

CCT396, Fall 2011

Database Design and Implementation

Yuri Takhteyev
University of Toronto



This presentation is licensed under Creative Commons Attribution License, v. 3.0. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>. This presentation incorporates images from the Crystal Clear icon collection by Everaldo Coelho, available under LGPL from <http://everaldo.com/crystal/>.

Week 10

Relational Databases and Documents

CERTIFICATION OF LIVE BIRTH

STATE OF HAWAII
HONOLULU



DEPARTMENT OF HEALTH
HAWAII U.S.A.

CERTIFICATE NO. [REDACTED]

CHILD'S NAME
BARACK HUSSEIN OBAMA II

DATE OF BIRTH
August 4, 1961

HOUR OF BIRTH
7:24 PM

SEX
MALE

CITY, TOWN OR LOCATION OF BIRTH
HONOLULU

ISLAND OF BIRTH
OAHU

COUNTY OF BIRTH
HONOLULU

MOTHER'S MAIDEN NAME
STANLEY ANN DUNHAM

MOTHER'S RACE
CAUCASIAN

FATHER'S NAME
BARACK HUSSEIN OBAMA

FATHER'S RACE
AFRICAN

DATE FILED BY REGISTRAR
August 8, 1961

OHSM 1.1 (Rev. 11/01) LASER

This copy serves as prima facie evidence of the fact of birth in any court proceeding. [HRS 338-13(b), 338-19]

ANY ALTERATIONS INVALIDATE THIS CERTIFICATE

CHILD'S NAME

BARACK HUSSEIN OBAMA II

DATE OF BIRTH

August 4, 1961

HOUR OF BIRTH

7:24 PM

SEX

MALE

CITY, TOWN OR LOCATION OF BIRTH

HONOLULU

ISLAND OF BIRTH

OAHU

COUNTY OF BIRTH

HONOLULU

MOTHER'S MAIDEN NAME

STANLEY ANN DUNHAM

MOTHER'S RACE

CAUCASIAN

FATHER'S NAME

BARACK HUSSEIN OBAMA

FATHER'S RACE

AFRICAN

DATE FILED BY REGISTRAR

August 8, 1961

Digital

```
name: Barack Hussein Obama II  
date_of_birth: August 4, 1961  
sex: male  
location: Honolulu  
mother: Stanley Ann Dunham  
father: Barack Hussein Obama  
hour_of_birth: 7:24 pm  
...
```

can be digitally signed

Full Text Search

Matching sub-strings

Can be done with LIKE

Bag-of-words search

Like doing LIKE for every term – very slow without a proper index!

Natural language search

Stop words ("the", "of", "to", "end")

Stemming ("monkey" vs "monkeys")

Synonyms ("monkey" vs "simian" vs "Affe")

Ranking

"tf-idf", "Okapi BM25", probabilistic models

Can be useful. SQL support is coming.

Full Text in MySQL

```
use common;  
select num from analects  
  where match(contents)  
         against("chariots business");  
select num from analects  
  where match(contents)  
         against("+chariots -business"  
                 in boolean mode);
```

This requires setting up a "fulltext" index first and use "MyISAM" engine – more on those in two weeks.

1-5 A

Depot van

FAKTUUR

N^o 918

STANDARD

Antwerpen, den 4 4 1947

AMERICAN PETROLEUM COMPANY, NAAMLOOZE VENNOOTSCHAP

OLIËN en VETTEN

Maatschappelijke Zetel: 101, Frankrijklei

ANTWERPEN

Telefonen:
ANTWERPEN: 210.20
(8 lijnen)Postcheckrekening Nr 368
Handelsregister van
Antwerpen: 680Telegram-Adres:
PETROLEA - ANTWERPEN

Esso

Essolube

DEBET voor verkoop en levering van:

PRODUKT en KWALITEIT	Hoeveelheid in liters	Prijs per liter	BEDRAG	TOTALEN
Essolube SAP 40 IX 60 TP	60	16.50	Fr. 990	
		16.35	981	
			64.60	1026
TOTAAL				Fr. 1026
Saldo te betalen aan den klient.				Fr. 1026

Waarborg der pakgoederen die
echter eigendom der klant blijven.

GOEDEREN ONTVANGEN

Hoeveelheid juist bevonden,
DE KLIENT.De verkoop onzer produktien is gedaan aan de
algemeene voorwaarden aangeduid op de achterzijde
dezer faktuur.

KOMPTANTE WAARDE.

Wij verzoeken onze klanten, ten einde hunne ver-
antwoordelijkheid te dekken, na te zien of de flakke
regels wel op de rekening geplakt zijn door den
voorman.

BETALING ONTVANGEN

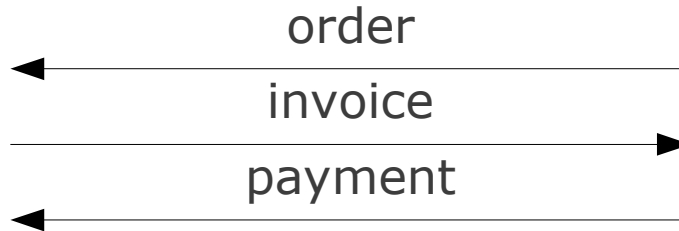
De depothouder

De voorman

D & P 54129

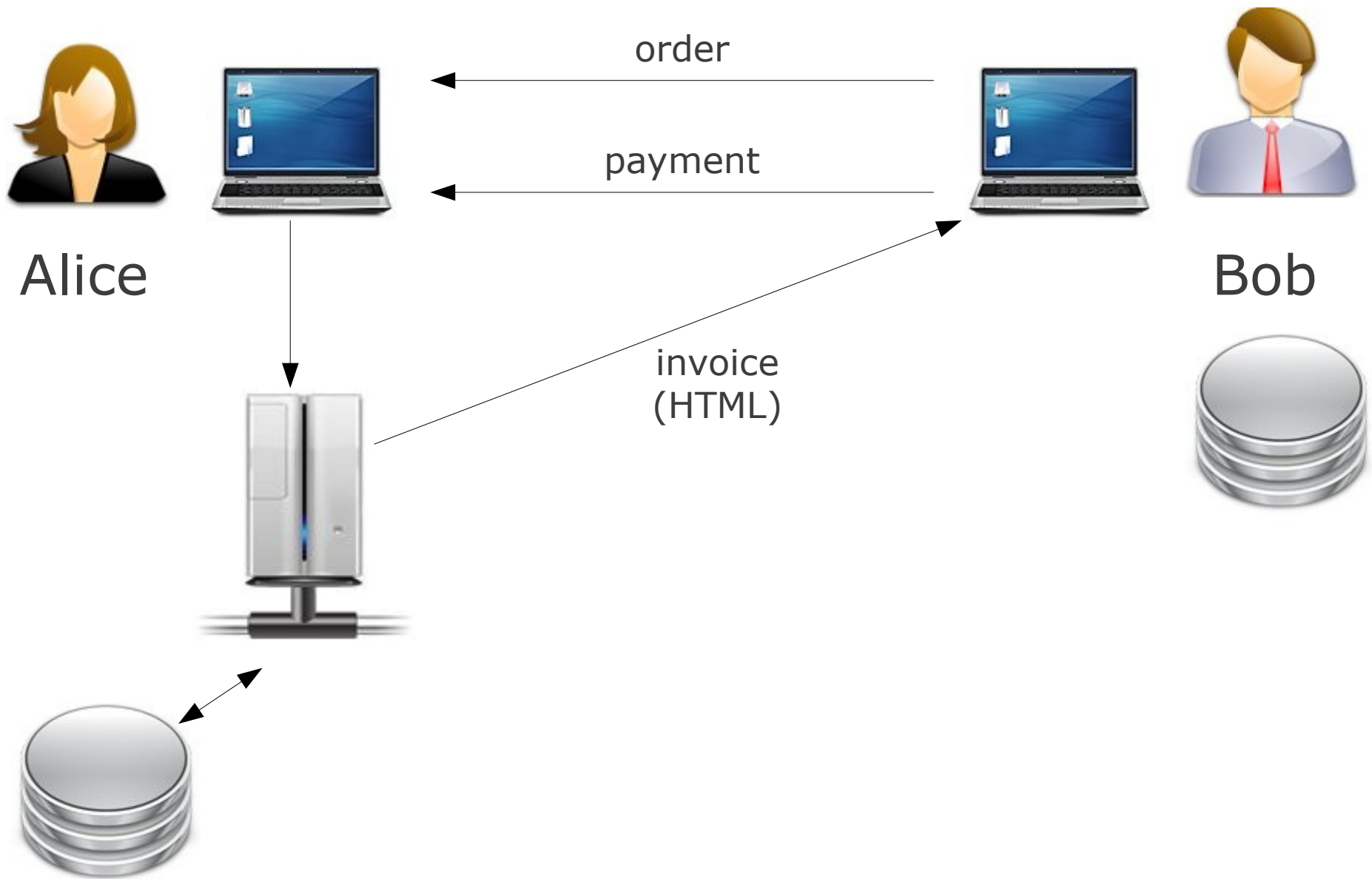


Alice



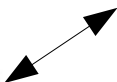
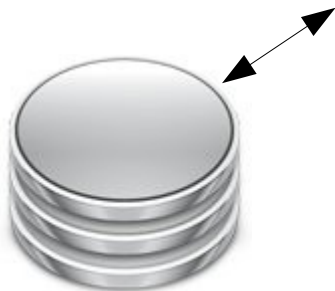
Bob



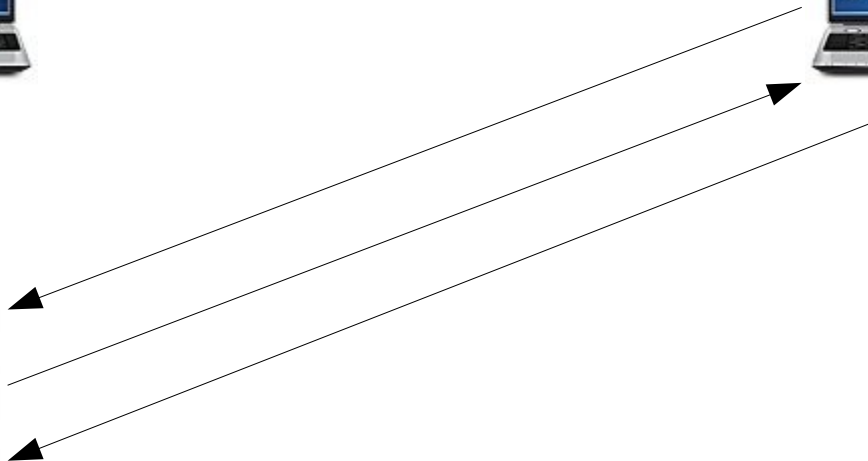


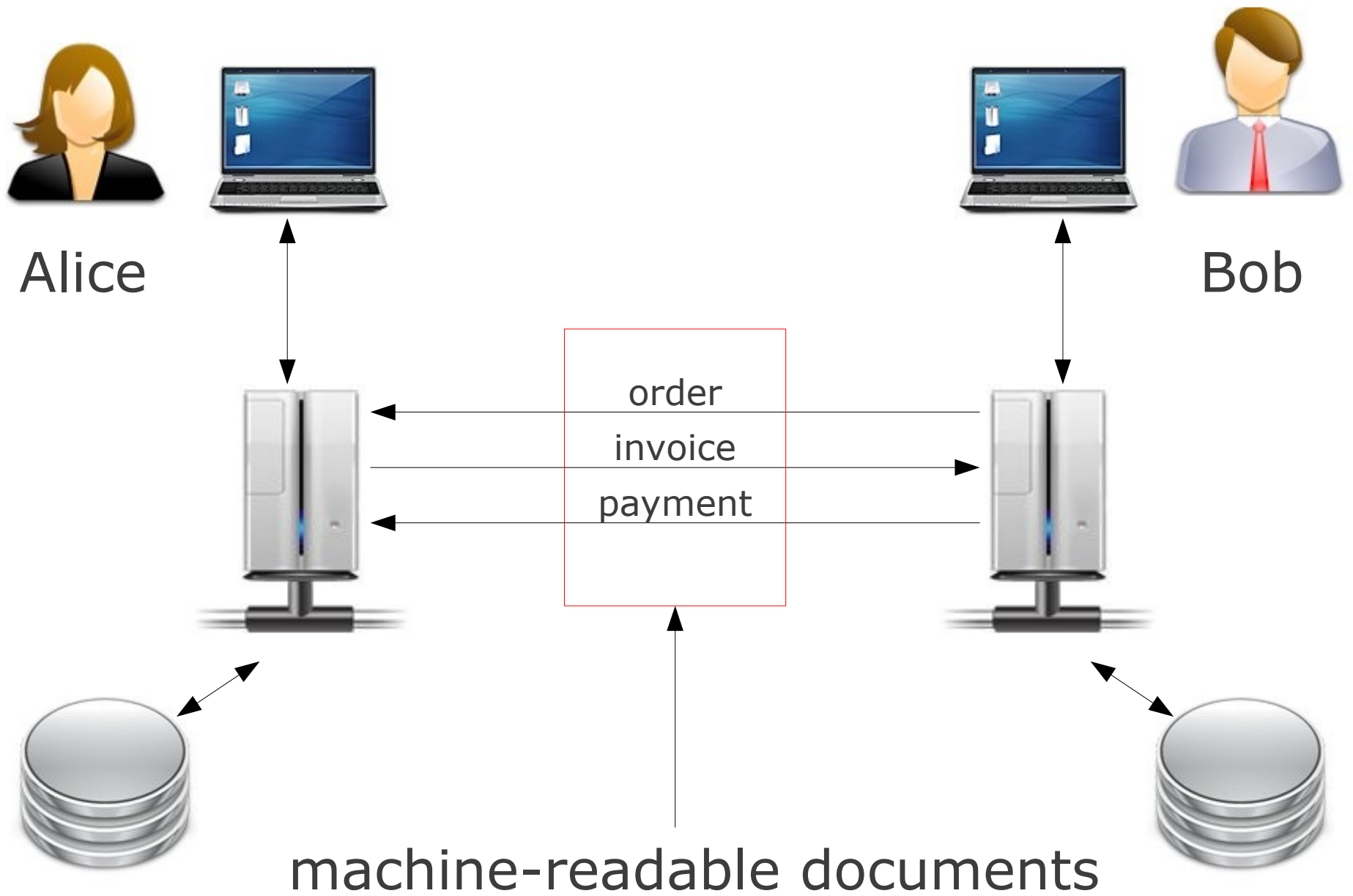


Alice



Bob





XML

```
<?xml version="1.0" encoding="UTF-8"?>
<Invoice>
  <ID>9834562</ID>
  <IssueDate>2003-03-14</IssueDate>
  <Buyer>
    <Name>Bills Microdevices</Name>
    <StreetName>Spring St.</StreetName>
    <BuildingNumber>413</BuildingNumber>
    <CityName>Elgin</CityName>
    <PostalZone>60123</PostalZone>
    <CountrySubentityCode>IL</CountrySubentityCode>
  </Buyer>
  <Seller>
    <Name>Joes Office Supply</Name>
    <StreetName>Lakeshore Dr</StreetName>
    <BuildingNumber>32 W.</BuildingNumber>
    <CityName>Chicago</CityName>
    <PostalZone>60022</PostalZone>
    <CountrySubentityCode>IL</CountrySubentityCode>
  </Seller>
```


XML

```
<Total>
  <TotalAmount amountCurrencyID="USD">479.50</TotalAmount>
  <TaxExclusive amountCurrencyID="USD">479.50</TaxExclusive>
  <TaxInclusive amountCurrencyID="USD">527.45</TaxInclusive>
</Total>
<InvoiceLine>
  <ID>1</ID>
  <InvoicedQuantity unitCode="PKG">5</InvoicedQuantity>
  <Amount amountCurrencyID="USD">12.50</Amount>
  <Item>
    <Description>Pencils, box #2 red</Description>
    <SellersItemIdentification>
      <ID>32145-12</ID>
    </SellersItemIdentification>
    <BasePrice>
      <PriceAmount amountCurrencyID="USD">2.50</PriceAmount>
    </BasePrice>
  </Item>
</InvoiceLine>
<InvoiceLine>
```

XML Anatomy

```
<InvoiceLine>
  <ID>1</ID>
  <InvoicedQuantity unitCode="PKG">5</InvoicedQuantity>
  <Amount amountCurrencyID="USD">12.50</Amount>
  <Item>
    <Description>Pencils, box #2 red</Description>
    <SellersItemIdentification>
      <ID>32145-12</ID>
    </SellersItemIdentification>
    <BasePrice>
      <PriceAmount amountCurrencyID="USD">2.50</PriceAmount>
    </BasePrice>
  </Item>
</InvoiceLine>
```

XML Anatomy

```
<InvoiceLine>
  <ID>1</ID>
  <InvoicedQuantity unitCode="PKG">5</InvoicedQuantity>
  <Amount amountCurrencyID="USD">12.50</Amount>
  <Item>
    <Description>Pencils, box #2 red</Description>
    <SellersItemIdentification>
      <ID>32145-12</ID>
    </SellersItemIdentification>
    <BasePrice>
      <PriceAmount amountCurrencyID="USD">2.50</PriceAmount>
    </BasePrice>
  </Item>
</InvoiceLine>
```

Validation

Syntactic: form

`<Amount amountCurrencyID="USD">12.50</Amount`

Semantic: meaning

`<Amount amountCurrencyID="USD">Male</Amount>`

Namespaces

```
<?xml version="1.0" encoding="UTF-8"?>
<Invoice xmlns:cbc="urn:oasis:names:specification:ubl:sc...">
  <ID>9834562</ID>
  <cbc:IssueDate>2003-03-14</cbc:IssueDate>
  <TaxPointDate>2003-03-14</TaxPointDate>
  <cac:OrderReference>
    <cac:BuyersID>20031234-1</cac:BuyersID>
    <cac:SellersID>154135798</cac:SellersID>
    <cbc:IssueDate>2003-02-02</cbc:IssueDate>
  </cac:OrderReference>
  <cac:BuyerParty>
    <cac:Party>
      <cac:PartyName>
        <cbc:Name>Bills Microdevices</cbc:Name>
      </cac:PartyName>
      <cac:Address>
        <cbc:StreetName>Spring St.</cbc:StreetName>
        <cbc:BuildingNumber>413</cbc:BuildingNumber>
        <cbc:CityName>Elgin</cbc:CityName>
        <cbc:PostalZone>60123</cbc:PostalZone>
        <cac:CountrySubentityCode>IL</cac:CountrySubentityCode>
```

XML and RDBMS

Externalization:

export, import, messaging

Internal use:

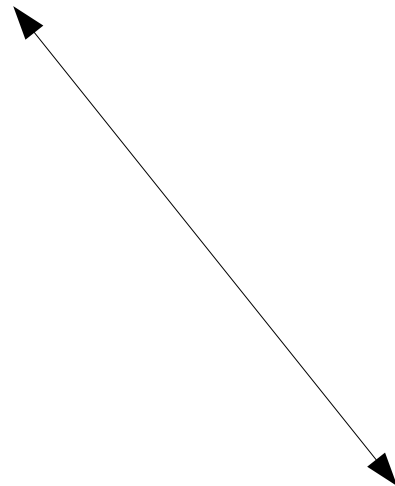
XML storage and retrieval
(less common for now)



Alice



Bob



HTML

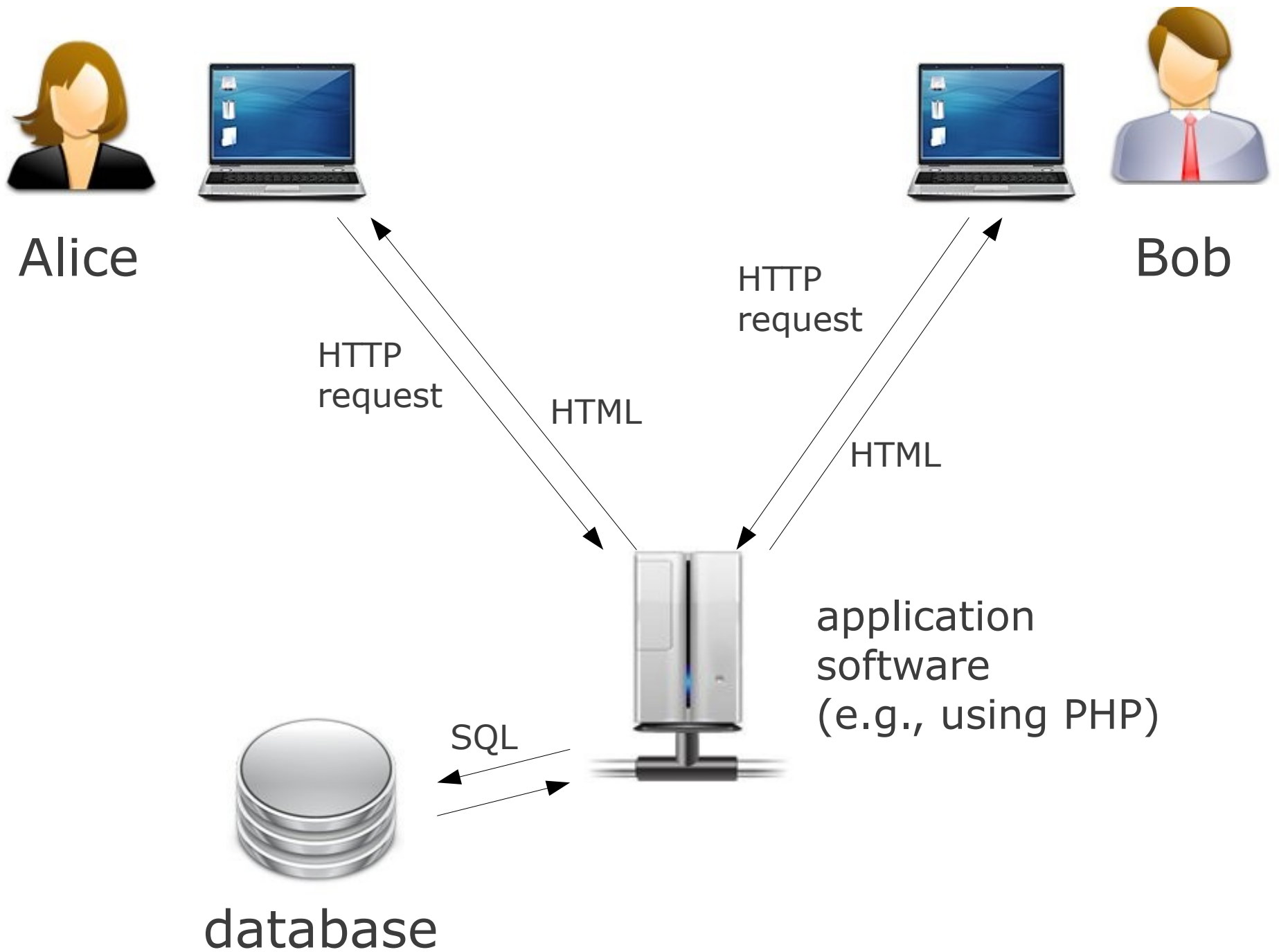


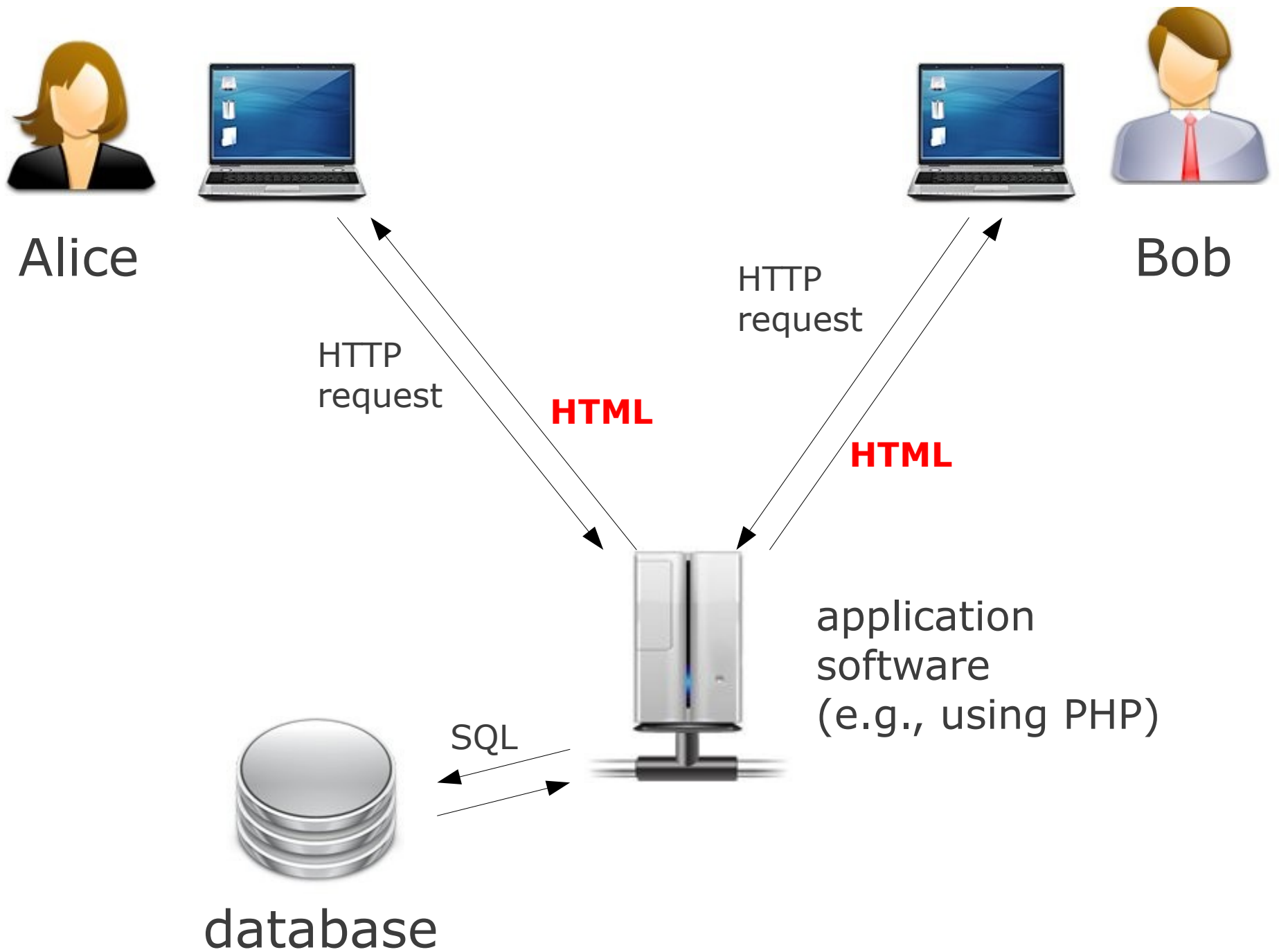
application
software
(e.g., using PHP)

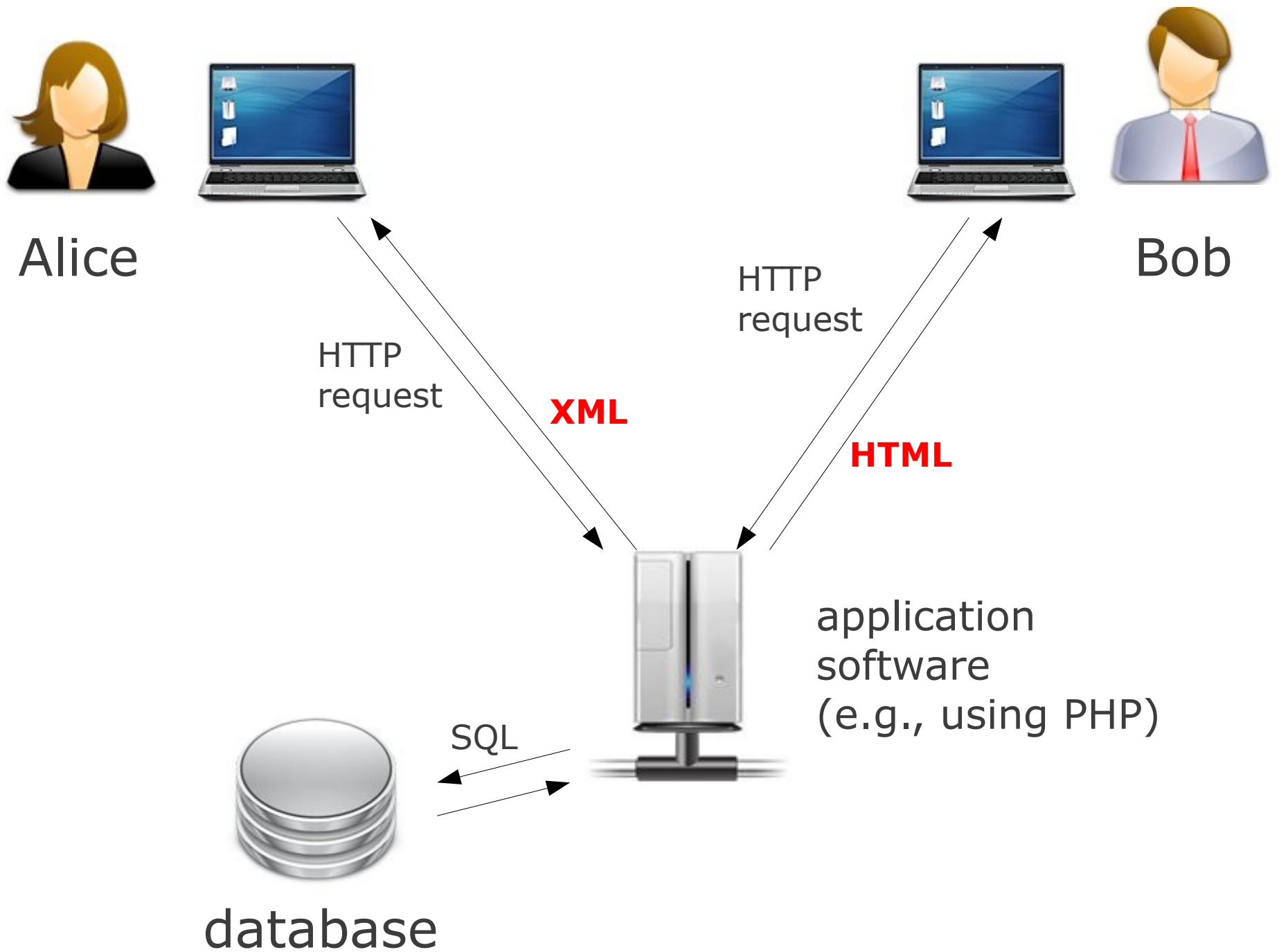


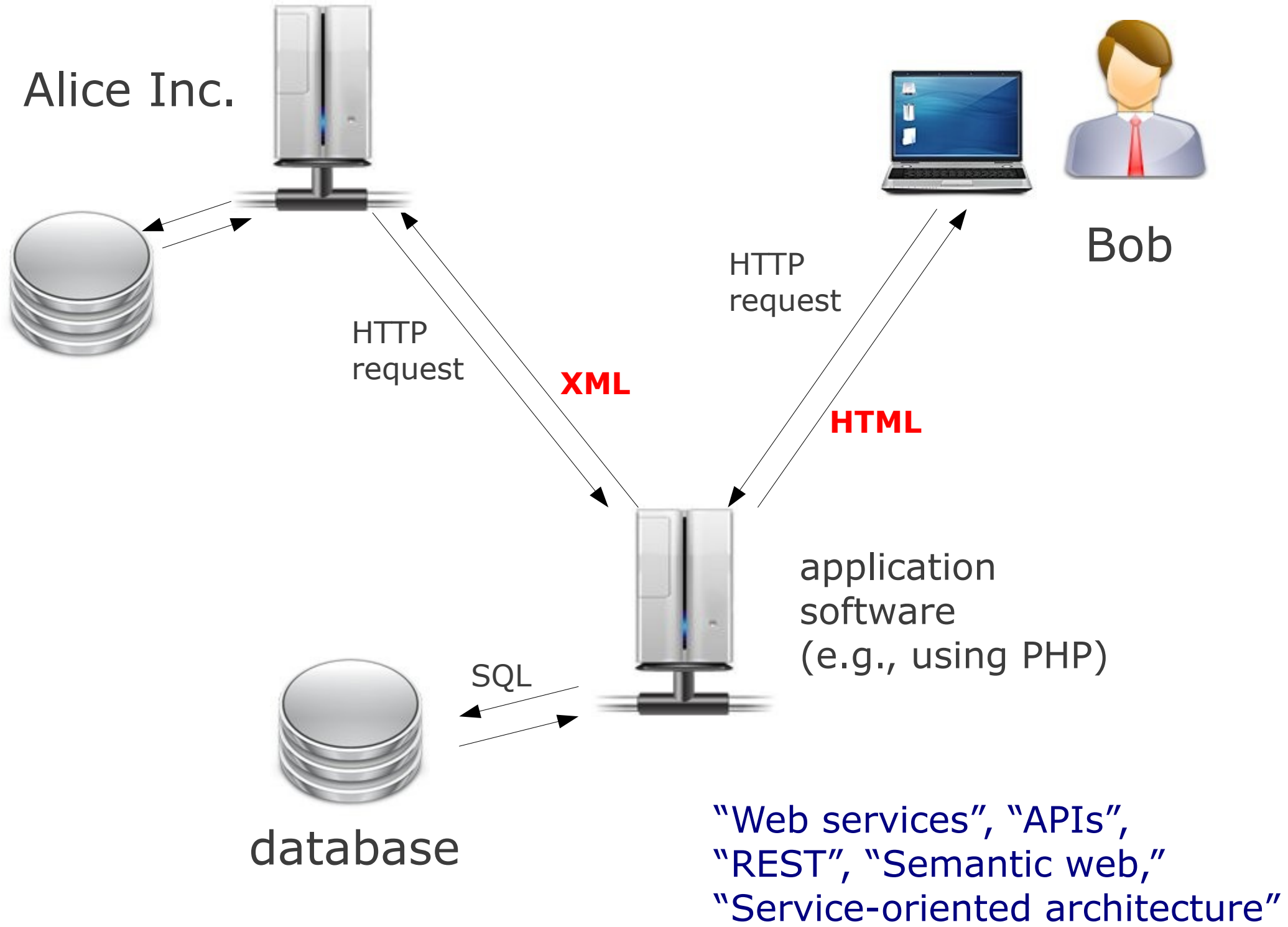
database











Examples

Twitter:

http://twitter.com/public_timeline

vs.

http://api.twitter.com/1/statuses/public_timeline.xml

Recent Public Tweets

What everyone on Twitter is talking about!



[REDACTED] mai nap.. regel miután felekeltem, ittam kávét...
... suliba beérek csöngetésre..*faszaa.* első óra offő.....

<http://yamm.hu/l/4vH>

less than 5 seconds ago via Yamm!



less than 10 seconds ago via ÜberTwitter



ShortmanScore Kanyes most certainly a genius but i truly do not care 4 his wave, the people around him keep him on a pedestal, but he makes good shit

less than 20 seconds ago via ÜberTwitter



pais_libre Con la expropiación de la Owens-Illinois se dejará de proveerle envases de vidrio a [#Polar](#)...que juego tan sucio! [#Venezuela](#)

half a minute ago via TweetDeck



Kelincimadu Cemburu loo cin ? :p RT [@NoegrohoBayu](#): Amin RT [@Kelincimadu](#): Ya allah semogaaa dia nggak kenapa2 , masih peduli kok ama dia !!!

half a minute ago via ÜberTwitter



studiovideochat Radio Etno Romania .. inregistrare domenii gratis ,gazduire radio ... <http://bit.ly/9e31hw>

half a minute ago via angajari videochat

Create Your Account

Join!

Already using Twitter from your phone? [Click here.](#)

Select Language ... ▼

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
  <head>
    <meta http-equiv="X-UA-Compatible" content="IE=8">

    <script type="text/javascript">
//<![CDATA[
(function(g){var a=location.href.split("#")[1];if(a){window.location.hash = "";g.location.pathname = g.HBR = a.replace(/^(^/)/,"/$1");}})(window);
//]]>
</script>
    <script type="text/javascript" charset="utf-8">
      if (!twtr) {
        var twtr = {}
      }

      // Benchmarking load time.
      // twtr.timeTillReadyUnique = '1288106691-86330-54446';
      // twtr.timeTillReadyStart = new Date().getTime();
    </script>

    <script type="text/javascript">
//<![CDATA[
var page={};var onCondition=function(D,C,A,B){D=D;A=A?Math.min(A,5):5;B=B||100;if(D()){C()}else{if(A>1){setTimeout(function(){onCondition(D,C,A-1,B)},
//]]>
</script>
    <meta content="text/html; charset=utf-8" http-equiv="Content-Type" />
    <meta content="en-us" http-equiv="Content-Language" />
    <meta content="Twitter is without a doubt the best way to share and discover what is happening right now." name="description" />
    <meta content="no" http-equiv="imagetoolbar" />
    <meta content="width = 780" name="viewport" />
    <meta content="4FTTxY4uvo0RZTMQqIyhhl8HsepyJOctQ+XT0ulzsfE=" name="verify-v1" />
    <meta content="1" name="page" />
    <meta content="NOODP" name="robots" />
    <meta content="n" name="session-logged-in" />
    <title id="page_title">Twitter / Public Timeline</title>
    <link href="http://a1.twimg.com/a/1288039940/images/twitter_57.png" rel="apple-touch-icon" />
    <link href="/oexchange.xrd" rel="http://oexchange.org/spec/0.8/rel/related-target" type="application/xrd+xml" />
    <link href="http://a1.twimg.com/a/1288039940/images/favicon.ico" rel="shortcut icon" type="image/x-icon" />

    <link href="http://a1.twimg.com/a/1288039940/stylesheets/twitter.css?1288056348" media="screen" rel="stylesheet" type="text/css" />
    <link href="http://a1.twimg.com/a/1288039940/stylesheets/geo.css?1288056348" media="screen" rel="stylesheet" type="text/css" />
    <link href="http://a2.twimg.com/a/1288039940/stylesheets/buttons_new.css?1288056348" media="screen" rel="stylesheet" type="text/css" />
    <style type="text/css">

      body { background: #CODEED url('http://a3.twimg.com/a/1288039940/images/bg-clouds.png') repeat-x; }

```



```
<div class="content-bubble-arrow"></div>
```

```
<table cellpadding="0" class="columns">
  <tbody>
    <tr>
      <td id="content" class="round-left column">
        <div class="wrapper">
```

```
<div class="tab">
  <h2 class="timeline-header"> Recent Public Tweets</h2>
  <div class="badgebox">
    <h3 class="timeline-subheader">What everyone on Twitter is talking about!</h3>
  </div>
  <ol id='timeline' class='statuses'>
    <li class="hentry u- status" id="status_28800833000"

```

```
<span class="thumb vcard author"><a href="http://twitter.com/" class="tweet-url profile-pic url"><img alt="alekszaa." class="photo fn" height="40px" data-bbox="100 500 150 540" /></a></span>
<span class="status-content">
  <strong><a href="http://twitter.com/" class="tweet-url screen-name">alekszaa_</a></strong>

  <span class="entry-content">mai nap..
```

```
regel miután felekeltem, ittam kávét... ... s...liba beérek csöngetésre..*faszaa.* első óra offő..... <a href="http://yamm.hu/l/4vH" class="tweet-url screen-name">yamm!</a></span>
```

```
<span class="meta entry-meta" data='{}'>
<a class="entry-date" rel="bookmark" href="http://twitter.com/status/28800833000">
  <span class="published timestamp" data="{time:'Tue Oct 26 15:24:49 +0000 2010'}">less than 5 seconds ago</span></a>
<span>via <a href="http://yamm.hu" rel="nofollow">Yamm!</a></span>

</span>
```

```
<ul class="meta-data clearfix">
</ul>
</span>
</li>
```

```
<li class="hentry u-williedkang status" id="status_28800825000">
  <span class="thumb vcard author"><a href="http://twitter.com/williedkang" class="tweet-url profile-pic url"><img alt="Willie D. " class="photo fn" height="40px" data-bbox="100 880 150 920" /></a></span>
  <span class="status-content">
    <strong><a href="http://twitter.com/williedkang" class="tweet-url screen-name">williedkang</a></strong>
```

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<statuses type="array">
```

```
<status>
```

```
  <created_at>Tue Oct 26 15:24:49 +0000 2010</created_at>
```

```
  <id>28800833000</id>
```

```
  <text>mai nap..
```

```
regel miut&#225;n felekeltem, ittam k&#225;v&#233;t...
```

```
... suliba be&#233;rek cs&#246;nget&#233;sre..*faszaa.* els&#337; &#243;ra off&#337;..
```

```
  <source>&lt;a href=&quot;http://yamm.hu&quot; rel=&quot;nofollow&quot;&gt;Yamm!&lt;/a&gt;</source>
```

```
  <truncated>>false</truncated>
```

```
  <favorited>>false</favorited>
```

```
  <in_reply_to_status_id></in_reply_to_status_id>
```

```
  <in_reply_to_user_id></in_reply_to_user_id>
```

```
  <in_reply_to_screen_name></in_reply_to_screen_name>
```

```
  <retweet_count></retweet_count>
```

```
  <retweeted>>false</retweeted>
```

```
  <user>
```

```
    <id>141942404</id>
```

```
    <name></name>
```

```
    <screen_name></screen_name>
```

```
    <location>r&#233;pcelak.</location>
```

```
    <description>meow.:3</description>
```

```
    <profile_image_url>http://a2.twimg.com/profile_images/1146245330/Photo4915_normal.jpg</profile_image_url>
```

```
    <url></url>
```

```
    <protected>>false</protected>
```

```
    <followers_count>27</followers_count>
```

```
    <profile_background_color>CODEED</profile_background_color>
```

```
    <profile_text_color>333333</profile_text_color>
```

```
    <profile_link_color>0084B4</profile_link_color>
```

```
    <profile_sidebar_fill_color>DDEEF6</profile_sidebar_fill_color>
```

```
    <profile_sidebar_border_color>CODEED</profile_sidebar_border_color>
```

```
    <friends_count>43</friends_count>
```

```
    <created_at>Sun May 09 13:34:31 +0000 2010</created_at>
```

```
    <favourites_count>11</favourites_count>
```

```
    <utc_offset></utc_offset>
```

```
    <time_zone></time_zone>
```

```
    <profile_background_image_url>http://a1.twimg.com/profile_background_images/159997094/0f1fbafa72cff199e075f02bda52ac0d_21006592.jpg</profile_backg
```

```
    <profile_background_tile>true</profile_background_tile>
```

```
    <profile_use_background_image>true</profile_use_background_image>
```

```
    <notifications></notifications>
```

```
    <geo_enabled>>false</geo_enabled>
```

```
    <verified>>false</verified>
```

```
    <following></following>
```

```
    <statuses_count>540</statuses_count>
```

```
    <lang>en</lang>
```

```
    <contributors_enabled>>false</contributors_enabled>
```

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<statuses type="array">
```

```
<status>
```

```
<created_at>Tue Oct 26 15:24:49 +0000 2010</created_at>
```

```
<id>28800833000</id>
```

```
<text>mai nap..
```

regel miután felekeltem, ittam kávét... ... suliba beérek csönget&

```
<source>&lt;a href=&quot;http://yamm.hu&quot; rel=&quot;nofollow&quot;&gt;Yamm!&lt;/a&gt;</source>
```

```
<truncated>>false</truncated>
```

```
<favorited>>false</favorited>
```

```
<in_reply_to_status_id></in_reply_to_status_id>
```

```
<in_reply_to_user_id></in_reply_to_user_id>
```

```
<in_reply_to_screen_name></in_reply_to_screen_name>
```

```
<retweet_count></retweet_count>
```

```
<retweeted>>false</retweeted>
```

```
<user>
```

```
<id>141942404</id>
```

```
<name>[REDACTED]</name>
```

```
<screen_name>[REDACTED]</screen_name>
```

```
<location>r&#233;pcelak.</location>
```

```
<description>meow.:3</description>
```

```
<profile_image_url>http://a2.twimg.com/profile_images/1146245330/Photo4915_normal.jpg</profile_image_url>
```

```
<url></url>
```

```
<protected>>false</protected>
```

```
<followers_count>27</followers_count>
```

```
<profile_background_color>CODEED</profile_background_color>
```

```
<profile_text_color>333333</profile_text_color>
```

```
<profile_link_color>0084B4</profile_link_color>
```

```
<profile_sidebar_fill_color>DDEEF6</profile_sidebar_fill_color>
```

```
<profile_sidebar_border_color>CODEED</profile_sidebar_border_color>
```

```
<friends_count>43</friends_count>
```

```
<created_at>Sun May 09 13:34:31 +0000 2010</created_at>
```

```
<favourites_count>11</favourites_count>
```

```
<utc_offset></utc_offset>
```

Examples

Yahoo! Geocoding API:

[http://maps.yahoo.com/#&q1=3359+Mississauga+Rd+N.
%2C+Mississauga%2C+ON](http://maps.yahoo.com/#&q1=3359+Mississauga+Rd+N.%2C+Mississauga%2C+ON)

vs.

[http://where.yahooapis.com/geocode?
q=3359+Mississauga+Rd+N,+Mississauga,+ON](http://where.yahooapis.com/geocode?q=3359+Mississauga+Rd+N,+Mississauga,+ON)



New User? Register | Sign In | Help

Trending: Miniskirt Ban

Yahoo!

YAHOO! LOCAL
Maps

Search

GET MAP AND DIRECTIONS

Clear

A 3359 Mississauga Rd N., Mississauga, ON

B Address, City, State

Reverse Directions | Round-Trip

Go

FIND A BUSINESS ON THE MAP

Clear

Find Restaurants, Hotels...

Search

SPONSOR RESULTS

[Cabo San Lucas Villas](#)

Choose from over 140 private luxury villas in Cabo San Lucas, Mexico.

www.cabovillas.com

[cabo san lucas](#)

\$89 Beachfront Luxury Inclusive! 99 Hr Sale, Book Los Cabos Today

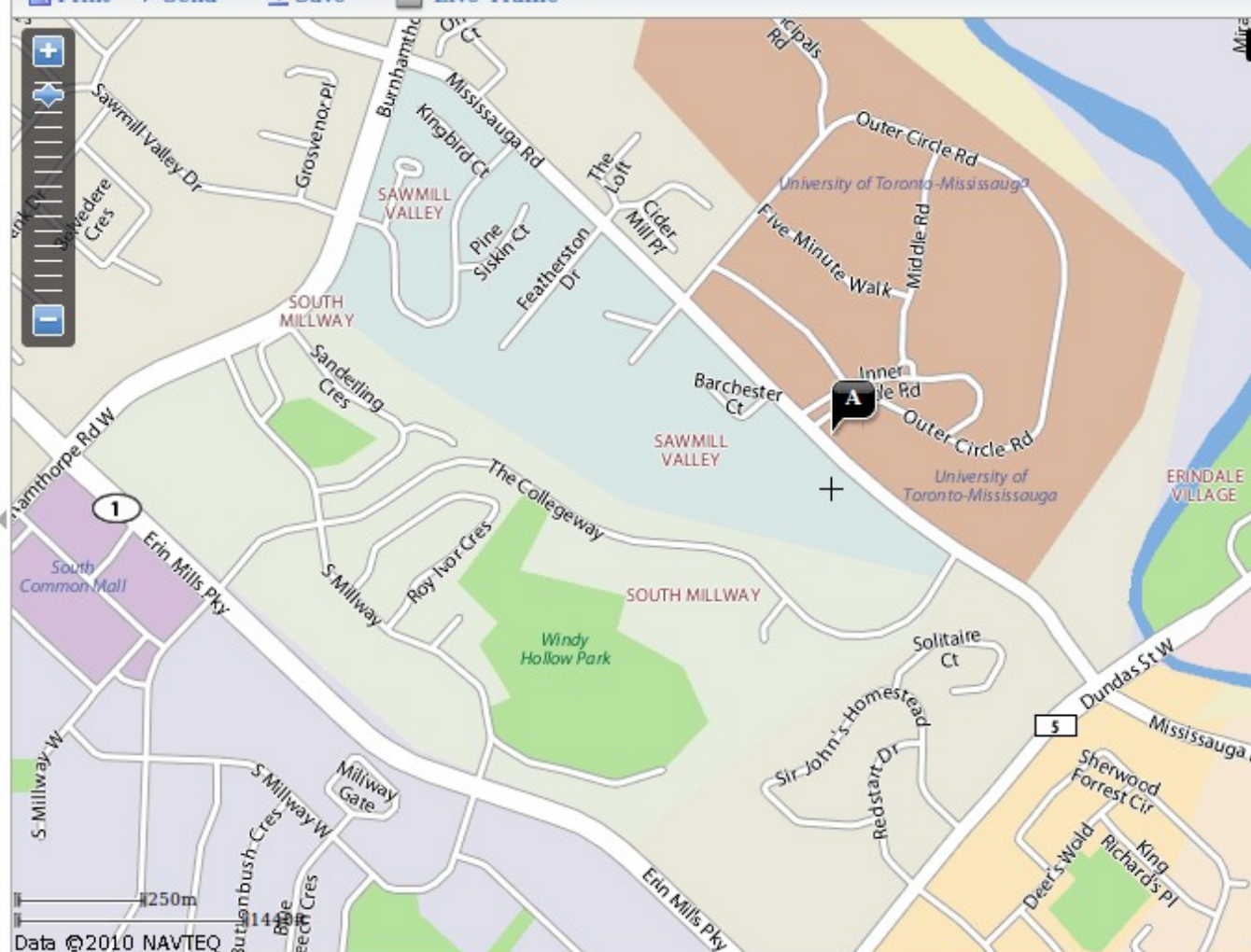
BookIt.com/LosCabos

[Find Maps & Directions](#)

Is There A Road Trip In Your Future? Create Custom Directions And Maps

www.Local.com/Maps

Print Send Save Live Traffic




```
<?xml version="1.0" encoding="UTF-8"?>
<ResultSet version="1.0"><Error>0</Error><ErrorMessage>No
error</ErrorMessage><Locale>us_US</Locale><Quality>87</Qu
ality><Found>1</Found><Result><quality>85</quality><latit
ude>43.546647</latitude><longitude>-
79.665735</longitude><offsetlat>43.546733</offsetlat><offs
etlon>-
79.665611</offsetlon><radius>500</radius><name></name><li
ne1>3359 Mississauga Rd</line1><line2>Mississauga, ON
L5L</line2><line3></line3><line4>Canada</line4><house>335
9</house><street>Mississauga
Rd</street><xstreet></xstreet><unittype></unittype><unit>
</unit><postal>L5L</postal><neighborhood></neighborhood><
city>Mississauga</city><county>Peel</county><state>Ontari
o</state><country>Canada</country><countrycode>CA</countr
ycode><statecode>ON</statecode><countycode></countycode><
uzip>L5L</uzip><hash>16F6288D6874C064</hash><woeid>1269755
7</woeid><woetype>11</woetype></Result></ResultSet>
<!-- gws14.maps.bf1.yahoo.com uncompressed/chunked Wed
Nov 9 10:23:36 PST 2011 -->
<!-- wws03.geotech.bf1.yahoo.com uncompressed Wed Nov 9
10:23:36 PST 2011 -->
```



```
<?xml version="1.0" encoding="UTF-8"?>
<ResultSet version="1.0">
  <Error>0</Error>
  <ErrorMessage>No error</ErrorMessage>
  <Locale>us_US</Locale>
  <Quality>87</Quality>
  <Found>1</Found>
  <Result>
    <quality>85</quality>
    <latitude>43.546647</latitude>
    <longitude>79.665735</longitude>
    <offsetlat>43.546733</offsetlat>
    <offsetlon>-79.665611</offsetlon>
    <radius>500</radius>
    <name></name>
    <line1>3359 Mississauga Rd</line1>
    <line2>Mississauga, ON L5L</line2>
    <line3></line3>
    <line4>Canada</line4>
    <house>3359</house>
    <street>Mississauga Rd</street>
    <xstreet></xstreet>
    <unittype></unittype>
```

Examples

Geonames Web Service

[http://ws.geonames.org/wikipediaBoundingBox?
north=45&south=42&east=-78&west=-82](http://ws.geonames.org/wikipediaBoundingBox?north=45&south=42&east=-78&west=-82)

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<geonames>
<entry>
<lang>en</lang>
<title>Toronto</title>
<summary>Toronto (, local pronunciation ) is the largest
city in Canada[http://wwp.canada-ca.net/ Canada],
Greenwich Mean Time. Retrieved on 2007-07-08. and is the
provincial capital of Ontario. It is located on the
northwestern shore of Lake Ontario.[http://www.city-
data.com/world-cities/Toronto.html Toronto, Ontario,
Canada, North America] Retrieved on 2007-07-08. With
over 2 (...)</summary>
<feature/>
<countryCode> </countryCode>
<population>0</population>
<elevation>0</elevation>
<lat>43.65</lat>
<lng>-79.3833</lng>
<wikipediaUrl>http://en.wikipedia.org/wiki/Toronto</wiki
pediaUrl>
<thumbnailImg>http://www.geonames.org/img/wikipedia/4700
0/thumb-46020-100.jpg</thumbnailImg>
```

Alternatives

JSON

[http://ws.geonames.org/wikipediaBoundingBoxJSON?
formatted=true&north=45&south=42
&east=-78&west=-82](http://ws.geonames.org/wikipediaBoundingBoxJSON?formatted=true&north=45&south=42&east=-78&west=-82)

JSON

```
[{"place":null,"geo":null,"truncated":false,"in_reply_to_status_id_str":null,"in_reply_to_status_id":null,"favorited":false,"in_reply_to_user_id_str":null,"source":"web","contributors":null,"in_reply_to_screen_name":null,"created_at":"Tue Oct 26 16:01:27 +0000 2010","retweet_count":null,"in_reply_to_user_id":null,"user":{"statuses_count":1491,"description":"grafia","show_all_inline_media":false,"profile_use_background_image":true,"favourites_count":57,"followers_count":124,"contributors_enabled":false,"notifications":null,"profile_background_color":"0c3bf5","geo_enabled":false,"profile_background_image_url":"http://a1.twimg.com/profile_background_images/159886578/tumblr_l8hz3wJ5lK1qb7u0po1_400.gif","profile_text_color":"eb1725","url":"http://www.fotolog.com.br/livinhacontreras","verified":false,"profile_background_tile":true,"follow_request_sent":null,"lang":"en","created_at":"Fri Jun 26
```

JSON

```
[
  { "place":null,
    "geo":null,
    "truncated":false,
    "in_reply_to_status_id_str":null,
    "in_reply_to_status_id":null,
    "favorited":false,
    "in_reply_to_user_id_str":null,
    "source":"web",
    "contributors":null,
    "in_reply_to_screen_name":null,
    "created_at":"Tue Oct 26 16:01:27 +0000 2010",
    "retweet_count":null,
    "in_reply_to_user_id":null,
    "user":
      { "statuses_count":1491,
        "description":"grafia",
        "show_all_inline_media":false,
        ...
```


SQL Support in RDBMS

Direct XML export

Direct XML import

Storing XML as a data

--xml

```
mysql --xml < humans.sql
```

```
mysqldump --xml -p starwars
```

LOAD XML

```
mysqldump --xml -p starwars  
persona > personas.xml
```

```
load xml local infile  
"personas.xml" into table  
personas2;
```

ExtractValue

```
use common;
select
  ExtractValue(xml,
    "/Locations/Location[4]/LocationName[1]")
  as name,
  ExtractValue(xml,
    "/Locations/Location[4]/Address[1]")
  as address
from xmltest;
```

Scaling

Facebook

.5 billion users

30 billion pieces of content per month

20 billion events *per day*

= 200,000 events per second

```
select url, count(*) from status join  
liking on status.id = liking.status_id  
group by status.url;
```

Instead:

[http://www.facebook.com/video/video.php?
v=707216889765&oid=9445547199&comments](http://www.facebook.com/video/video.php?v=707216889765&oid=9445547199&comments)

“NoSQL” Storage

Basic ideas:

- Key-Value Storage & Retrieval

- Documents or sparse relations

- Schema-free

- Distributed storage

“Wide Column Store”:

- BigTable, HBase, Cassandra

“Document Store”:

- CouchDB, MongoDB

MapReduce

Map

split the task into many parts

all parts should be equivalent

e.g.: count “likes” by URL for batches
of 1 mln hits

Reduce

put together the results

e.g: merge the batch counts into a total
count

Cf. with Rel. Model

Pros:

- horizontal scaling

Cons:

- more work

- “you are not Google”

Questions?