

Web Semântica e Processamento de Linguagem Natural na prática

Newton Calegari



PUC-SP

Newton Calegari
Bacharel Ciência da Computação
Mestre em Tecnologias da Inteligência e Design Digital

Newton Calegari
Bacharel Ciência da Computação
Mestre em Tecnologias da Inteligência e Design Digital

Data on the Web Best Practices

W3C Recommendation 31 January 2017

This version:

<https://www.w3.org/TR/2017/REC-dwbp-20170131/>

Latest published version:

<https://www.w3.org/TR/dwbp/>

Latest editor's draft:

<http://w3c.github.io/dwbp/bp.html>

Implementation report:

<http://w3c.github.io/dwbp/dwbp-implementation-report.html>

Previous version:

<https://www.w3.org/TR/2016/PR-dwbp-20161215/>

Editors:

Bernadette Farias Lóscio, [CIn - UFPE, Brazil](#)

Caroline Burle, [NIC.br, Brazil](#)

Newton Calegari, [NIC.br, Brazil](#)

Objetivo

O objetivo do curso é apresentar conceitos e fundamentos de Web Semântica e direcionar para o uso e implementações práticas das tecnologias que fazem parte desta abordagem.

Introdução e contexto

O surgimento da World Wide Web

A evolução da Web e o desenvolvimento de aplicações

A proposta da Web Semântica

Web Semântica e contexto

Os fundamentos da Web Semântica

A pilha tecnológica da Web Semântica

Ontologias e Vocabulários

Estado da arte e exemplos

RDF e o grafo de conhecimento

Triplas RDF

Grafos de dados

Resource Description Framework e as linguagens de
serialização

Bancos de dados RDF e a linguagem SPARQL

Conhecendo os triple stores

Utilizando bancos de dados RDF

Executando queries SPARQL

Ontologias e Vocabulários

Descrição de dados na Web

Modelos de domínio

Repositórios de ontologias e vocabulários

RDFa e JSON-LD para desenvolvedores web

Especificações RDFa e JSON-LD

Aplicação prática e testes com as linguagens RDFa e
JSON-LD

Desenvolvimento de aplicações semânticas

Frameworks com suporte para RDF
RDFa e JSON-LD para front-end
APIs para consumo de dados
Content-negotiation via http

Inteligência Artificial e Web Semântica

Modelos e abordagens de Inteligência Artificial
Machine learning, deep learning e Processamento de
Linguagem Natural
Intersecções com Web Semântica

Inteligência Artificial e Web Semântica

Utilização de framework PLN

Utilização de IA para busca de entidades em texto

Web Semântica como base para o futuro

Revisão do conteúdo do curso

Próximos passos e aplicações no curto, médio e longo prazo

Análise da aplicação de Web Semântica em áreas como Internet das Coisas e Realidade Virtual

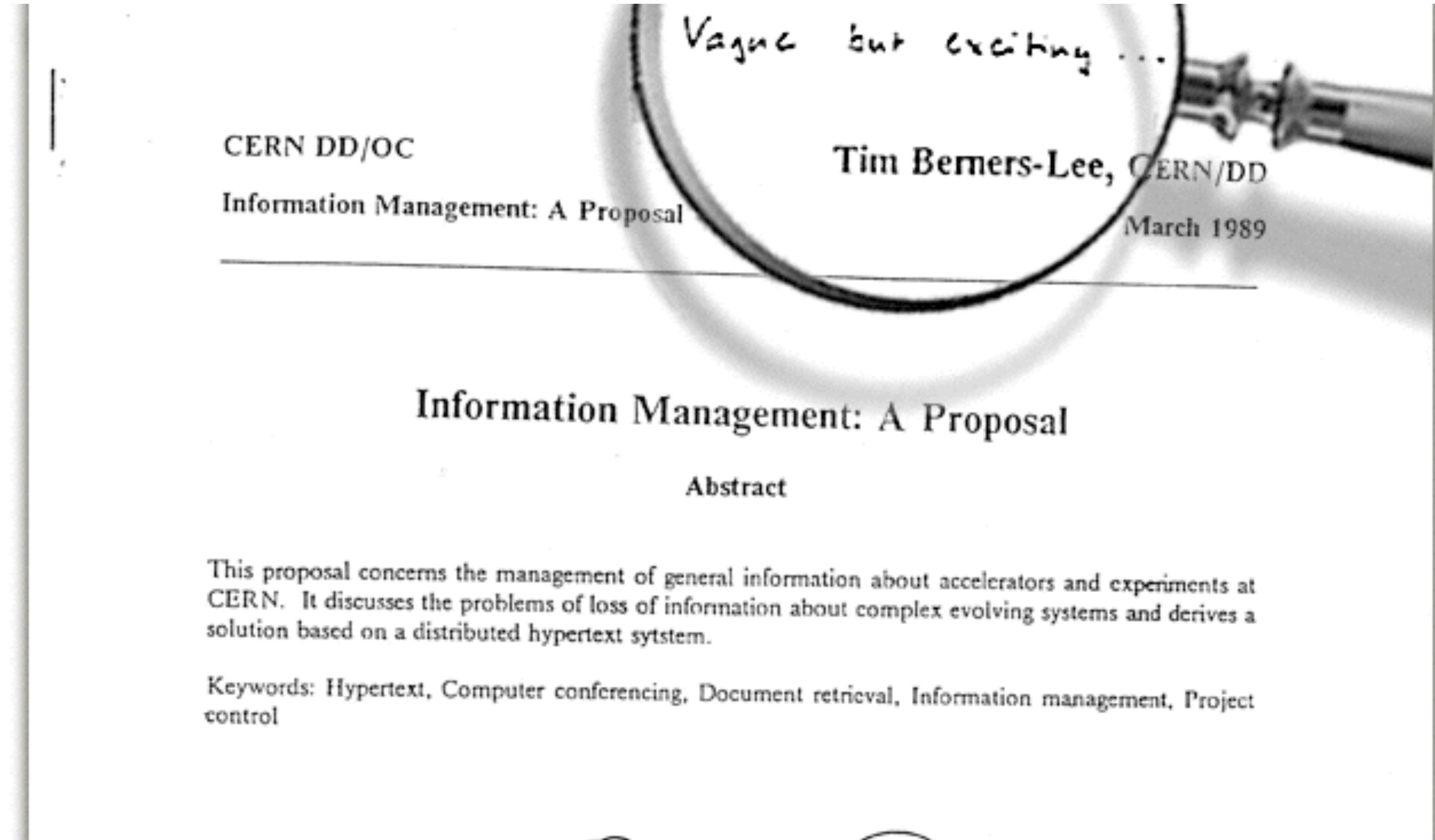
Introdução e contexto

O que é Web?

**Internet e Web
não são as mesmas
coisas**

As tecnologias fundamentais da Web

WORLD WIDE WEB VAGUE BUT EXCITING



<http://info.cern.ch/Proposal.html>

WORLD WIDE WEB TECNOLOGIAS FUNDAMENTAIS

- Páginas web (documentos de hipertexto)
- Protocolo HTTP
- Linguagem HTML
- Identificador de recursos (URI)

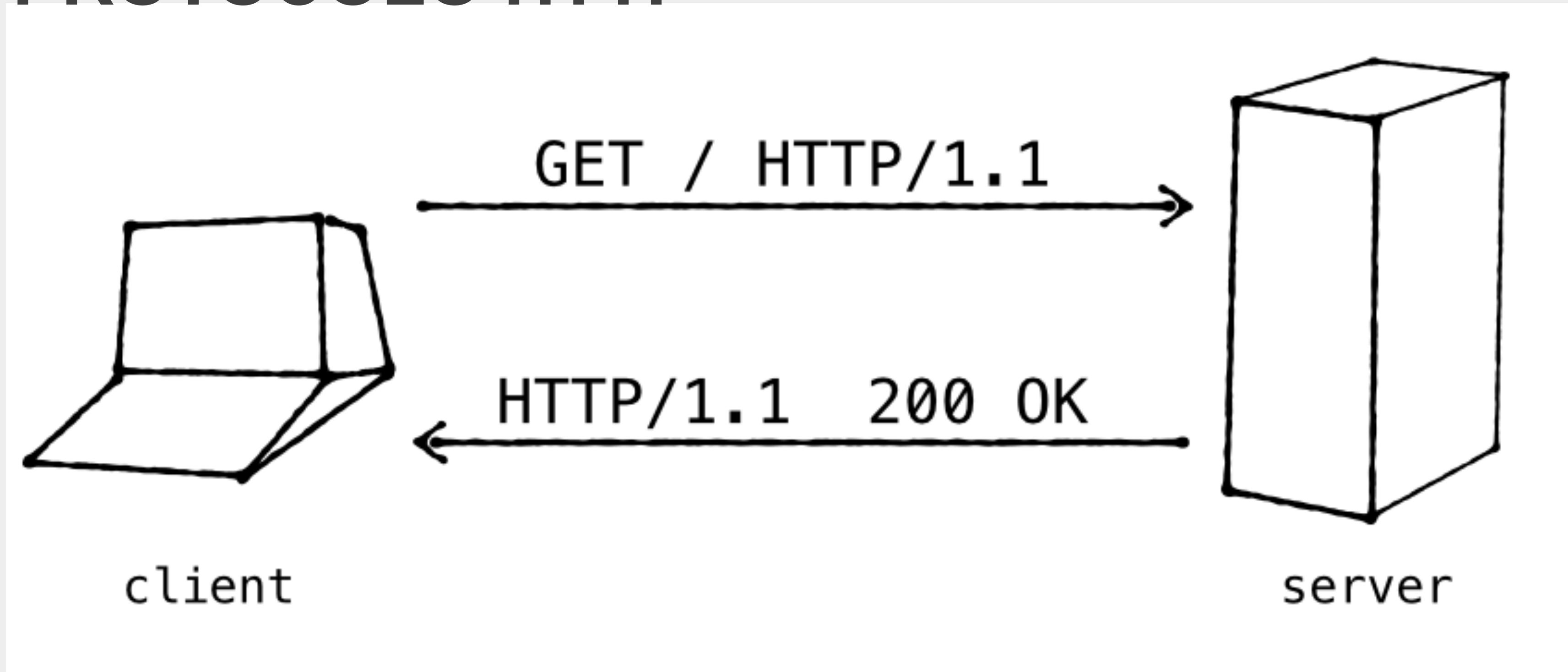
<http://info.cern.ch> - home of the first website

From here you can:

- [Browse the first website](#)
- [Browse the first website using the line-mode browser simulator](#)
- [Learn about the birth of the web](#)
- [Learn about CERN, the physics laboratory where the web was born](#)

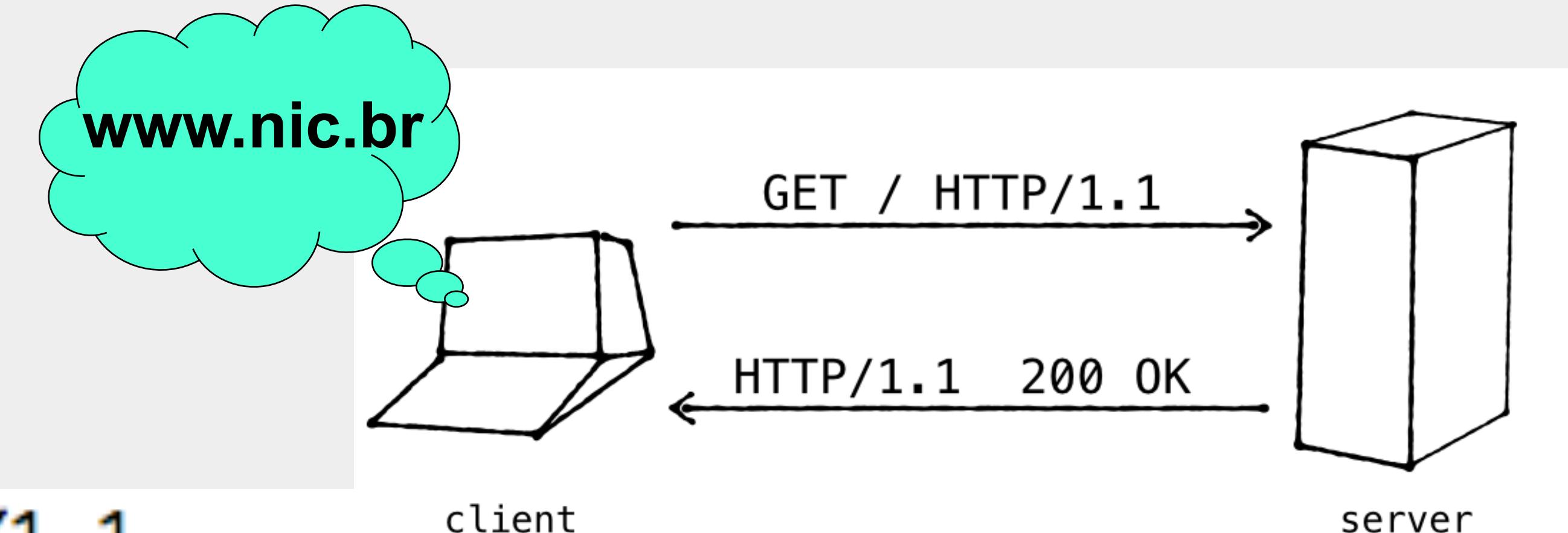
WWW

PROTOCOLO HTTP



- **HTTP 0.9** [Tim Berners-Lee 1991]
- **HTTP/1.0** [RFC 1945, Maio 1996]
- **HTTP/1.1** [RFC 2068, Janeiro 1997]
- **HTTP/2** [RFC 7540, Maio 2015]

HTTP REQUEST



GET / HTTP/1.1

Host: nic.br

Connection: close

User-Agent: Mozilla/5.0 Firefox/39.0

Accept-Encoding: gzip

Accept-Charset: ISO-8859-1,UTF-8;q=0.7,*;q=0.7

Cache-Control: no-cache

Accept-Language: de,en;q=0.7,en-us;q=0.3

HTTP RESPONSE

HTTP/1.1 200 OK

Date: Tue, 18 Aug 2015 00:26:31 GMT

Server: Apache/2.2.22

Last-Modified: Mon, 17 Aug 2015 19:21:31 GMT

Accept-Ranges: bytes

Content-Length: 42093

Vary: Accept-Encoding

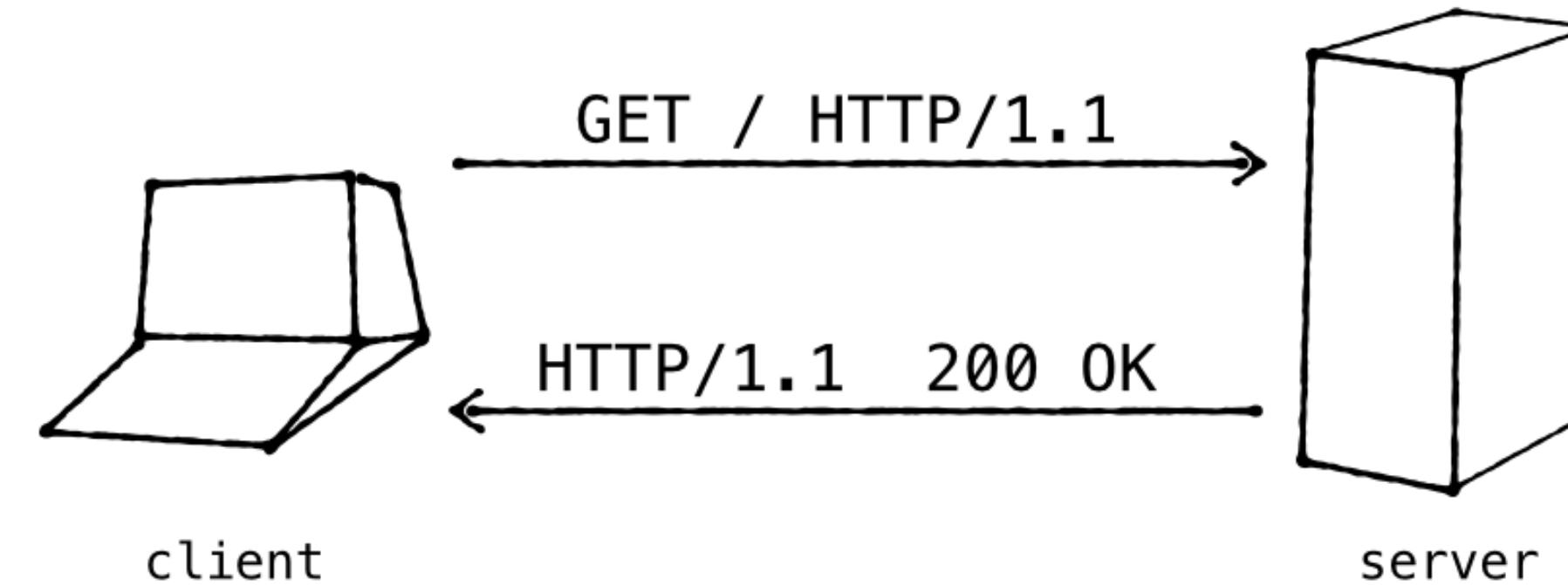
Content-Type: text/html

<!DOCTYPE html>

<html lang="pt-br">

 <head>

 <meta charset="utf-8">



WWW LINGUAGEM HTML

- Linguagem de hipertexto para documentos (páginas web)
- Baseada em SGML [ISO 8879:1986]
- Formada por tags (<html> <head> <body> ...)
- Possui link hipertexto ()
- Interpretada pelos navegadores/browsers
- Lista WWW-talk criada em 1991
- Atual versão HTML5

WWW LINGUAGEM HTML

```
<!DOCTYPE html>
<html lang="pt-br">
<head>
  <meta charset="utf-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="initial-scale = 1.0,maximum-
  <meta name="author" content="NIC.br">
  <meta name="keywords" content="nic, nic.br, internet no b
  <meta name="description" content="O Núcleo de Informação e
  <meta name="google-site-verification" content="nDjnC4ZPe8Y4
```

WWW URI, URL, DOMÍNIOS

- Nomes de domínios *
- DNS [RFC 882, Novembro 1983]
- URIs na WWW [RFC 1630, Junho 1994]
- URL: localizador de recursos na Web
- Importância para Web semântica
- Identificar recursos e conceitos

* Mais sobre TLDs em <https://www.iana.org/domains/root/db>

WWW URL

http://ceweb.br/contato/index.html

esquema://domínio:porta/caminho/recurso?query_string#fragmento

esquema: http

domínio: ceweb.br

porta: 80

caminho: /contato

recurso: index.html

WORLD WIDE WEB

TECNOLOGIAS FUNDAMENTAIS

- Páginas web (documentos de hipertexto)
- Protocolo HTTP
- Linguagem HTML
- Identificador de recursos (URI)

**Porque padrões web são
importantes?**

Os padrões estão consolidados

Tim Berners-Lee inventou a Web em 1989 e fundou o W3C em 1994

Já foram publicados mais de 110 padrões



A.M. Turing Award

amturing.acm.org

A.M. TURING CENTENARY CELEBRATION WEBCAST

acm

MORE ACM AWARDS

A.M. TURING AWARD

A.M. TURING AWARD WINNERS BY...

ALPHABETICAL LISTING

YEAR OF THE AWARD

RESEARCH SUBJECT

FEATURED AWARD WINNER:
Sir Tim Berners-Lee

For inventing the World Wide Web, the first web browser, and the fundamental protocols and algorithms allowing the Web to scale.

Inventor of World Wide Web Receives ACM A.M. Turing Award
Sir Tim Berners-Lee Designed Integrated Architecture and Technologies that Underpin the Web

ACM named Sir Tim Berners-Lee, a Professor at Massachusetts Institute of Technology and the University of Oxford, the recipient of the 2016 ACM A.M. Turing Award. Berners-Lee was cited for inventing the World Wide Web, the first web browser, and the fundamental protocols and algorithms allowing the Web to scale. Considered one of the most influential computing innovations in history, the World Wide Web is the primary tool used by billions of people every day to communicate, access information, engage in

O processo e os padrões são abertos

São desenvolvidos pela equipe do W3C,
membros e com participação ativa da
comunidade.

Status das especificações

Indicam o nível de estabilidade dos documentos

Working Draft (WD)

Last Call (LC)

Candidate Recommendation (CR)

Proposed Recommendation (PR)

Recommendation (REC)

Os padrões são interoperáveis

O uso de padrões proporciona que as páginas sejam acessadas por todos.

Garante compatibilidade

Sua página abre corretamente nos navegadores que seguem padrões.

Passado



Presente



A página em [http://www.google.com](#) diz:



ATENÇÃO ! Para o registro de ocorrências, utilize MICROSOFT INTERNET EXPLORER VERSÃO 4 ou superior.

OK

Para que a oco
corretos.

Durante a conferência, pode falar
pelo telefone info@... ou

Após, a ocorrência narrado, no caso Delegacia da locação "Dados Básicos".

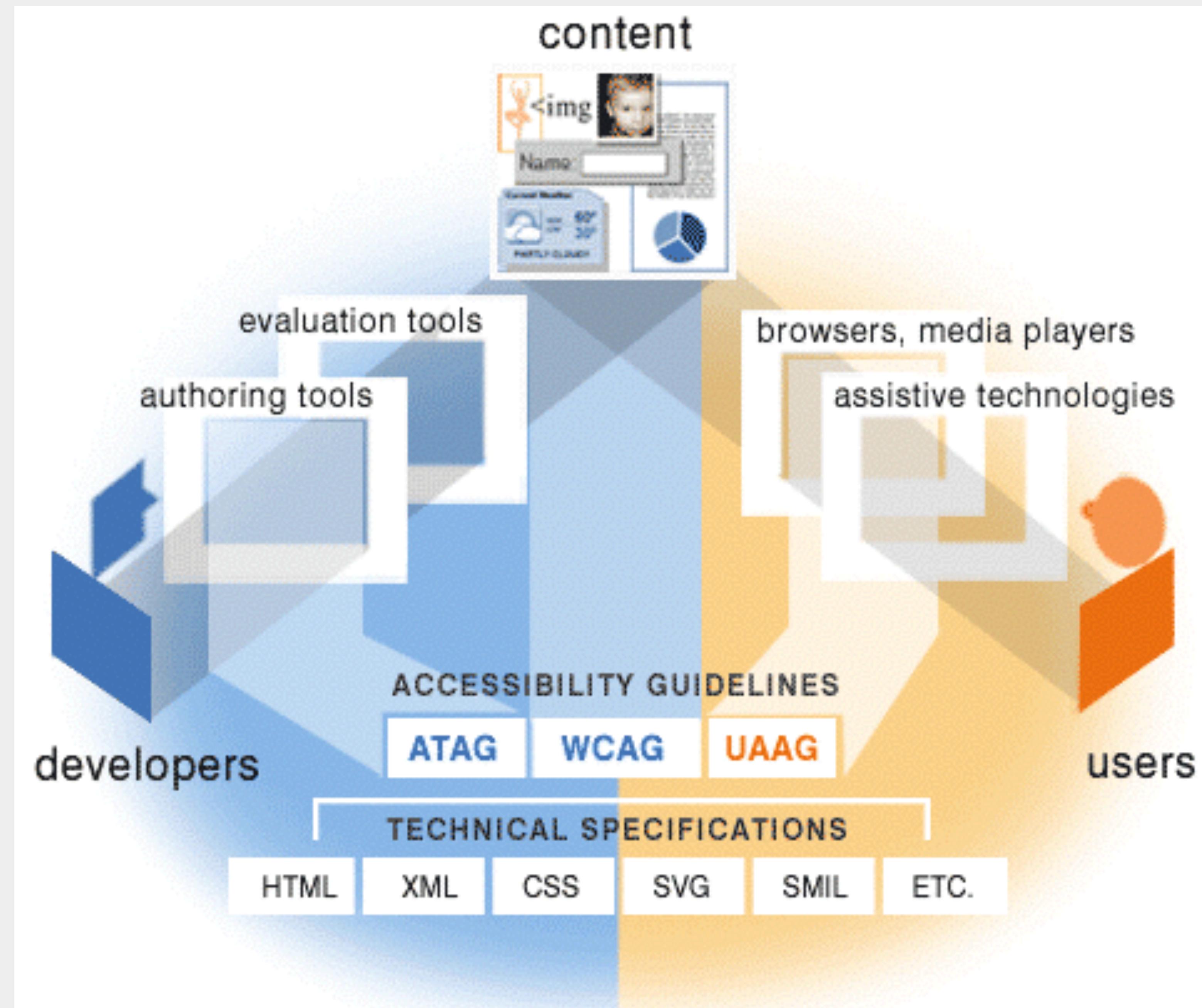
Esclarecemos que o seu computador será monitorado durante o registro virtual da ocorrência e lembramos que a comunicação falsa de crime ou contravenção tem pena prevista no artigo 340 do Código Penal Brasileiro.

Pessoa
Desaparecida

Ajuda

É acessível

Possibilita o acesso de pessoas com deficiências às informações.



Você também pode criar um padrão

Mas vai ter muito trabalho pela frente...

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



YEAH!

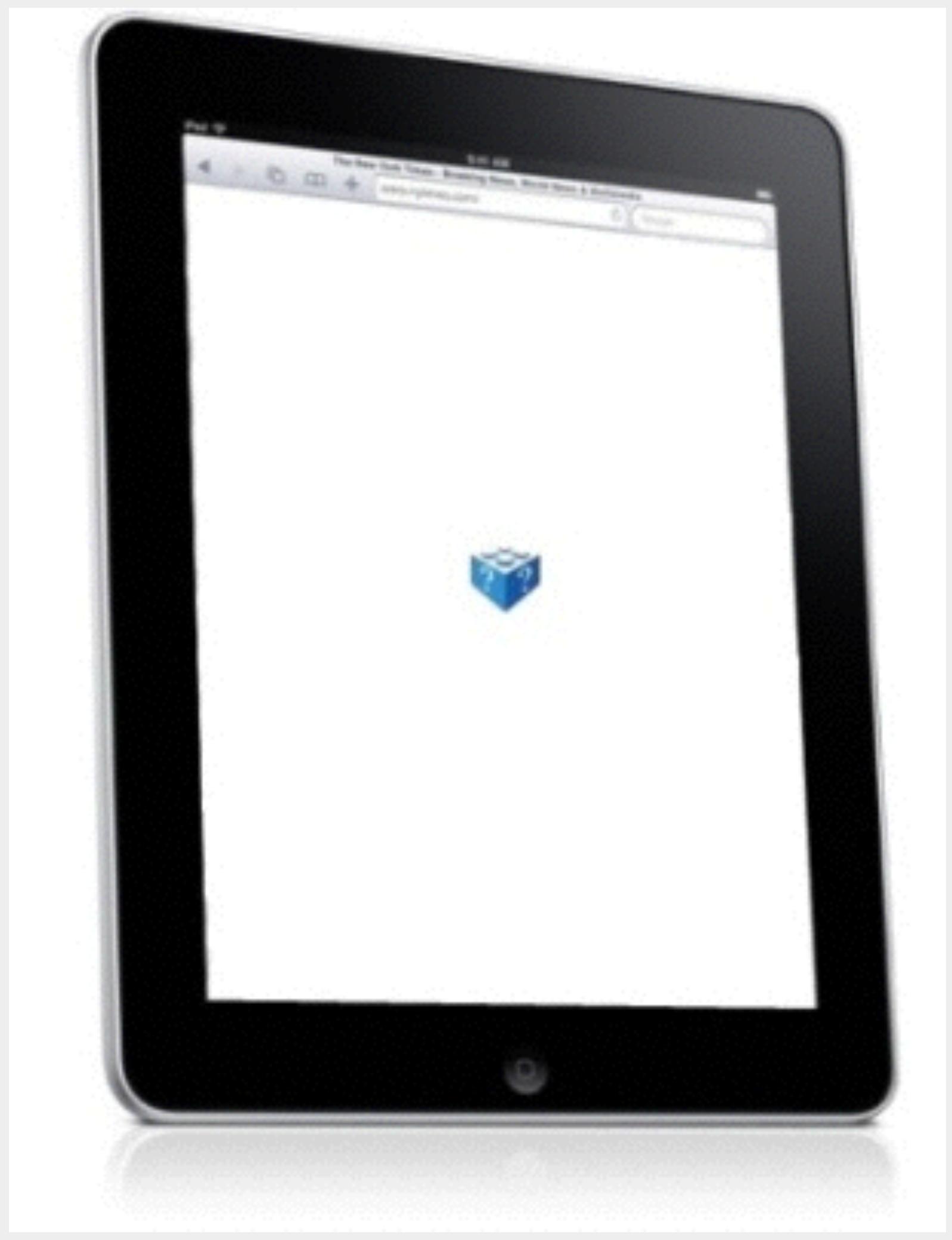
SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.



Compatibilidade

Instalação de plug-ins e dependência de sistemas operacionais



[Brasil](#) | [Mundo](#) | [Economia](#) | [Política](#) | [Esportes](#) | [Carros](#) | [Emprego](#) | [Educação](#) | [Saúde](#) | [Tech](#) | [Bizarro](#) | [Pop&Arte](#) | [MG](#) | [RJ](#) | [SP](#) | [Telejornais](#) | [Virada de ano](#)

| CES 2014

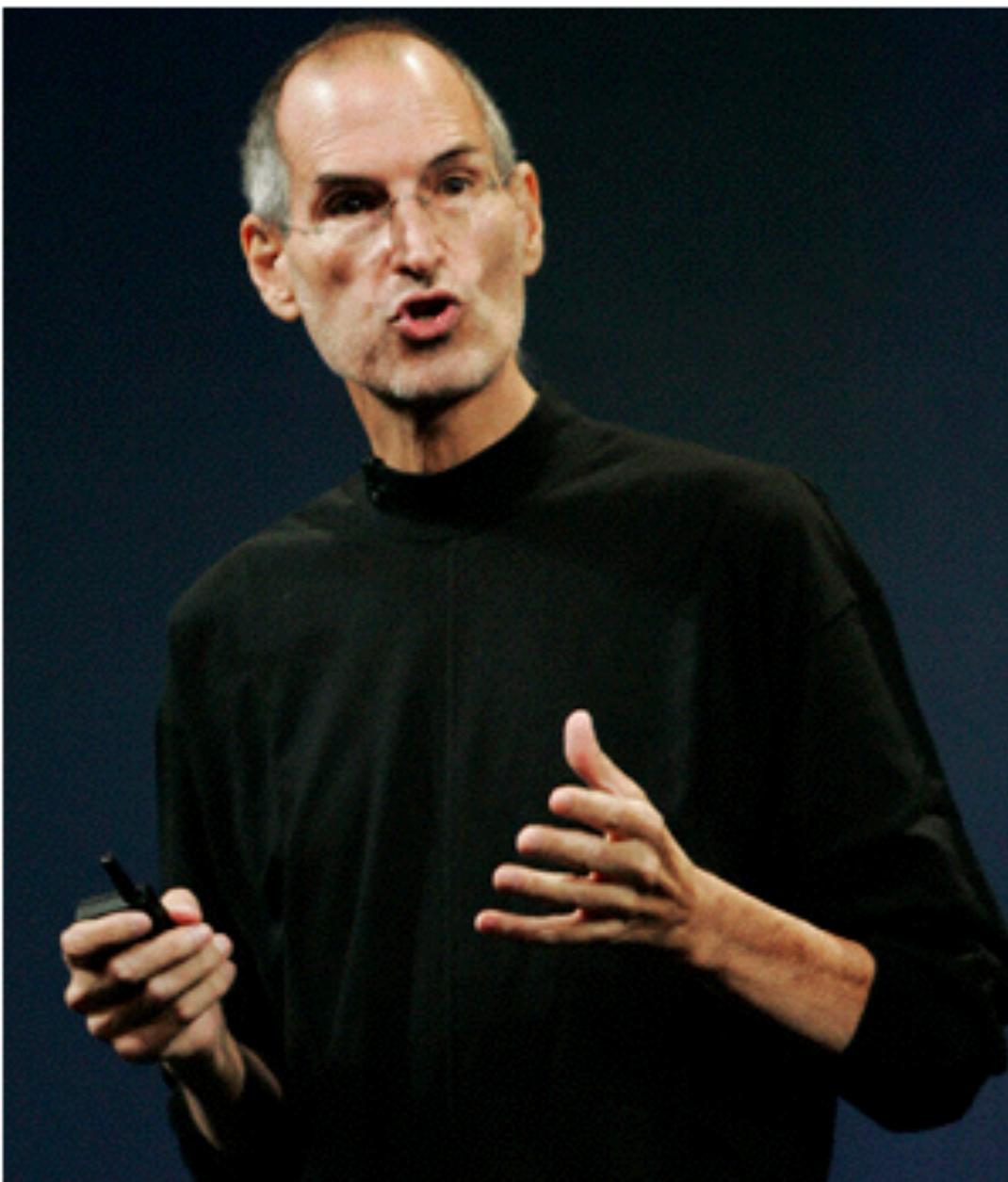
29/04/2010 17h39 - Atualizado em 20/05/2010 07h09

Steve Jobs explica razões para não usar Flash em iPhone e iPad

Carta aberta foi publicada no site da Apple nesta quinta-feira (29). 'Aplicativos mostram que o Flash não é necessário', diz Jobs.

Do G1, em São Paulo

imprimir



Em carta aberta publicada no site da Apple nesta quinta-feira (29), Steve Jobs listou algumas das razões para não utilizar o Adobe Flash em aparelhos como o iPad ou iPhone.

O documento, intitulado "Thoughts on Flash" ("Pensamentos sobre o Flash") começa reconhecendo a "longa relação" entre as companhias. "Nós conhecemos os fundadores da Adobe quando eles ainda estavam em sua garagem. A Apple foi seu primeiro grande cliente, adotando sua linguagem Postscript em nossas então novas impressoras Laserwriter. A Apple investiu na Adobe e possuiu cerca de 20% da companhia por muitos anos", escreveu Jobs.

Tecnologia e Games

17 AGO



10:31
Conheça os 'Zuckerbergs' latino-americanos; brasileiro criou...



08:17
Alphabet: veja lista de todas as letras da 'nova Google', de...



06:00
'Hotline Miami': Boneco baseado no game troca até de máscara; veja...

16 AGO



19:32
Mais lidas do dia:
16/agosto/2015



08:38
Como recuperar os arquivos que você apagou sem querer do...

PUBLICIDADE

“Nós realmente acreditamos que todos os padrões pertencentes à web deveriam ser abertos. Em vez de usar Flash, a Apple adotou o HTML5, CSS e JavaScript – todos padrões abertos”

<http://www.apple.com/hotnews/thoughts-on-flash/>

“Recentemente, a Symantec destacou o Flash por possuir um dos piores recordes de segurança em 2009. Nós também sabemos em que o Flash é a razão número um para crash nos Macs”

<http://www.apple.com/hotnews/thoughts-on-flash/>

CURRENT MEMBERS

As of 22 June 2012, the World Wide Web Consortium (W3C) has 412 members. [View the statistics about W3C Members](#) and view an [animation of current members](#).

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q
R | S | T | U | V | W | Y | Z

- [28msec Inc.](#)
- [Aalto University](#)
- [Academia Sinica](#)
- [ACCESS CO., LTD.](#)
- [ACORD Corporation](#)
- [Adesis Netlife S.L. — *testimonial*](#)
- [Adobe Systems Inc.](#)
- [AdTruth](#)
- [Afilias Limited](#)
- [Agencia para el Desarrollo del Gobierno de Gestión](#)
- [Electronica y la Sociedad de la Informacion y del Conocimiento \(AGESIC\)](#)
- [AgileDelta, Inc. — *testimonial*](#)
- [Akmin](#)
- [Alcatel-Lucent](#)
- [Algebraix Data Corporation](#)
- [Alisys Software](#)
- [Altova GmbH — *testimonial*](#)
- [Amadeus](#)
- [AME Info FZ LLC — *testimonial*](#)
- [ANBOTO — *testimonial*](#)
- [ANDANZA TECHNOLOGIES](#)
- [Apple, Inc.](#)
- [association BrailleNet — *testimonial*](#)
- [Astra Zeneca](#)
- [AT&T](#)
- [Atal Behari Vajpayee - Indian Institute of Information Technology and Management, Gwalior](#)



A web não morreu

Mas ela está mudando

Tim Berners-Lee: we need to re-decentralise the web

POLITICS / 06 FEBRUARY 14 / by LIAT CLARK



5606 shares
12 comments

Twenty-five years on from the web's inception, its creator has urged the public to re-engage with its original design: a decentralised internet that at its very core, remains open to all.

Speaking with Wired editor David Rowan at an event launching the magazine's **March issue**, Tim Berners-Lee said that although part of this is about keeping an eye on for-profit internet monopolies such as search engines and social networks, the greatest danger is the emergence of a balkanised web.



Tim Berners-Lee with David Rowan *Chris Woods / chrismwoods.com*

<http://www.wired.co.uk/news/archive/2014-02/06/tim-berners-lee-reclaim-the-web>

A web está evoluindo

O que esperar do futuro da Web?



O futuro da Web

que já é quase presente

[Todas](#)[Notícias](#)[Vídeos](#)[Imagens](#)[Shopping](#)[Mais ▾](#)[Ferramentas de pesquisa](#)

Stranger Things / Elenco

Winona Ryder
Joyce

Millie Brown
Eleven

Natalia Dyer
Nancy

David Harbour
Chief Hopper

Charlie Heaton
Jonathan

Noah Schnapp
Will

Matthew Modine
Dr. Martin Brenner

Cara Buono
Karen

G...
Dus...

Notícias em destaque



[De "Stranger Things": veja Millie Brown, Gaten Matarazzo e o restante do elenco na vida real!](#)

Purebreak Brasil - 2 dias atrás

De "Stranger Things": Gaten Matarazzo é tão fofo quanto o seu ... De

"Stranger Things": David Harbour é mais estiloso que o Chief Hopper, né?

De "Stranger Things": Natalia Dyer e Nancy Wheeler são ...

[Nova febre entre os amantes de séries, 'Stranger Things' recupera anos 1980 - 31/07/2016 - Ilustrada - Folha de S.Paulo](#)

Folha de S.Paulo - 2 dias atrás

[De "Stranger Things": Larissa Manoela, Chay Suede, Mel Maia e mais no elenco brasileiro da série!](#)

Purebreak Brasil - 1 dia atrás

[Mais notícias sobre elenco stranger things](#)

[5 coisas "estranhas" \(e fofas!\) sobre as crianças de "Stranger Things" ...](#)

[tvefamosos.uol.com.br/listas/5-coisas-estranhas-e-fofas-sobre-as-criancas-de-stranger-t... ▾](#)

5 dias atrás - Veja cenas da série "Stranger Things", da Netflix8 fotos ... Bobby Brown e Gaten Matarazzo em cena de "Stranger Things" Imagem: ... algo semelhante ao que acontece com o elenco de "Todo Mundo Odeia o Chris". ... Não foi difícil para Finn Wolfhard, o Mike, se ambientar à época de "Stranger Things".

Stranger Things

Seriado de ficção científica



Stranger Things é uma websérie americana dos gêneros suspense, ficção científica e terror criada pelos irmãos Matt e Ross Duffer e distribuída pela Netflix. [Wikipédia](#)

Emissora original: [Netflix](#)

Número De Episódios: 8

Número De Temporadas: 1

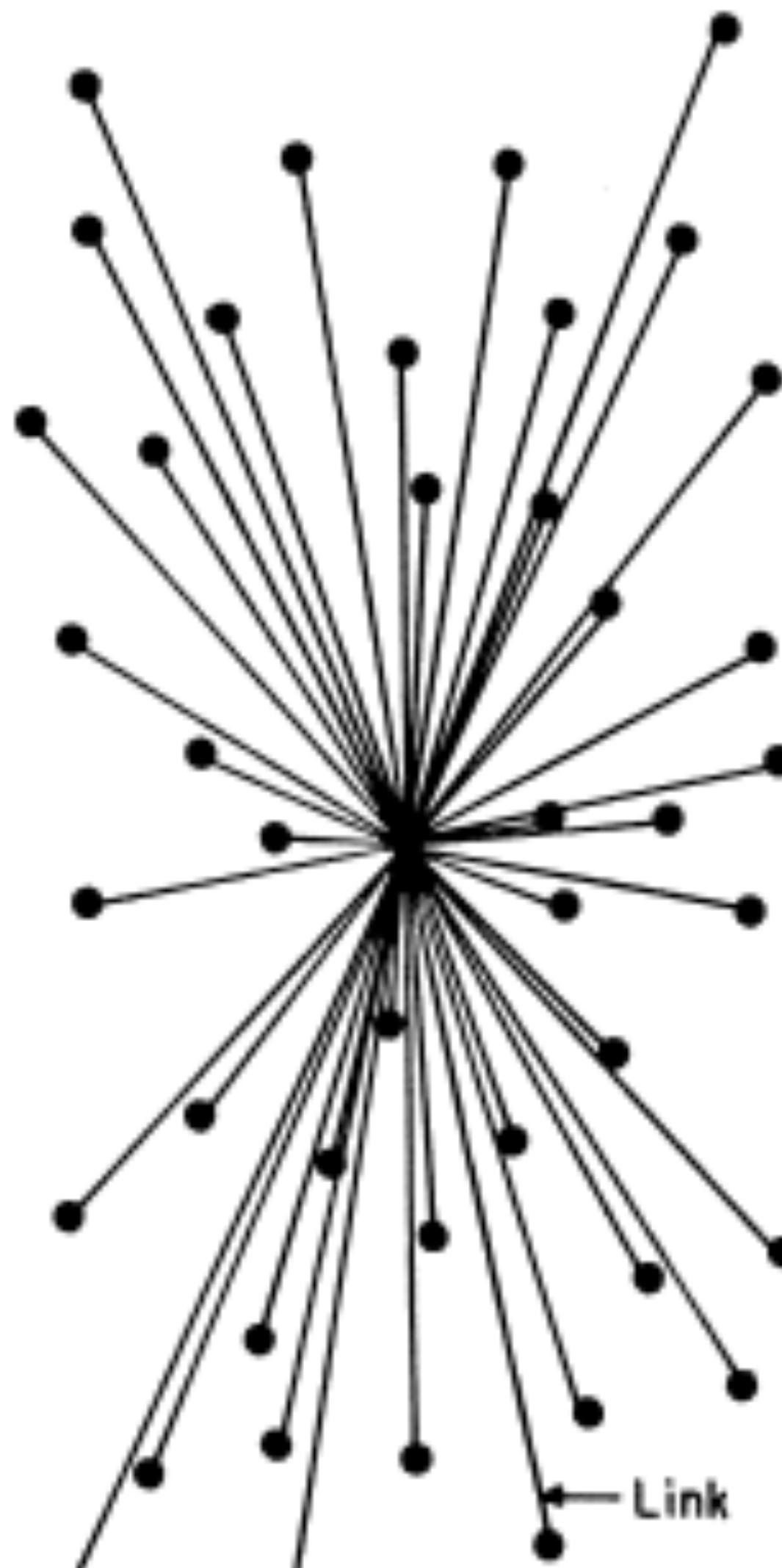
Idioma: Língua inglesa

Gêneros: Thriller, Ficção científica, Terror, Mistério, Supernatural fiction

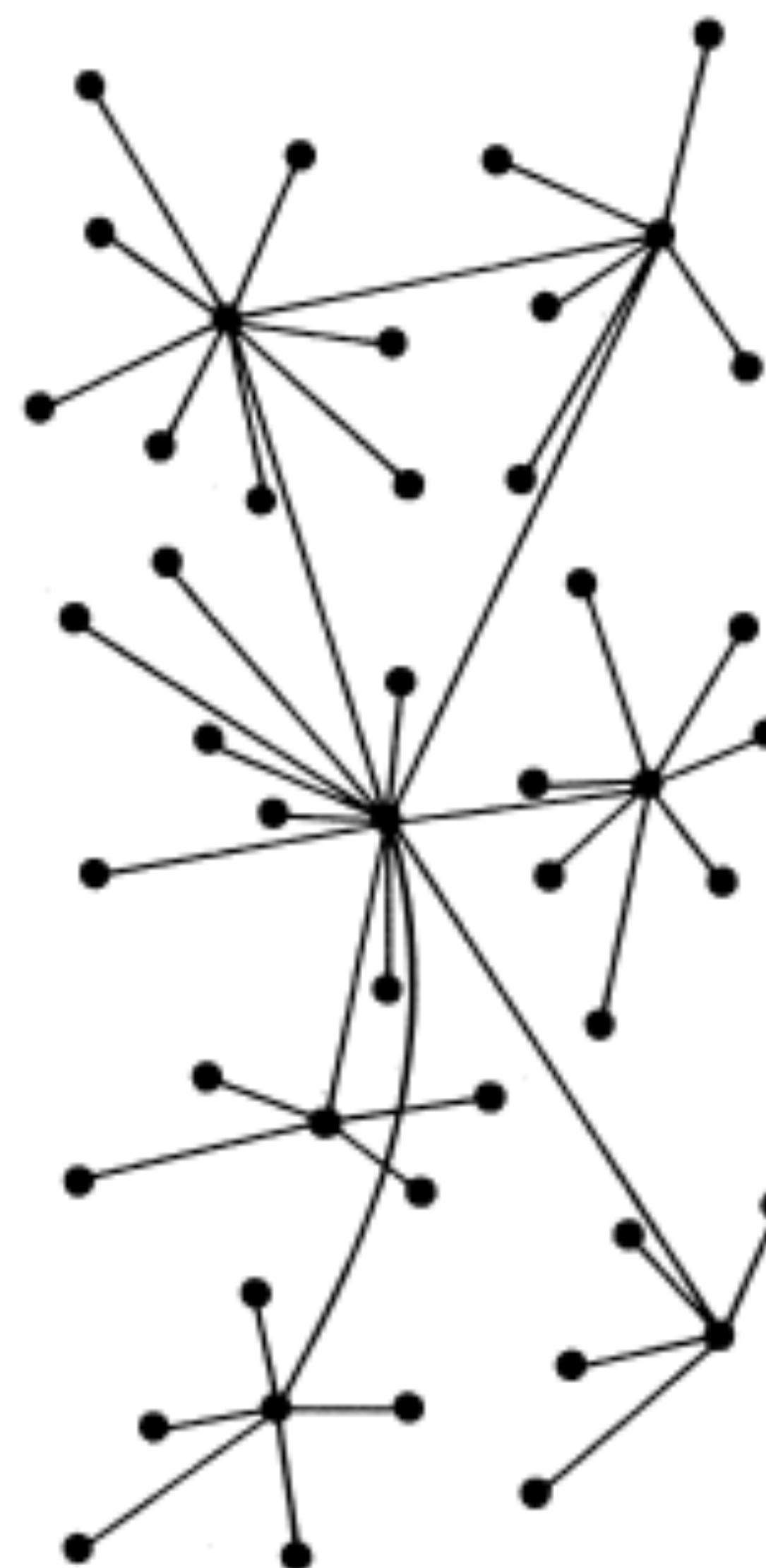
Pesquisas relacionadas



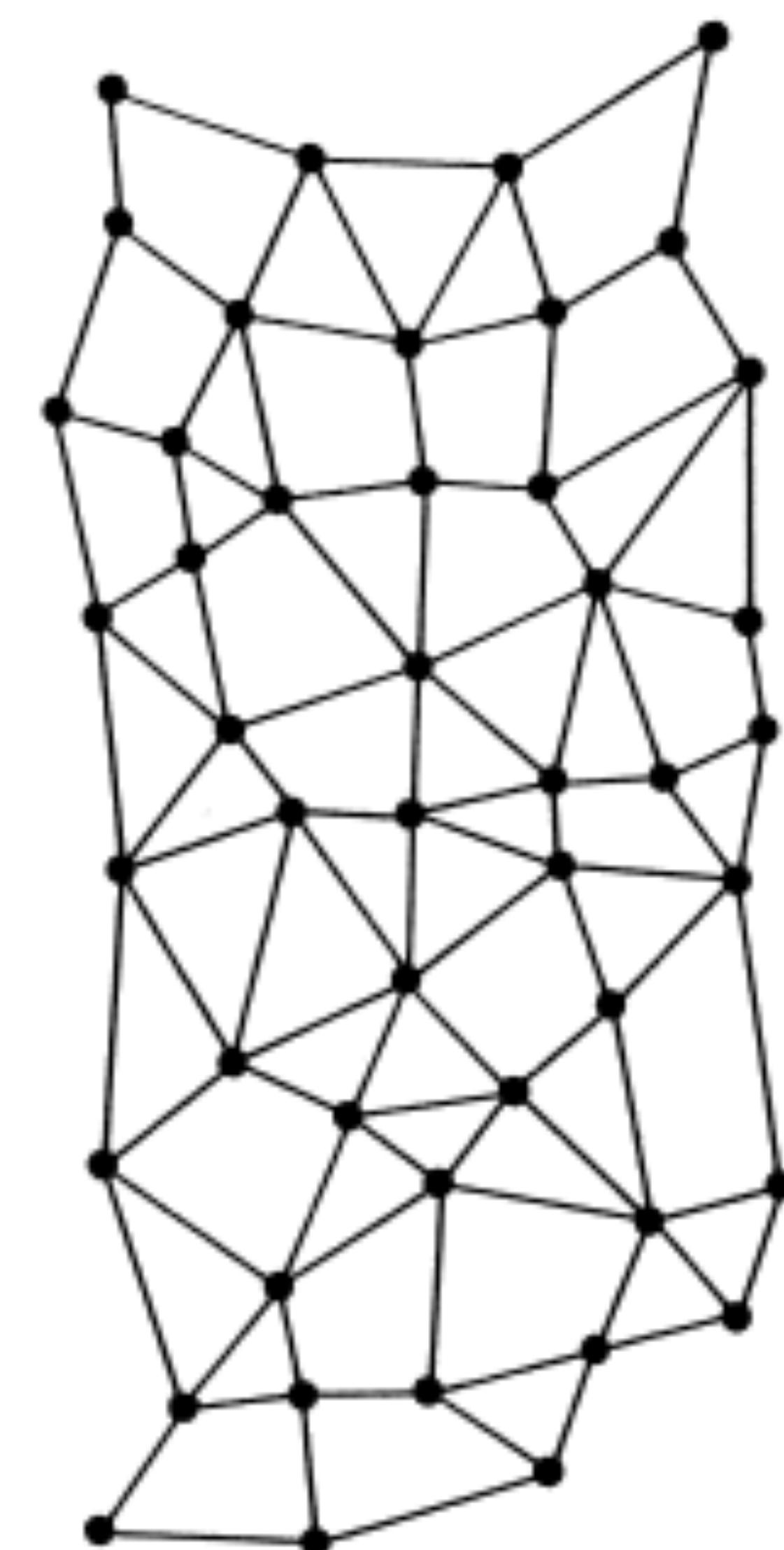




CENTRALIZED
(A)



DECENTRALIZED
(B)



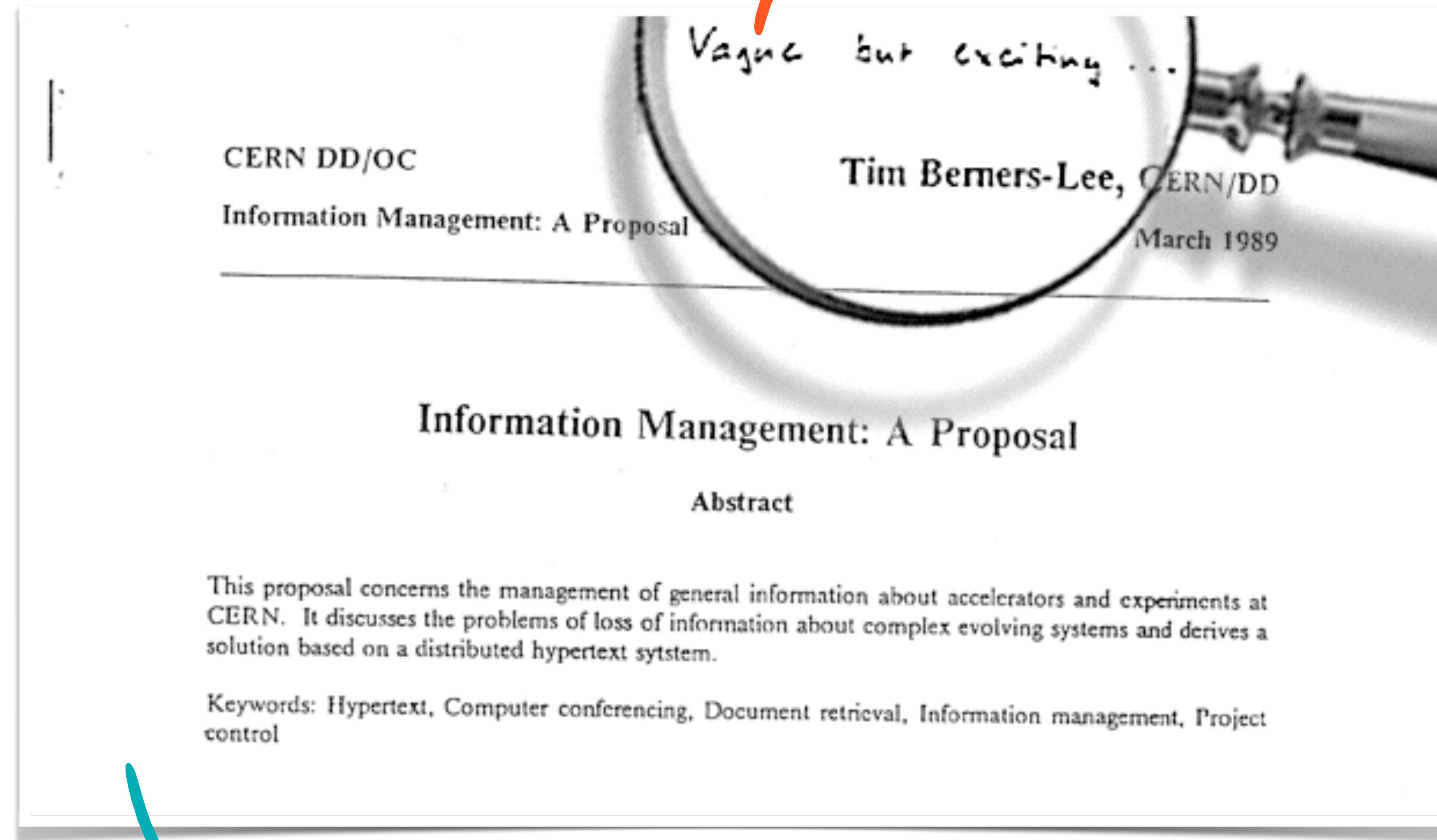
DISTRIBUTED
(C)

FIG. I – Centralized, Decentralized and Distributed Networks



Web Semântica

“vague but exciting”



a proposta da World Wide Web

Web

of



Documents

Data

Things

Web Semântica

Berners-Lee et al., 2001

“The Semantic Web aims to describe the meaning of information published on the Web to enable retrieval based on an accurate understanding of that information’s semantics”

(RODRÍGUEZ-ROCHA et al., 2015)

LONG READ | SECRETS OF THE MOST PRODUCTIVE PEOPLE

At Sundar Pichai's Google, AI Is Everything—And Everywhere

A year into his tenure as CEO of Google, the low-key company is, where it's going, and how it gets things



The wide scope represented by that lineup may suggest that Pichai isn't picking his competitive battles. But his key ambitions are bound by a core prediction: that the world is moving from the smartphone age into, in Pichai's phrase, an "AI-first" era, in which Google products will help people accomplish tasks in increasingly sophisticated, even anticipatory ways. The Pixel phones and Google Home, for instance, are the first devices with embedded support for Google Assistant, a rival to Apple's Siri and Amazon's Alexa that is designed not only to handle straightforward commands but also fuzzier requests such as "Play that

EXCLUSIVE: WARP DRIVE UNDERWATER • ARCTIC OIL VS. WILDLIFE

SCIENTIFIC AMERICAN

Get the Idea?
(TOMORROW'S WEB WILL)



PLUS:

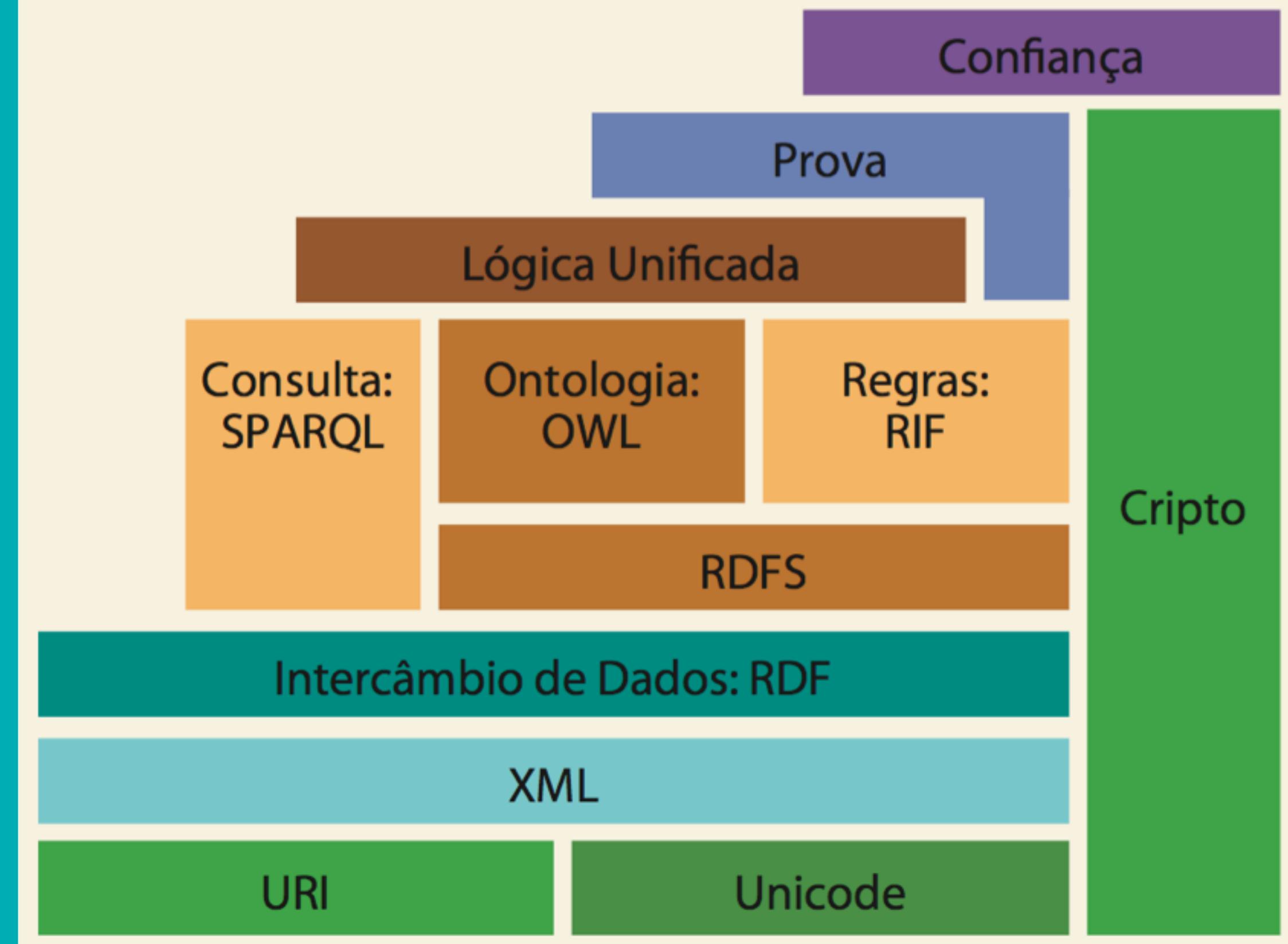
Antibiotics' Dim Future
Rorschach: A Waste of Ink
The Oldest Stars

MAY 2003
www.sciam.com \$4.95

~ 10 anos

... para vermos grandes aplicações

Interface do Usuário & Aplicações



Semantic Web Stack



RDF



sujeito

Demi Getschko

predicado

conhece

objeto

Vint Cerf



... a evolução da Web

Web

of



Documents

Data

Things

interoperabilidade semântica

Web Semântica e Processamento de Linguagem Natural na prática

Newton Calegari



PUC-SP

newton@nic.br

fb.com/newtoncalegari

twitter: @newtoncalegari