Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Institute of Computer Science
Chair of Communication Networks
Prof. Dr.-Ing. P. Tran-Gia

Crowdsourcing – evolution of work organization in Internet ?

Phuoc Tran-Gia, Matthias Hirth, Tobias Hoßfeld

Institute of Computer Science
University of Würzburg
Germany

Crowdsourcing –
spam or serious business model ?

**Phuoc Tran-Gia, Matthias Hirth, Tobias Hoßfeld**

Institute of Computer Science
*University of Würzburg*
Germany

# Crowdsourcing & evolution of work organization in Internet

▶ Motivation: Serious Business or Hype?

▶ Definition of Crowdsourcing & Evolution of Work Organization

▶ Applications, Use Cases and Potential

▶ Quo vadis

MISH'S
Global Economic
Trend Analysis

HOME   CONTACT   ADVERTISERS   SITKA PACIFIC

Google™ Custom Search

RECENT POSTS

42520 readers
BY FEEDBURNER

Subscribe By Email

RECENT POSTS

Trichet Hikes, Two More Expected; Bank of England Holds; Fundamental Factors Affecting Currencies; Is Trichet's Move the Right Move?

Too Many Bureaucrats and They Are Paid Too Much - Part II

Vibrant to Vacant: Mall Vacancies Highest in 11 Years; Online Retail Sales Hit 12%

Too Many Bureaucrats and They Are Paid Too Much

California Governor to Start Dog and Pony

Monday, April 26, 2010 11:53 PM

"Crowd-Sourcing" IBM to Cut 3/4 of its Permanent Staff by 2017?

Computer Weekly and its sister publication, Personnel Today, claim **IBM crowd sourcing could see employed workforce shrink by three quarters**

Multinational firms saddled with huge people costs are considering downsizing their permanent workforce and hiring sub-contractors on a scale never seen before - presenting an "enormous" management task for HR.

IT giant IBM told Personnel Today that the firm's global workforce of 399,000 permanent employees could reduce to 100,000 by 2017, the date by which the firm is due to complete its HR transformation programme.

Tim Ringo, head of IBM Human Capital Management, the

ABOUT MIKE SHEDLOCK

Mike Shedlock / Mish i registered investment representative for Sitka Pacific Capital Manage Continue reading...

FREE

MISH'
Global Econom

http://www.wired.com/wired/archive/14.06/crowds.html

Favoriten   Mail Reply Re Dear Phuoc   Vorgeschlagene Sites   Kostenlose Hotmail

Seite   Sicherheit   Extras

Sign In | RSS Feeds

All Wired

Issue 14.06 - June 2006

Subscribe to WIRED magazine and receive a FREE gift!

The Rise of Crowdsourcing

Remember outsourcing? Sending jobs to India and China is so 2003. The new pool of cheap labor: everyday people using their spare cycles to create content, solve problems, even do corporate R & D.

By Jeff Howe   Page 1 of 4 next »

1. The Professional

Story Tools

PRINT   MAIL

Story Images

Click thumbnails for full-size image:

Claudia Menashe needed pictures of sick people. A project director at the National Health Museum in Washington, DC, Menashe was putting together a series of interactive kiosks devoted to potential pandemics like the avian flu. An exhibition designer had created a plan for the kiosk itself, but now Menashe was looking for images to accompany the text. Rather than hire a photographer to take shots of people suffering from the flu, Menashe decided to use preexisting images – stock photography, as it's known in the publishing industry.

Feature:

The Rise of Crowdsourcing

Plus:

5 Rules of the New Labor Pool

Look Who's Crowdsourcing

http://www.wired.com/blogs

Internet   100%

VentureBeat

INTERPRETING INNOVATION

VentureBeat Profiles | Events | Jobs | Videos | Newsletters | Entrepreneur Corner | Mobile App Spotlight

MAIN   MOBILEBEAT   GREENBEAT   GAMESBEAT   DEALSBEAT   DEMOBEAT   SOCIALBEAT   MEDIABEAT

CrowdFlower raises $7M to crowd-source freelance work

March 22, 2011 | Matthew Lynley   Add a Comment

Like   One person likes this.   Tweet

Looking for something?

Have news to share? Launching a startup?
Email: tips@venturebeat.com

BUSINESS LISTING VERIFICATION

Correct inaccurate business listings.

SEARCH RELEVANCE

Assess the relevance of your search results.

PRODUCT CATEGORIZATION

Categorize large data sets.

CUSTOM SOLUTIONS

Tailored solutions to fit your needs.

Crowdflower, a firm that specializes in crowd-sourcing freelance workers and making it easy for companies to quickly find freelance programmers and other types of professionals, announced today that it has raised $7 million in its second round of funding led by Harmony Partners.

Growth of Microworkers Platform

▶ Commercial Crowdsourcing,

▶ Platform initiated in 2009

▶ August 2012:
~ 300k users from 200
countries

▶ Growth can be
approximated
by a logistic growth model
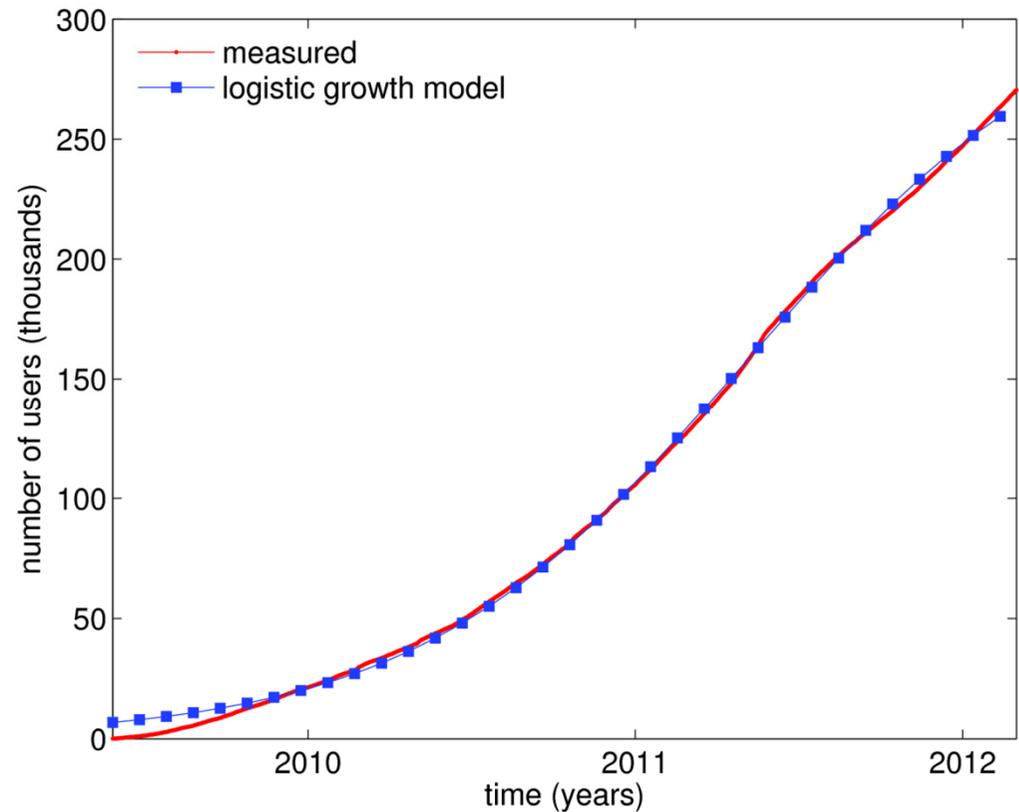
Crowdsourcing & evolution of work organization in Internet

▶ Motivation: Serious Business or Hype?

▶ Definition of Crowdsourcing & Evolution of Work Organization

▶ Applications, Use Cases and Potential

▶ Quo vadis

Phuoc Tran-Gia, Tobias Hoßfeld, Matthias Hirth

Evolution of Work Organization
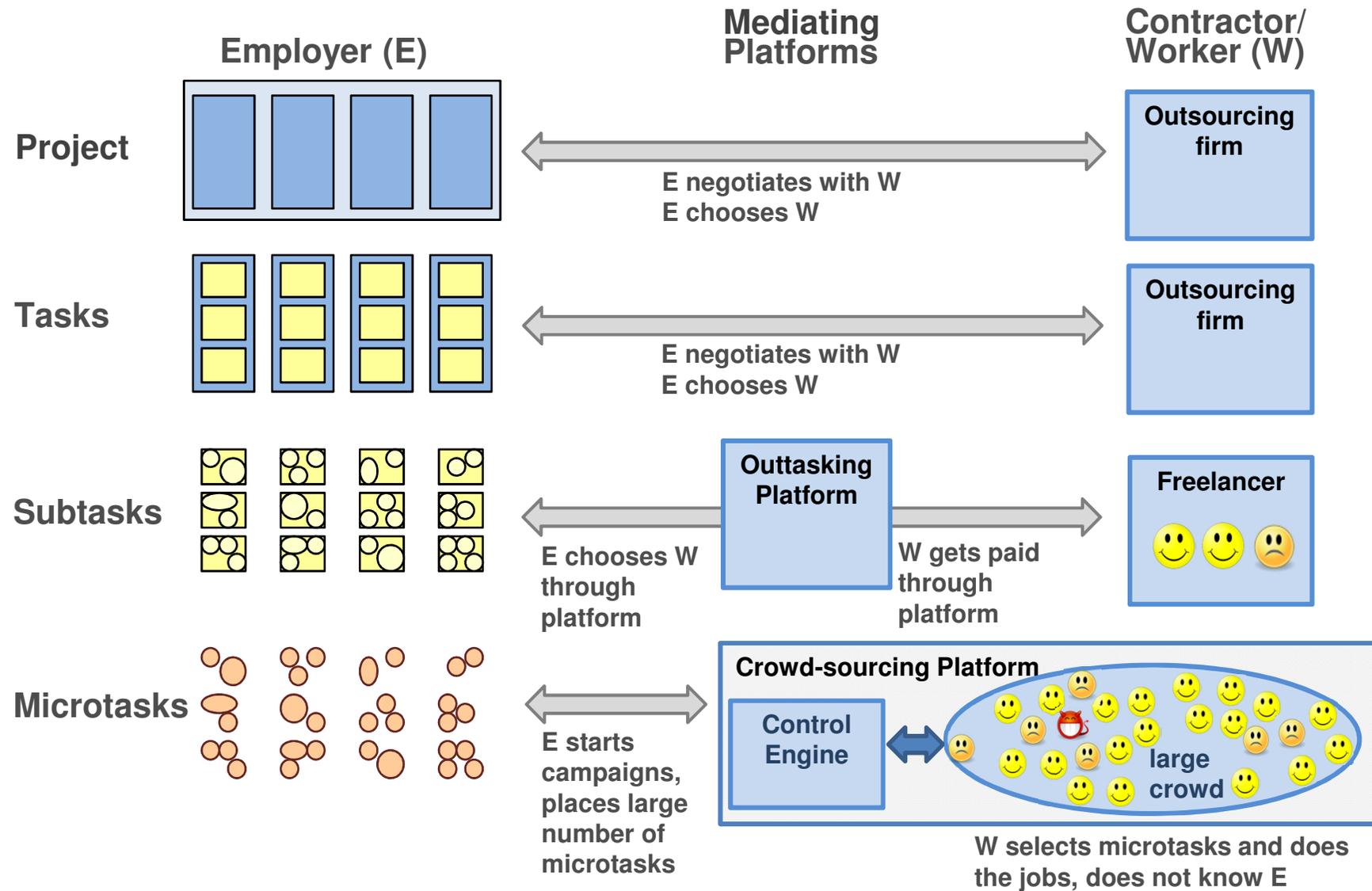
Crowdsourcing Workflow

1 Submit task

Crowd-sourcing Platform

2 Pull task

4 Deliver task

3 Submit task

5 Approve and pay

Employer

Control Engine

Worker Crowd
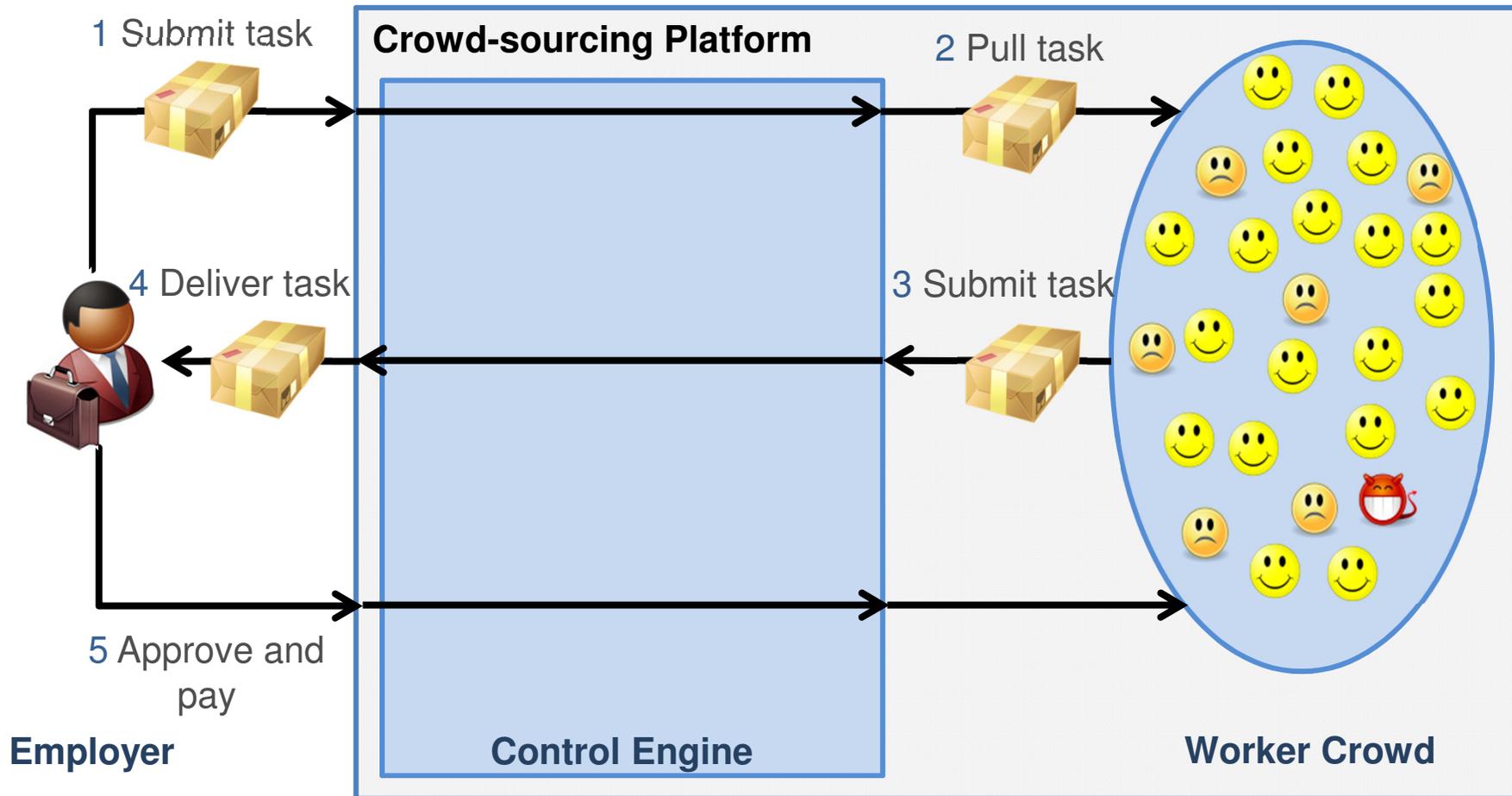
Classic outsourcing and outtasking platforms

rather classical approaches

▶ oDesk

- Routine tasks, complex tasks, or creative tasks
- Employer chooses workers dependent on the job demands and worker profiles
- Average payment: per hour $ 9.70, per fixed-price jobs: $ 296

▶ Elance

- Routine tasks, complex tasks, or creative tasks
- Employer post tenders and choose workers

▶ Innocentive

- Target: creative tasks
- Employer chooses best solution submitted by workers and rewards the corresponding worker
- Payment: Fixed rewards depending on the challenge $ 5000 - $ 1,000,000

much more platforms on the market

Major crowdsourcing platforms

Crowdsourcing

… hype at the moment …

▶ Amazon Mechanical Turk (Mturk)
- routine and simple tasks
- pure crowdsourcing platform
- micropayments per task

▶ microWorkers
- international crowds of 300000
- growth rate of 5-10% per month
- pure crowdsourcing platform
- Micropayments: paypal, alertpay, moneybookers, …

▶ … rapidly growing number of platforms

Nationality of workers in MechanicalTurk and microWorkers

Nationality

U.K. 1%
Pakistan 0.5 %
Romania 0.5 %
Philippines 1%
Canada 3%
Other 5%
India 32%
U.S. 57%

MTurk

Nationality

Other 19%
Indosenia 19%
Nepal 2%
Malaysia 2%
Poland 2%
U.K. 2%
Pakistan 2%
India 14%
Egypt 3%
Romania 4%
Philippines 5%
Bangladesh 13%
U.S. 13%

microWorkers

Phuoc Tran-Gia, Tobias Hoßfeld, Matthias Hirth
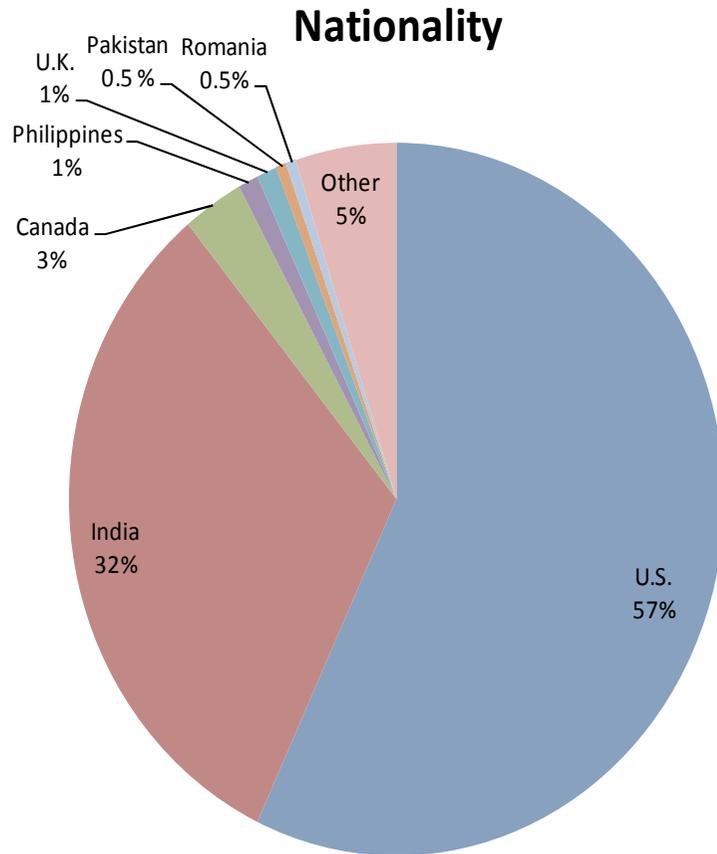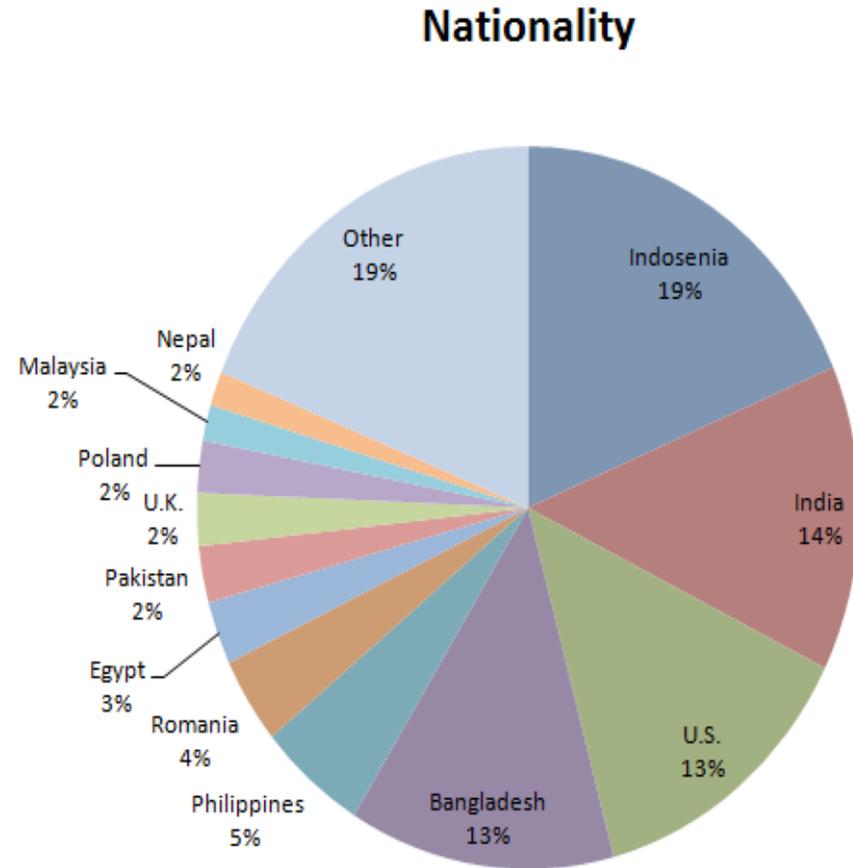
Definition

Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call.

(Jeff Howe 2008 http://www.crowdsourcing.com/)

What enables crowdsourcing?

Enabling technologies for Crowdsourcing

▶ Widespread availability of Internet access

▶ Search engines & Internet marketing

- Pagerank & Search Engine Result Page (SERP)

- New currencies in the Internet (page impression, click-through-rate, etc)

- Marketing in Internet (Google Adsense&Adwords, Briteads, …)

- Search Engine Optimization (SEO)

▶ Micropayments

- Paypal, Alertpay, Moneybooker, …

▶ Communities and social networks

- Rapid growths of social networks

- Change of information diffusion speed

▶ Reputation management techniques

Crowdsourcing & evolution of work organization in Internet

▶ Motivation: Serious Business or Hype?

▶ Definition of Crowdsourcing & Evolution of Work Organization

▶ Applications, Use Cases and Potential

▶ Quo vadis

Phuoc Tran-Gia, Tobias Hoßfeld, Matthias Hirth

Recent Tasks for Crowdsourcing in Current Internet

▶ Online reputation improvements

▶ Image labeling / recognition

▶ Acquisition of new users / joining new platforms

▶ Online advertisement / search engine optimization

  ▪ Posting twitter information

  ▪ Creation of back links

▶ User surveys for marketing / scientific studies

▶ Content creation

  ▪ Forum posts, blog entries

  ▪ Articles

▶ Data extraction & categorization

▶ Research and development

Routine tasks

Complex tasks

Creative tasks

Knowledge Collection and Generation

▶ Idea:
- Collection and post-processing of existing data
- Use "Wisdom of the crowd" to generate new knowledge

▶ Realizations:
- reCAPTCHA
- Wikipedia

▶ Examples:

Digitalization of old texts using reCAPTCHA
- Captchas are used to distinguish between automated scripts and humans (cf. email registration)
- Idea: Use words from book digitalization process which cannot be recognized by high quality OCR and add distortion
  → result is impossible to be recognized by any other script or program
- Use one known word to check if the input is done by a human, additionally one unknown word is digitalized

Integrated OCR and Crowdsourcing Digitalization Process

▶ Digitalized version of literature fosters and supports research in linguistics, but

- current OCR software cannot identify old German typefaces (like "Fraktur") reliably
- manual transcription is to cost and time intensive

➜ Using a combined approach is promising

Original text

Scanning

Scanned pages

OCR

Editable text including information about the recognition confidence

Text with low OCR recognition confidence is post-processed

human crowd

Corrected text

Digitalized pages

Automatic post processing and finalization

Digitalized text
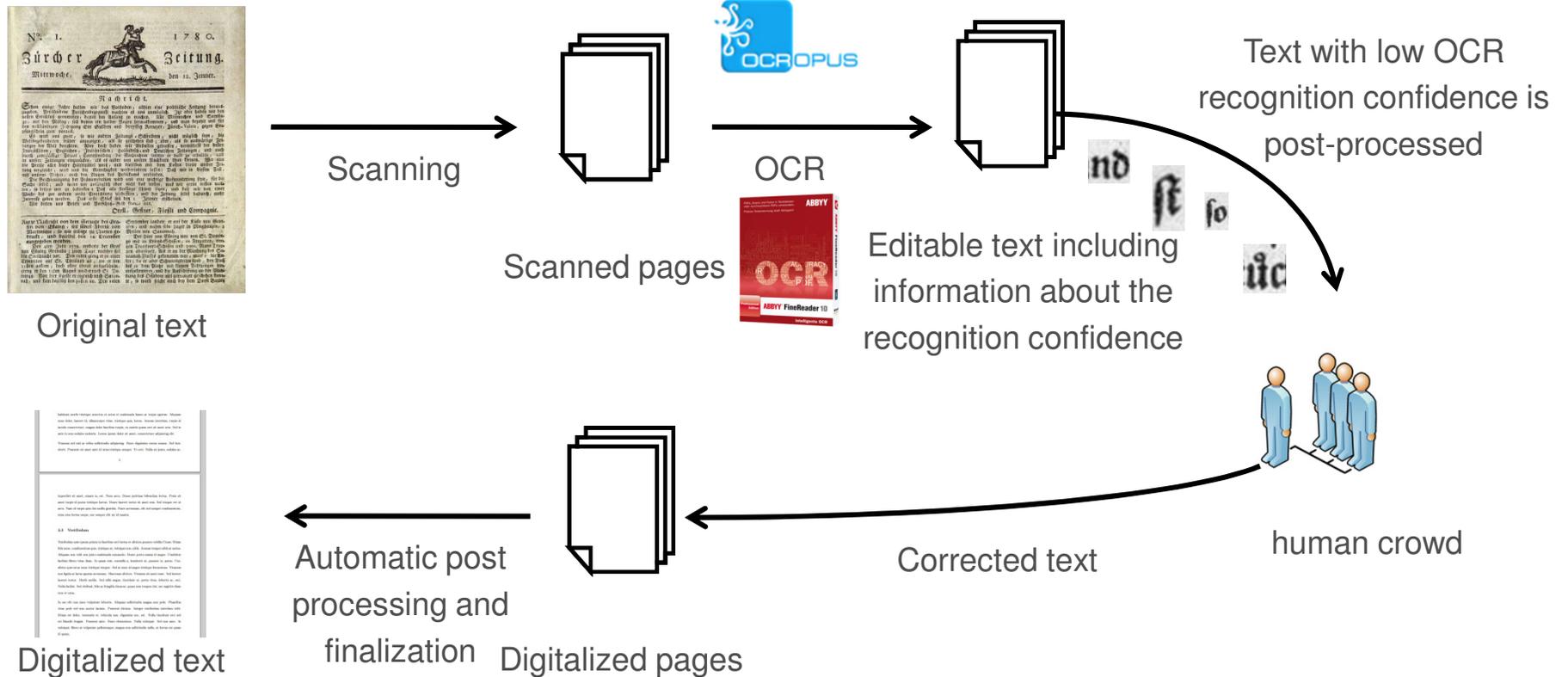
Crowd-Sensing

▶ Basic Ideas:
  ▪ Use humans as mobile sensors
  ▪ Collect data in the real world (in contrast to data collection from web pages)

▶ Realization:
  ▪ Smartphones or specialized mobile devices as sensors
  ▪ Different types of user interaction possible
    – Automatic sensing
    – Explicit sensing
    – Implicit sensing
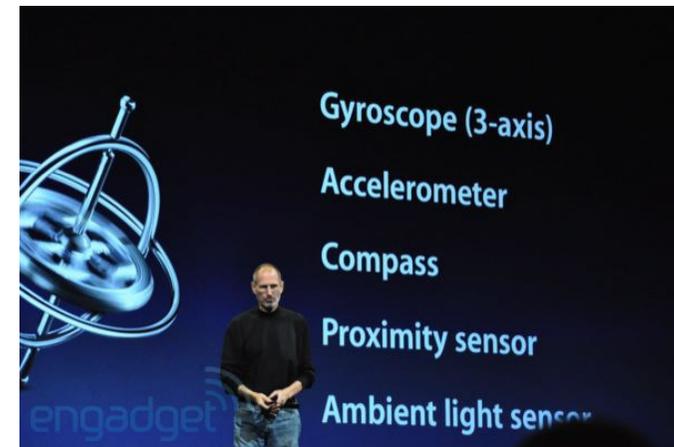
© www.engadget.com

▶ Example:
  Smartphone based road damage detection
    – "Feel" the bump when driving though a road whole using the smartphone's accelerometer
    – Ask the driver if he just passed a road hole
    – Process the answer and report the whole including GSP coordinates to a central server

Creativity and Innovation Design

▶ Idea:
- Developing new products, software and business ideas
- Creation and evaluation of new designs and user interfaces (UI)

▶ Realizations:
- 99designs.com
- uTest.com

▶ Example:

Testing of a new software UI

  – Usually UI test are performed by co-workers or designated tests
  - ➔ Limited resources
  - ➔ Different expectations than regular user
  - ➔ Better trained than regular users

  – Crowdsoucing UI test enable real user feedback
  - ➔ Faster and more realistic feedback
  - ➔ Feedback for users with a different point of view

Problem Solving

▶ Idea:
- Use a broad audience to solve specific R&D challenges
- Use crowdsourcing to find specialist for advices in specialized tasks like decontamination after a chemical accident

▶ Realization:
- Innocentiv.com
- Kaggle.com

▶ Example:
- Tasks from the aforementioned platforms
  - *Identify patients who will be admitted to a hospital within the next year, using historical claims data.*
    Reward $3,000,000
  - *Develop an algorithm for matching latent fingerprints.*
    Reward $100,000
  - *Recycling Challenge: How to Recover Glass From Buildings?*
    Reward $ 5,000

Layers of Crowdsourcing Processes

Company — Microsoft, Google, Amazon, EBay

Specialized mediator — microTASK

Crowdsourcing platforms — microWorkers (work & earn or offer a micro job), amazon mechanical turk (Artificial Artificial Intelligence), CrowdFlower, uTest, Machine clouds and mobile crowdsourcing Platforms

International workers — Worker at specific locations, Worker with special skills, Worker with special devices

Julius-Maximilians-UNIVERSITÄT WÜRZBURG

Decomposing an Outtasking Process
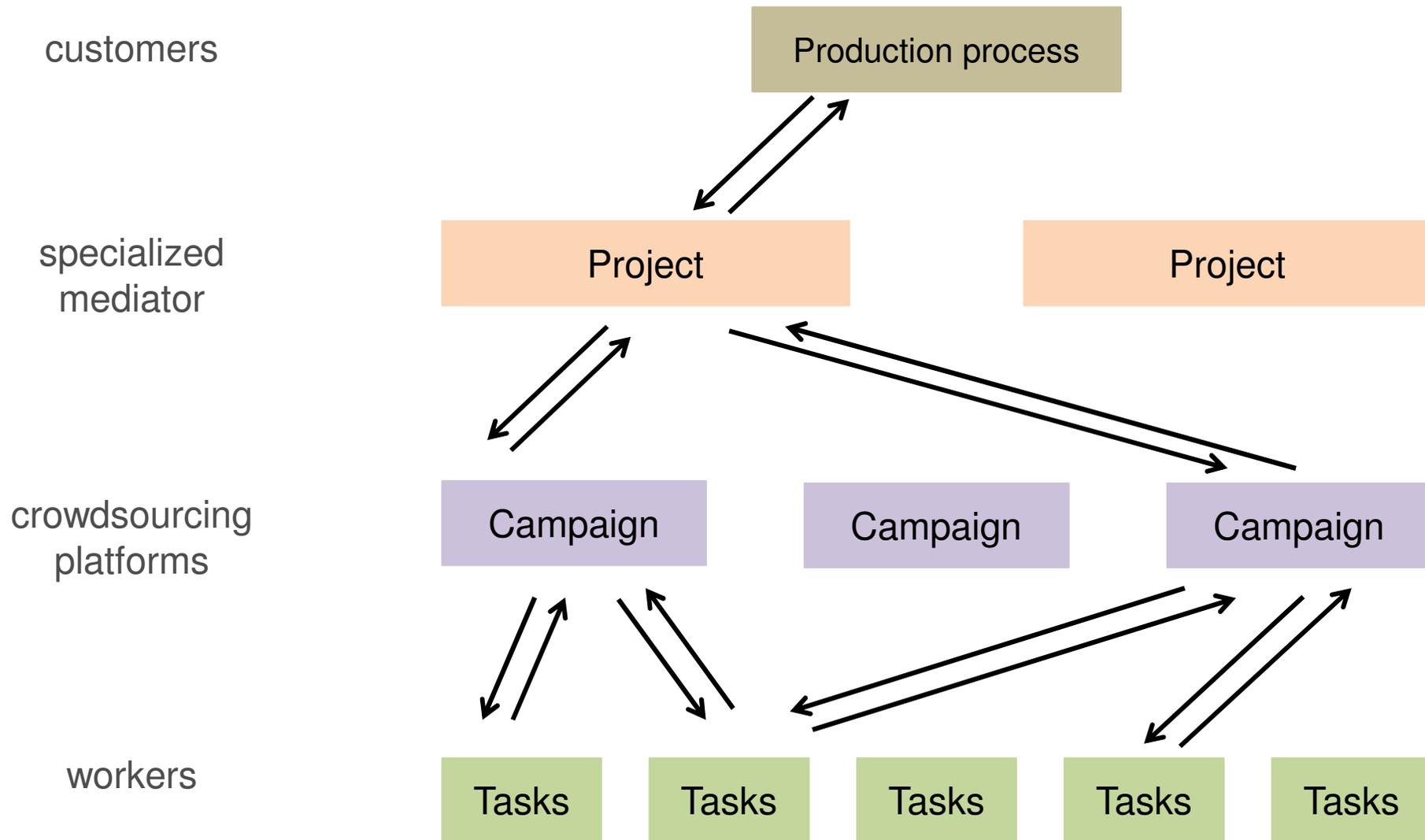
customers

Production process

specialized
mediator

Project

Project

crowdsourcing
platforms

Campaign

Campaign

Campaign

workers

Tasks

Tasks

Tasks

Tasks

Tasks

Crowdsourcing & evolution of work organization in Internet

▶ Motivation: Serious Business or Hype?

▶ Definition of Crowdsourcing & Evolution of Work Organization

▶ Applications, Use Cases and Potential

▶ Quo vadis

Human cloud vs machine cloud
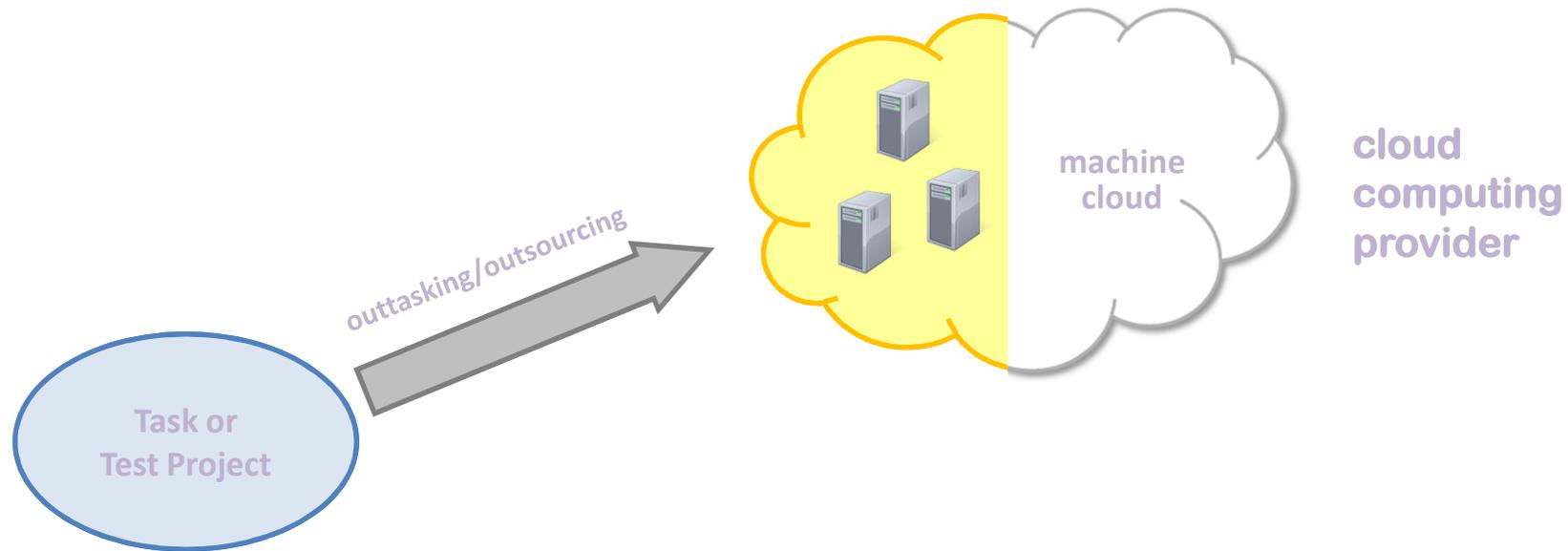
Human cloud vs machine cloud

Human cloud vs machine cloud

machine cloud

cloud computing provider

Task or Test Project

outtasking/outsourcing

Crowdsourcing (intercloud outtasking)

crowdsourcing

human cloud

crowdsourcing platforms

Human cloud vs machine cloud

Phuoc Tran-Gia, Tobias Hoßfeld, Matthias Hirth
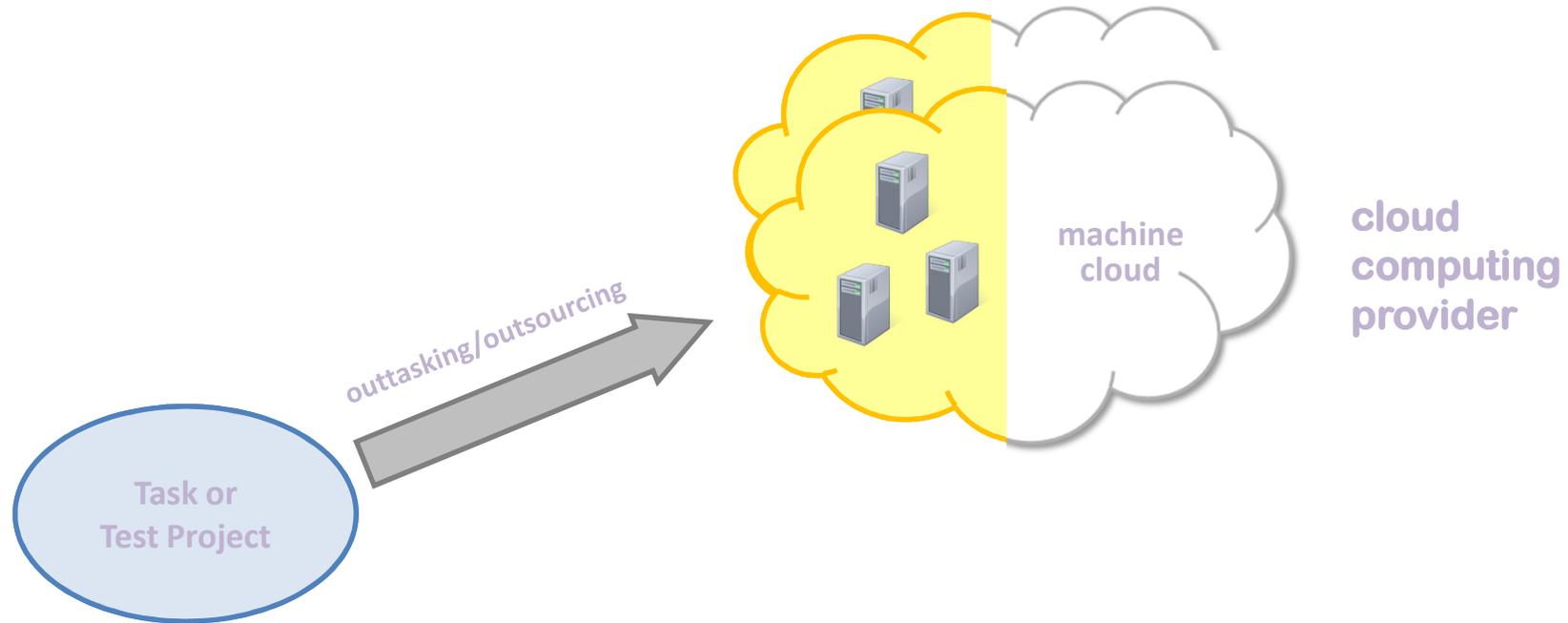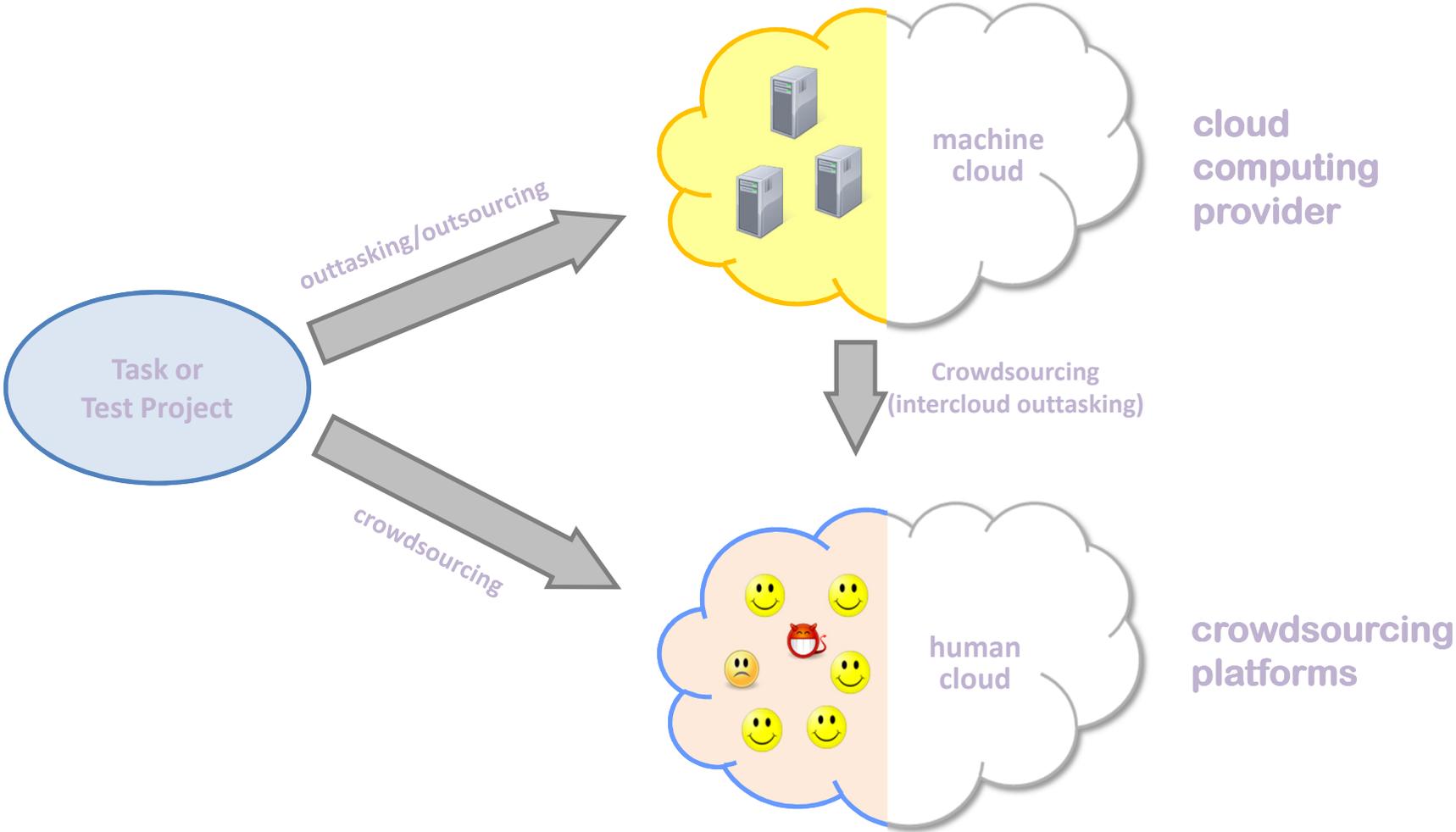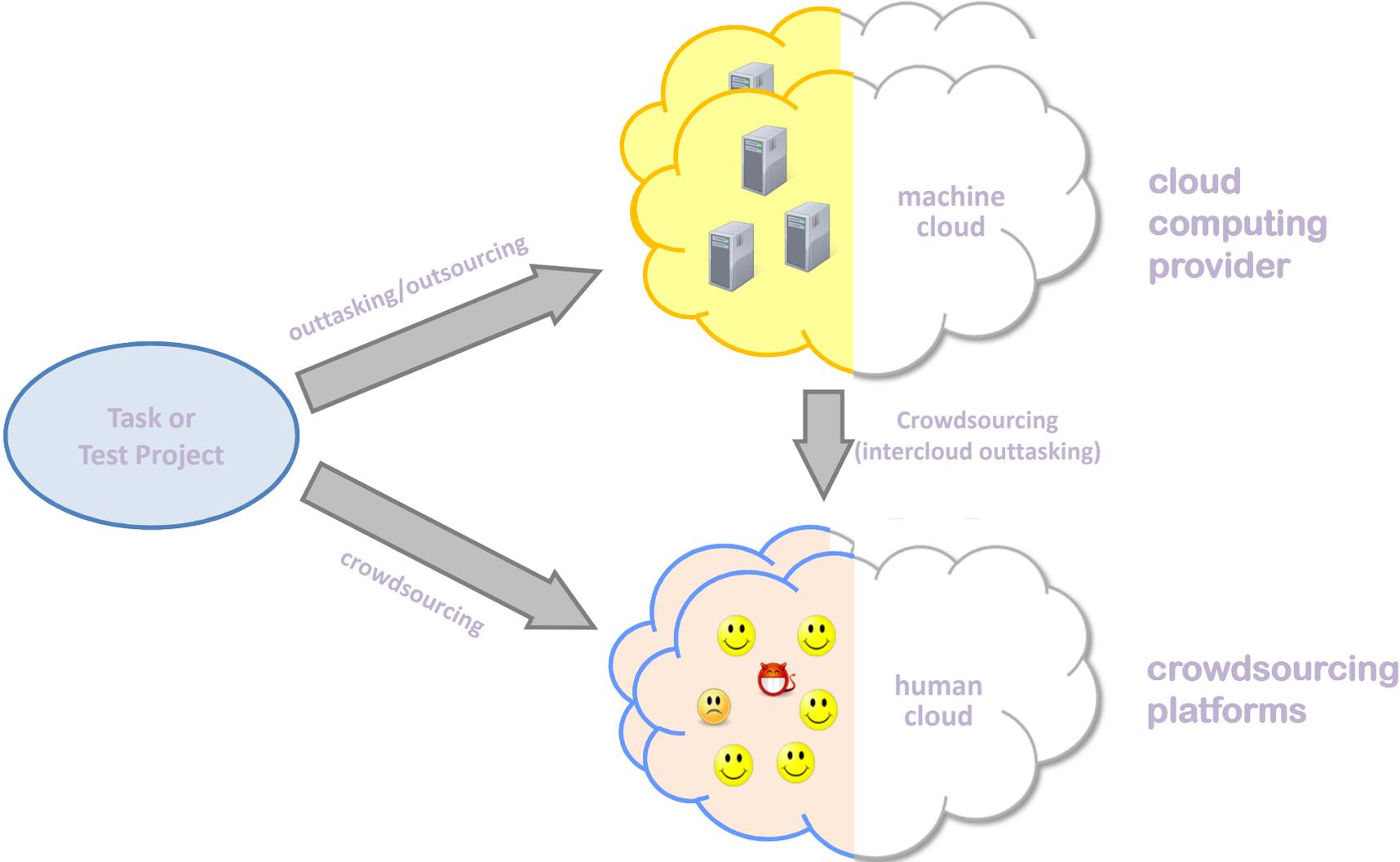
Application Areas & Challenges

▶ Current Internet Crowdsourcing

- Web based platforms for various purposes
- *Main challenges*
  - Massive amount of data generated
  - Huge number of users

▶ Real-time Crowdsourcing

- Crowdsourcing task which are time critical (transcription of images for blind people)
- *Main challenges*
  - Real time constrains on the underlying infrastructure (clouds)
  - 24/7 crowd required (huge size of the crowds, diversity of the crowd in terms of time zone location)

Application Areas & Challenges

▶ Mobile Crowdsourcing

- Crowdsourcing applications using mobile devices for sensing and communication with the users

- *Main challenges*

  – Diversity of user devices

  – Validation of location base data in spare populated areas

▶ Enterprise Crowdsourcing

- Using crowdsourcing based approaches for company internal processes

- *Main Challenges*

  – Redesign existing process to make it "crowdsourcable" way

  – Protection of internal data/security issues

geigercrowd.net

what is project geigercrowd?
ガイガークラウドプロジェクトとは?

geigercrowd is an approach to fill in an information gap. radiation measurements should be open data and provided by the government in an easy to access and uncensored manner. the current situation in japan shows a different picture. websites providing up-to-date meassurements are either down or don't show valid values. this is where we want to fill in. we crowdsource the meassuring to people in japan who operate and own automated or handheld geiger counters or other equipment that delivers data on radiation levels. this way we all together can try to fill the currently white map on this page with up-to-date radiation measurements that help people all over japan to assess the current risk situation. geigercrowd.net is backed by a capable content delivery network to hopefully handle more load than the other available resources that currently provide this data on the web.

放射線計測値は公にされるべきもので、行政の手によって、容易にアクセス可能なデータ改ざんされないような形で公開されるべきです。しかし現在の日本の状況はこれとは違う様相を見せています。最新の計測値を提供するウェブサイトはダウンしてしまっているか、信頼できる値を載せていません。我々はそういう状況を埋め合わせようとしています。我々は自動／手動読み取りのガイガーカウンター、または他の放射線レベル計測機器を所持している日本にいる多数の人々のデータを集約しようとしています。そうすることで、集約された最新の放射線計測値をこのサイトの日本地図にプロットし、日本全体の人々がだれでも現時点の状況や危険性にアクセスできるようになります。これは多数の人が少しずつ協力して一つの問題を解決するというクラウドソースという手法です。ガイガークラウド.netはこれを実現するために十分なコンテンツ配信ネットワークによってサポートされており、願わくば現在インターネット上で提供されている他の情報源よりも多くのデータを取り扱うようになりたいと考えています。

current status
現在の状況

how to participate?
どのように参加することができますか?

what is project geigercrowd?
ガイガークラウドプロジェクトとは?

Presenting Organizations

MINISTRY OF YOUTH AND CULTURE

World Bank LAC
opportunities for all

Sponsors

Development Bank of Jamaica Limited

RIM BlackBerry

USAID
FROM THE AMERICAN PEOPLE

Participants

freelancer

samasource

microWorkers
work & earn or offer a micro job

WILDFIRE

Vielen Dank !

Backup

Digital Jam 2.0

▶ Three day event in July 2012 to overcome youth unemployment in Jamaica (in 2009 at 27.3%)

▶ Organizers:
- Government of Jamaica
- World Bank
- and other governmental and non-governmental institutions

▶ Approach:
- Demonstrate opportunities of the virtual economy for unemployed or underemployed youth in Jamaica
- Youths can enter the global labor market via the internet and earn an income by performing paid tasks in the virtual economy
  - "micro-work" (tasks that are broken into component parts and then outsourced online)
  - "e-lancing" (where young people offer their services via online portals)
  - "crowdsourcing" (where they can participate in finding collaborative solutions for large projects launched by companies and other entities online)

Growth of Microworkers Platform

▶ Logistic growth model
  - Initial population $N_0 = 6.75$, limit $K = 316.16$, growth rate $r_0 = 1.97$

▶ Square growth model
  - $f(t) = 30.81\, t^2 + 17.94$

▶ A look into crystal ball: Sep 14, 2013
  - 313k users (logistic)
  - 647k users (square)

▶ Increasing importance in the future
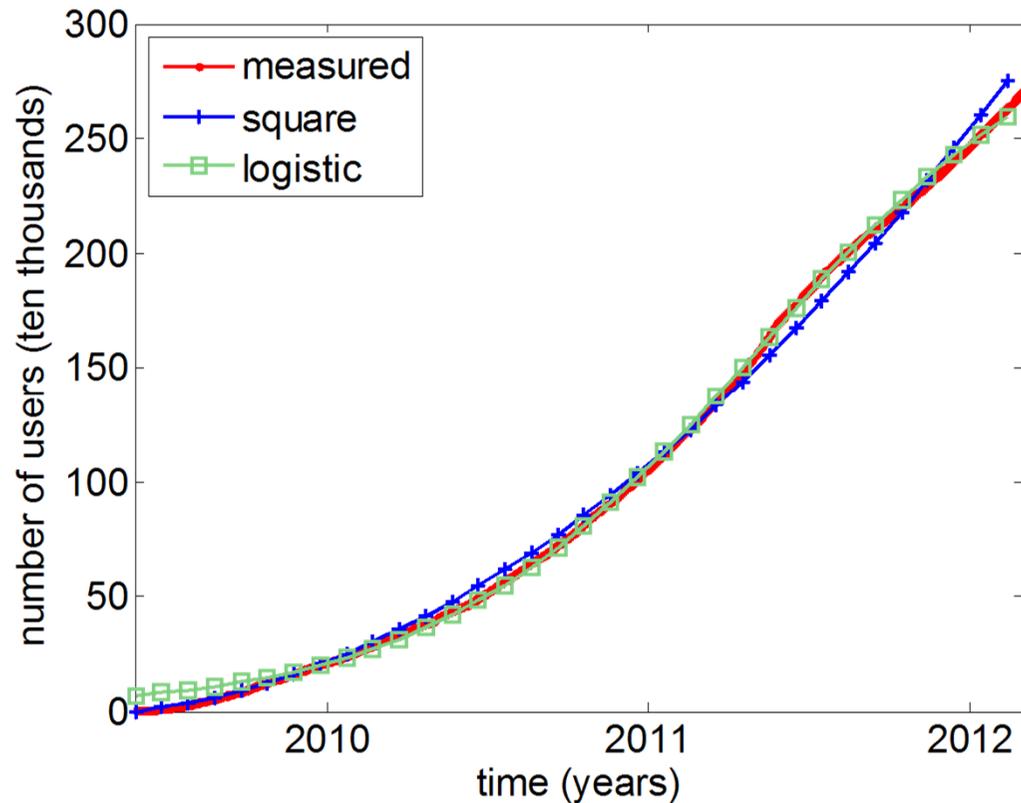
Definition of Crowdsourcing (alternative)

Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of

activity undertaken.

"Towards an Integrated Crowdsourcing Definition" - Enrique Estellés-Arolas et al.

Consequences of Crowdsourcing

▶ Individual
- ▪ "Everyone is self-employed"
  - – Pros
    - • More flexibility in terms of working hours, work place and type of work
    - • Possibilities for a better income for higher skilled worker
  - – Cons
    - • No job guarantees
    - • World wide work marked decreases the wages
- ▪ Privacy issues in crowdsourcing sensing applications

▶ Society
- ▪ New legal solutions are require to deal with a world wide task marked (Platform located in country A, worker located in country B, employer originating form country C)
  - – Taxation
  - – Trade unions and worker rights
  - – Application of which laws?
- ▪ Changes in education system, as in general no proof of a profession is required to work on a certain task
- ▪ Enables people to easily contribution to the common wealth, e.g. reporting of pollution and submit solutions for community problems

Consequences of Crowdsourcing

▶ Companies
- Enables new business models and business cases, e.g. consulting on how to crowdsource tasks
- Fosters competition and globalization
- Scalable, flexible and diverse workforce
  – Faster completion of tasks
  – On demand experts
- Security of business data

▶ Science
- Crowdsourcing as a tool
  – Easy access to huge test panels
  – New ways to perform measurements
- Crowdsourcing as research topic
  – Legal issues :
    • Creative tasks: Intellectual rights
    • Wrong data, hired flash crowds: liability
  – Computer science:
    • Quality assurance mechanisms
    • Task and worker recommenders
- Current projects
  – The EveryAware Project (Crowd sensing)

Emerging Applications and Potential of Crowdsourcing

- ▶ Crowd sensing:
  - ▪ Data collection
    - – human sensor – mobile devices, smartphones with sensors
    - – Example:
  - ▪ Realizations:
    - – Automatic sensing
      - • no interaction of the user necessary, data is gathered automatically (position of crowd → prediction of electricity consumption of the mobile access network)
    - – Explicit sensing
      - • Users manually input and share data (geolocation-aware image collection e.g. of road damages)
    - – Implicit sensing
      - • Relies on manual inputs, however sensor data are by-products of other user actions (rfid based train ticket control can be used to predict crowding of train station)

- ▶ Knowledge Collection and Generation
  - ▪ Wikipedia, Wisdom of the crowd
  - ▪ IBM's Spoken Web system offering knowledge sharing fro farmers in rural India via mobile phones

- ▶ Creativity and Innovation Design
  - ▪ Developing new business ideas
  - ▪ Developing new products (Tchibo Ideas)

- ▶ Problem Solving
  - ▪ Research and Development
  - ▪ Crowdsourcing based disaster management (Allowing professional volunteers to offer their skills to local governments during desasters)

Applications

► Applications

- Helping handicaped people
  - http://www.crowdsourcing.org/editorial/crowdsourcing-iphone-app-to-help-the-blind-see/4430
  - http://abclocal.go.com/kgo/story?section=news/technology&id=8760115

► Enterprises:

- Digitalisierung von Krankenakten (IBM)

Application Areas of Crowdsourcing and Resulting Challenges

▶ Current Internet CS
  - Web based platforms for various purposes
  - Main challenges
    – Massive amount of data generated
    – Huge number of users

▶ Mobile CS
  - Crowdsourcing applications using mobile devices for sensing and communication with the users
  - Main Challenges
    – Diversity of user devices
    – Validation of location base data in spare populated areas

▶ Real-life CS
  - Crowdsourcing application which require geographical presence (painting a room, buying food…)
  - Main Challenges
    – Acquiring enough participants (workers and employers) in the same area
    – …

▶ Real-time CS
  - Crowdsourcing task which are time crucial (Transcription of images for blind people)
  - Main Challenges
    – Ubiquitous access to the CS platform
    – Real time constrains on the underlying infrastructure (clouds)
    – 24/7 crowd required (huge size of the crowds, diversity of the crowd in terms of time zone location)

▶ Enterprise CS
  - Using crowdsourcing based approaches for company internal processes
  - Main Challenges
    – Redesign existing process to make it "crowdsourcable" way
    – Protection of internal data/security issues

quo vadis ?

▶ Sandbox or serious business models?

▶ Building Platforms and Customizing Use Cases
- Transform a use case into campaigns
- Preprocessing of data and postprocessing of results
- Building internal crowdsourcing platforms within companies
- Supporting automatic creation of microtasks from large projects

▶ Building crowd
- Forming specialized crowds for predefined jobs
  - Workers with special qualifications for difficult tasks
  - Workers with specific demographic properties, e.g., for surveys
- Forming "internal" crowd in companies and enterprises
- Developing cheat detection techniques