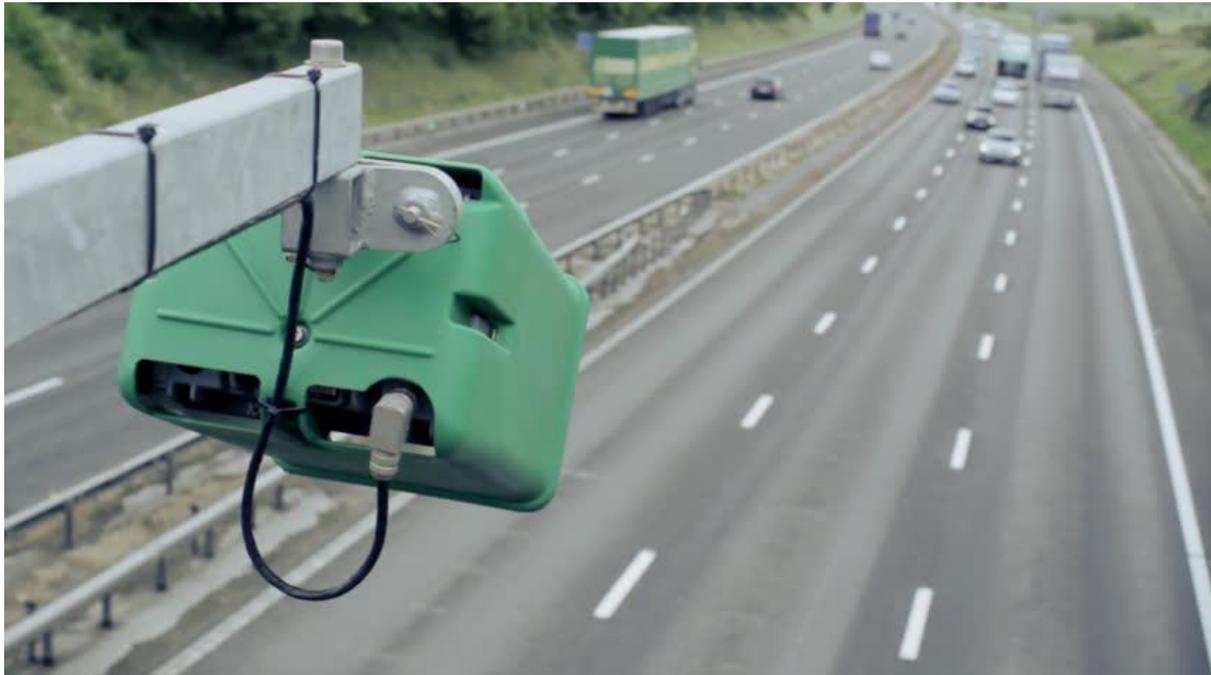
- Avaliação intrínseca é mais barata.
 - se houver dados e métricas disponíveis.
- Em muitos casos é a única técnica viável.
- Avaliação extrínseca complementa a análise e enriquece a pesquisa.
 - útil também quando os resultados de outras formas de avaliação não foram totalmente conclusivos.
- Idealmente deveriam ser realizadas ambas! 😊

Comparar com o quê?

- Os resultados do sistema ou artefato proposto só fazem sentido se representarem um ganho significativo em relação a um termo de comparação plausível.
 - Sistema de baseline
- Um *benchmark* fornece resultados prontos que podem representar o estado-da-arte.
 - Na ausência destes, será preciso propor – e possivelmente implementar - um ou mais sistemas de baseline **robustos**.
- A significância do ganho observado é avaliada com testes estatísticos específicos para a natureza dos dados / medida de comparação.
 - Respeitada a prática da área.

Exercício

- Seja um novo sistema de reconhecimento de imagens que, dada a fotografia de uma placa de automóvel, identifique a sua numeração.
- Que tipos de avaliação intrínseca e extrínseca você sugere?



Considerações

- Todo trabalho de pesquisa precisa ser avaliado
 - Com métricas adequadas
 - Comparando-o com o estado-da-arte (um baseline dito “robusto”)
 - Com teste de significância estatística adequados ao que está sendo medido.
- Mas a contribuição do meu trabalho é “teórica”...
 - Um levantamento de dados?
 - Um mapeamento de um determinado fenômeno?
 - Uma avaliação de alguma coisa?
 - Uma análise crítica do estado-da-arte?



E na minha área?

Entrega e02: avaliação de artigo

- Escolha um artigo relevante para seu projeto para analisar o método de avaliação utilizado, preferencialmente publicado em periódico ou conferência de alto impacto **e que tenha necessariamente algum tipo de avaliação.**
- Pergunta-se:
 1. Qual tipo de avaliação foi feito?
 2. Foram utilizadas métricas de avaliação? Quais?
 3. Houve uso de testes estatísticos? Quais?
 4. Houve comparação com trabalhos similares e/ou sistemas de baseline? Se sim:
 - a) Estes trabalhos, se houver, foram propostos por terceiros, ou foram desenvolvidos pelos mesmos autores do artigo?
 - b) São recentes?
- Elabore uma **apresentação de 10 minutos** em formato **PDF não compactado**. A apresentação deve fazer uma breve introdução ao tema e ao objetivo do artigo, explicando brevemente (para uma audiência não-especialista) o que foi feito, e então usar a maior parte do tempo para discutir os pontos acima.
- Alguns exemplos serão selecionados para apresentação em aula - esteja preparado.

Referências

- Dodig-Crnkovic, Gordana (2002) *Scientific Methods in Computer Science*. Proc. Conf. for the Promotion of Research in IT at New Universities and at University Colleges in Sweden.
- Sociedade Brasileira de Computação (2006) *Grandes Desafios da Computação 2006-2016*. Disponível em www.sbc.org.br
- Jacques Wainer (2007) *Métodos de pesquisa quantitativa e qualitativa para a Ciência da Computação*. In: Tomasz Kowaltowski and Karin Breitman. (Org.). *Atualização em informática 2007*. Sociedade Brasileira de Computação e Editora PUC-Rio, pp. 221-262.
- Ted Pederson (2008) *Empiricism is not a matter of faith*. *Computational Linguistics*(34) vol.3.
- Anja Belz (2009) *That's Nice . . . What Can You Do With It?* *Computational Linguistics* (35) vol.1, pp. 111-118.
- Humphreys (2004) *Extending ourselves: Computational Science, Empiricism and Scientific Method*. Oxford University Press.



Obrigado