

## Annual report 2010

for

**iAD – Information Access Disruptions**  
Centre for Research-based Innovation (CRI)  
at Fast Search & Transfer ASA

*Enabling technology and emerging services in the information age*

**Microsoft** | Development Center  
Norway



**Netview**  
Technology



**Cornell University**



**Induct**



**UNIVERSITY  
OF OSLO**



**BI** NORWEGIAN SCHOOL  
OF MANAGEMENT



**NTNU**

Norwegian University of  
Science and Technology

**Table of Contents**

<b>Centre Vision and Objectives .....</b>	<b>3</b>
<i>Introduction Summary .....</i>	3
<i>iAD Objectives.....</i>	3
<b>Research Strategy.....</b>	<b>4</b>
<b>Centre Organisation.....</b>	<b>5</b>
<i>Consortium Partners.....</i>	5
<i>International Collaboration Partners .....</i>	5
<i>Organisational Structure.....</i>	6
<i>Key Personnel.....</i>	6
<i>Collaboration between Partners.....</i>	9
<b>Activities and Results .....</b>	<b>11</b>
<i>Business Context.....</i>	12
<i>Technology and Academic Context .....</i>	12
<b>International Collaboration.....</b>	<b>16</b>
<b>Recruitment in 2009 .....</b>	<b>17</b>
<b>Communication / Dissemination .....</b>	<b>18</b>
<b>Summary of Scientific Activities in 2008 .....</b>	<b>22</b>
<i>Sub-project 1: Schema agnostic indexing services.....</i>	22
<i>Sub-project 2: Processing high-speed data streams.....</i>	24
<i>Sub-project 3: Scalable infrastructure for push and pull based computing .....</i>	26
<i>Sub-project 4: Extreme precision &amp; recommendation in multimedia access.....</i>	27
<i>Sub-project 5: Understanding and managing the disruptive potential of iAD.....</i>	30
<b>Appendix A: Personnel at the academic partners .....</b>	<b>32</b>
<b>Appendix B: Statement of Accounts.....</b>	<b>38</b>
<b>Appendix C: Publications .....</b>	<b>39</b>

## Centre Vision and Objectives

---

### Introduction Summary

Information access technology has emerged as one of the most innovative technology areas, impacting a wide range of industries, business models and even transforming social patterns. Information access is also a highly prioritized research area in the European and international research programs. Google, Yahoo and Microsoft (GYM) have proven both the strategic and short term business value of search based systems in a consumer setting. FAST has emerged as the recognized leading provider of search based systems. This includes both embedding of search as an IT infrastructure component and search as a mission critical part of new services. Furthermore, search technology represents a disruptive threat to the database vendors such as Oracle and IBM that add search capabilities with an incremental approach.

Dramatic changes in technology, business models and services are foreseen in information access. It is our belief that Norway has a unique opportunity to build a strong science and business platform to play a key role in the future of the information age. Integrating academic and corporate efforts with international centres of excellence will be of key importance in this effort. During the course of 2010, FAST has become fully integrated with Microsoft Corporation and as of 1. January 2011 changed its name to Microsoft Development Center Norway (MDCN). While this continues to represent a well of new business opportunities emerging, the support for iAD and related research programs also remains at the same level, with the explicit support of Microsoft.

### iAD Objectives

The Information Access Disruptions (iAD) Centre targets core research for next generation precision, analytics and scale in information access. Furthermore, iAD builds international networks to identify and execute on global disruption opportunities enabled by emerging services in the information age.

iAD enables:

- A Norwegian computer science cluster in information access
- Strong Norwegian presence in the global information age transformation
- Proven international networks in information access for both science and business development
- Excellent international research networks within areas prioritized by the EU research programs
- Strong trans-Atlantic research cooperation
- Strong innovation focus secured by corporate hosting and full alignment with the vision and business goals of the corporate partners
- Integration of computer science and business research to translate scientific breakthroughs into concrete innovation results
- Emerging disruptive services such as RFID analytics, multimedia access, next generation telecom, information warehousing and streaming computing
- Norwegian industry cluster development in enabling technology and emerging services for the information age

## Research Strategy

---

iAD spends all public funding from the Research Council of Norway in the university sector driving academic research to:

- Create schema agnostic indexing services fusing structured, unstructured and multimedia content in precision, analytics and scale optimised information access services
- Develop scalable and fault-tolerant system architectures including data processing and mining platforms for capturing, cleaning, and extracting knowledge from high-speed data streams
- Develop and validate in real environments next generation infrastructure for distributed information access
- Develop extreme precision solutions for access to multimedia
- Identify disruptive processes either within information access or enabled by information access solutions that can be used as a cluster foundation for Norwegian IT innovation

iAD creates tight relationships between already cooperating centres of excellence inside universities and corporate partners. In total, the iAD corporate partners FAST, Accenture have more than 200 R&D people working on productization within the iAD field. Hence, iAD prioritizes all external funds and added corporate grants to the university groups. A key objective for iAD is to stimulate research and student recruitment across a university based national cluster in information access. This national cluster has unique international research networks and access to challenging customer cases even within the consortium. iAD is organized in 5 major research tasks which are described in the Chapter on [Activities and Results](#). Each of these tasks has dedicated university owners.

iAD adopts a scientific approach to innovation and business planning. BI Norwegian School of Management studies the disruptive processes within the iAD field and creates a basis for cluster development within information age technologies in Norway. iAD combines computer science and business research to explore the diverse innovation opportunities enabled by new search based paradigms.

FAST continues to disseminate business value globally together with Accenture and Schibsted, based on iAD results. Breakthroughs in iAD will be measured against the following FAST development practice:

- The innovation must have a visual and practical impact on the end-user search experience
- The usage of the innovation must have benefits across a wide range of search driven applications
- The innovation must fundamentally change or add to what problems that can be efficiently addressed with search based technology
- The benefits must be validated in customer settings

## Centre Organisation

---

### Consortium Partners

iAD is hosted by FAST Search & Transfer<sup>1</sup>. FAST is recognised by Gartner in the Magic Quadrant as a leading provider of information access solutions for enterprises. iAD is fully aligned with FAST's company vision. Furthermore, iAD consists of university and corporate partners with a proven track record in a global setting and a history of successful collaboration within the iAD consortium.

iAD creates a national cluster for computer science research in information access by integrating the leading groups at the universities in Tromsø<sup>2</sup> (UiTø), Trondheim<sup>3</sup> (NTNU) and Oslo<sup>4</sup> (UiO). BI<sup>5</sup> Norwegian School of Management studies the disruptive business processes within the iAD field. Accenture<sup>6</sup> provides a laboratory for validating enterprise solutions in Norway. Furthermore, Accenture has established an international business unit, Accenture Information Management Services (AIMS), which provides iAD with a global network to experiment and deploy information access solutions. The iAD Lab provides a physical and virtual meeting place for researchers and clients to further enhance collaboration.

<b>iAD Centre Director</b>	Bjørn Olstad (CTO in FAST & Prof. II NTNU)	
<b><u>iAD Consortium:</u></b>	<b>University partners</b>	
<b>Corporate partners</b> Microsoft Development Center Norway (MDCN) Accenture ANS	Computer science NTNU, IDI (Trondheim) UiTø (Tromsø) UiO, IFI (Oslo)	
<b><i>FAST acts as the host institution</i></b>	<i>Business</i> BI Norwegian School of Management (Oslo)	
<b><u>Collaboration partners:</u></b>	<b><u>iAD Lab partners</u></b>	
Cornell University (USA) Dublin City University (Dublin, Ireland)	Creuna Netview	Induct Funcom

### International Collaboration Partners

The formalised and active participation of Cornell University<sup>7</sup> gives iAD a strong international foundation that also accelerates quality in Norwegian IT research. Cornell is recognised as a worldwide leader in computer science in general and the field of scalable and trustworthy distributed systems in particular. The Dublin City University<sup>8</sup> (DCU) adds a strong academic track record in innovative multimedia access and social networks.

---

<sup>1</sup> <http://www.fastsearch.com>

<sup>2</sup> <http://www.uit.no>

<sup>3</sup> <http://www.ntnu.no>

<sup>4</sup> <http://www.uio.no>

<sup>5</sup> <http://www.bi.no>

<sup>6</sup> <http://www.accenture.no>

<sup>7</sup> <http://www.cornell.edu/>

<sup>8</sup> <http://www.dcu.ie/>

## Organisational Structure

iAD is led by Centre Director Bjørn Olstad (MDCN CTO and NTNU professor), assisted by Research Program Manager Odd Petter Slyngstad. The Centre Director has overall responsibility for operational day-to-day management and is supported by additional management teams.

iAD has established a centre board to align and revise research and innovation objectives including associated recruiting. The board ensures that the intentions and plans underlying the contract for the centre are fulfilled and that the corresponding funding plan is completed within the adopted time frame. The board further ensures that the interaction between the iAD centre, the host Institution and the other collaborating consortium participants functions smoothly.

The current iAD board members are:

- Bjørn Olstad, MDCN
- Kirsti Kierulf, Accenture
- Kjell Bratbergsengen, Norwegian University of Science and Technology (NTNU)
- Gunnar Hartvigsen, University of Tromsø (UiTø)
- Morten Dæhlen, University of Oslo (UiO)
- Espen Andersen, Norwegian School of Management (BI)

## Key Personnel

Name	Background / Role
------	-------------------

### NTNU

Svein Erik Bratsberg	Professor
Svein-Olaf Hvasshovd	Professor
Magnus Lie Hetland	Førsteamanuensis
Jon Krogstie	Professor
Jon Atle Gulla	Professor
Stein Løkke Thomassen	Postdoc fellow
Simon Jonassen	PhD student
Bilegsaikhon Naidan	PhD student
Mohammad Ali Norozi	PhD student
Truls Amundsen Bjørklund	PhD student
Nils Grimsmo	PhD student
Ole Edsberg	PhD student
Hongli Yang	PhD student
Jon Marius Venstad	PhD student

### UiT

Dag Johansen	Professor
Steffen Valvåg	PhD student
Håvard Johansen	Researcher
Åge Kvalnes	Assistant professor
Audun Arnesen	PhD student
Joseph Hurley	PhD student
Kjetil Jacobsen	PhD student
Johannes Gehrke	Assoc. prof at large

Annual report 2010

**iAD – Information Access Disruptions** Centre for Research-based Innovation at  
Fast Search & Transfer AS

Cathal Gurrin	Senior researcher at large
---------------	----------------------------

**UiO**

Pål Halvorsen	Associate professor
Carsten Griwodz	Associate professor
Paul B Beskow	PhD student
Andreas Petlund	PhD student
Pengpeng Ni	PhD student
Håvard Espeland	PhD student
Tomas Kupka	PhD student
Håkon Kvale Stensland	PhD student

**BI**

Espen Andersen	Associate Director, BI and European Research Director, The Concours Institute
Ingunn Myrtveit	Professor
Erik Stensrud	Professor II
Vedrana Jez	PhD student
Yang Zhenyu	PhD student

**MDCN**

Bjørn Olstad	iAD Centre Director
Odd Petter N. Slyngstad	Research Program Manager
Øystein Torbjørnsen	Senior Research Scientist
Eirik Knutsen	Chief Architect

**Accenture**

Kirsti Kierulf	Director Innovation Lab
Jens Mosbergvik	Manager – Technical Expert Enterprise Search
Lars Sverre Gjølme	Senior Manager – Business-model responsible for Next Generation Workplace
Kjersti Solem	Manager
Nils Øveraas	Senior Executive – global partner
Vegard Kolbjørnsrud	Senior Manager – strategy – PhD student
Ida Irene Eriksen	Analyst – PhD student
Rayid Ghani	Senior Manager – research in ATL Chicago
Chad Kumbly	Consultant – research in ATL Chicago
Robert Groenvell	Manager – research ATL Sophia
Sølve Fredheim	Senior Technology Consultant
Gøran Karlsson	Director
Dan O'Hara	Senior Manager
Francis D'Silva	Senior Manager

**Collaboration Partners****Dublin City University**

Cathal Gurrin	Project leader
Sorin Sav	Postdoctoral Research Fellow
David Scott	PhD student
JinLin Guo	PhD student
Peter Wilkins	Postdoc fellow
Colum Foley	Postdoc fellow
Paul Ferguson	Postdoc fellow

**Cornell University**

Johannes Gehrke	Associate Professor
Fred B. Schneider	Professor
Robbert van Renesse	Principal Research Scientist
Christoph Koch	Associate Professor
Lucja Kot	Postdoctoral Associate

Tuan Cao  
Marcos Vaz Salles  
Michaela Goetz  
Ben Sowell  
Yin Lou

PhD student  
Postdoctoral Associate  
PhD student  
PhD student  
PhD student

## Collaboration between Partners

Collaboration among the partners in the iAD Centre is supported on various levels. The design of the five research sub-projects is fundamentally based on broad participation of multiple partners with different background. Furthermore, the mobility of researchers is ensured by a formalised rotation scheme which allows e.g. PhD students to spend a significant share of their time in the scientific environment of other partners. Some concrete examples of collaboration between the partners are given in the following paragraphs.

NTNU has continued their collaboration with Johannes Gehrke, and also with Michaela Götz, at Cornell. The main focus has been on search in social media and this work is based on ideas developed as part of the previous collaboration.

In 2010, Cornell had very intensive collaboration with Norwegian partners:

1. Cornell has had on-going collaborations with the University of Tromsø, NTNU, and Microsoft. Two faculty members from the University of Tromsø (Dag Johansen and Åge Kvalnes) spent their sabbatical from 2009-2010 at Cornell University. Cornell also has an on-going collaboration with NTNU on the iAd subproject on social search. Finally, Cornell has on-going discussions with Microsoft about technology transfer.
2. Johannes Gehrke gave two full days of lectures in a class at the University of Tromsø in 2010.
3. PhD student Truls Bjorklund from NTNU visited Cornell from 8/2008 to 5/2009 as part of an iAd exchange visit. Truls worked with Johannes Gehrke on query processing for next-generation search engines. This collaboration continued even after Truls returned to Norway where he is finishing his PhD; Johannes Gehrke has been acting as informal PhD thesis advisor to Truls all throughout 2009 and now 2010. Several papers have resulted from this collaboration.
4. PhD student Michaela Götz, advised by Johannes Gehrke, has been actively collaborating with Truls Bjorklund from NTNU on combining search with access control.
5. Johannes Gehrke, Robbert VanRenesse, and Fred Schneider have been advising the Enterprise Search Group at Microsoft (based in Norway) in 2010.

iAD Tromsø collaborates closely with FAST R&D, Dublin City University, University of Oslo, and Cornell University.

With iAD industrial partners, UiT attempts to define research problems based on real-life, practical scenarios. They have cooperated closely with MDCN/FAST the last year to integrate video search and streaming into their next generation enterprise search engine FS 14 (resulting in vESP).

Two University of Tromsø faculty from the iAD group have been on sabbatical at Cornell for a year (July 2009 - June 2010) collaborating on data centre operating system issues. This cooperation is on-going with the professors back in Norway.

Two of UiTs current PhD students are, for instance, jointly supervised with Dr. Robbert van Renesse from Cornell.

Together with DCU, UiT is studying next generation digital library platforms with focus on management and access problems. Digital video is at focus, currently investigating real-time scenarios.

Together with UiO, UiT is studying application specific approaches to performance isolation. We are also collaborating closely on next generation video search, composition, and streaming technologies.

Together with DCU, next generation digital library platforms are being studied with focus on management and access problems. Digital video is in focus, in combination with investigating general video information retrieval problems.

UiO has had close collaboration with MDCN/FAST, UiT and DCU in the DAVVI activity. Additionally, UiT has participated in some of the work on our Nornir framework.

DCU collaborates closely with University of Tromsø (UiT) and the University of Oslo (UiO) with additional project specific collaborations with BI.

With UiT we are constantly collaborating on all aspects of our work. UiT bring a strong systems background to our research and we have been working closely together to develop scalable video analysis tools and deploying them in real-world environments. Dr Gurrin (DCU) has been teaching an Information Retrieval course at UiT (with Johannes Gherke from Cornell).

With UiO we are integrating their video streaming technologies and their scalable computing work into our video processing architectures to expedite the processing of video content to support real-time analysis and personalised access.

With BI we are involved in real-world evaluation of video search technologies. We worked with BI on the evaluation of our iPad video search engine prototype for TRECVID 2010.

The main collaboration during 2010 from BIs side has been with Accenture – working together to arrange seminars, courses and presentations for a number of Norwegian and international audiences on the topics of the iAD project, including Mannheim, PFIT and Sophia Antipolis.

The collocation of researchers in the centre is further supported by additional part-time positions at other partners. This enables e.g. a professor-at-large position for Johannes Gehrke (Cornell) and a senior-researcher-at-large position for Cathal Gurrin (DCU) at the University of Tromsø, further strengthening the collaboration among partners and knowledge transfer e.g. by means of focussed guest lectures. Similarly, cross-fertilisation between academia and industry has been a focus area in 2010.

Furthermore, collaboration involving the iAD commercial partners has materialized through the following projects

<b>Project Name</b>	<b>Duration</b>	<b>Objectives</b>
FindmyPicture	2010	Develop applications for multi-touch and search on the surface table. Test it under the student-week, package the application as component, disseminate into the NTNU environment for further research and student work.
Information Management and Access	2010- 2012	Research activities in Accenture Technology Labs focusing on how information management will be influenced by new technologies like search. Working with large data information systems like bank systems in finance and in the governmental industry to understand how to detect fraud earlier, how to use information in gap-analysis of knowledge skills in HR-systems
Service Innovation	2008 – 2012	Describe and develop models for innovation in Services. Lead by BI.
Value networks	2006- 2010	Describe and documents theory supporting the change in business models in industries as new technology enables them to move from value chains to value shops and value networks. 1 PhD is funding by Accenture Research Fund. Partnership with BI and Telenor. The work also was expanded with a summer intern doing research on the value network models for the Nordic media industry in the period of 1998 to 2008 and Schibsted was interviewed.
Accenture Interactive Manangement	2009- 2013	Develop and pilot and scale up a digital platform for marketing and brand maintenance in large brand organizations. Search is used as the driver in the front and in the backend to industrialize the reuse of brand-templates, reuse effect on multichannel complexity, maintain and update content in templates based on consumer behaviour. In pilot with first client fall 2010, to be full scale launched spring 2011. FAST/MDCN used as search engine, due to fact that FAST was introduced to our research environment through the IAD-program and the fact that Microsoft is building their new enterprise search strategy on the FAST platform on net.

## Activities and Results

---

### Business Context

The user revolution has elevated the expectations from both users and businesses on IT effectiveness in general and information access in particular. Focus has shifted from content to users, from publishing to consumption and from control to empowerment. FAST's vision is "***Empowering People and Business for the Information Age***". This vision is based on the fact that search is becoming the key technology for building successful online businesses. Search allows businesses to exploit intent analysis in the creation of unique user experiences. Furthermore, search can effectively monetize the traffic generated by such online services. Similarly, search has proven to enable substantial productivity improvements for the information worker. In summary, iAD has therefore been designed to drive the following trends from a technology perspective:

- ***Search will be everywhere.***  
Search will be available in virtually every application and interface, allowing people to quickly find who and what they need.
- ***Search will enable unique user experiences.***  
Search will enable rich, new user experiences by creating a conversation based on context—an understanding of who the user is, what information is available, and what business process is involved.
- ***Search will change the way people do business.***  
Search will create net new revenue streams, improve business decisions, and increase business productivity.

iAD technology developments have during 2010 enabled numerous deployments of disruptive information access solutions.

### Technology and Academic Context

FAST is in a transition to a next generation search platform. At the core, search is all about matching knowledge about the user (**intent**) with knowledge about **content** to produce unique user experiences. The next generation search platform is built on three pillars:

- **Interaction Management:** How to build algorithmic and intent-driven portals
- **Contextual Matching:** How to deliver unprecedented relevance and expressiveness
- **Content Analytics:** How to capture and enhance information across diverse content sources

These 3 pillars have been developed as independent solutions that can be fitted together for an end-end next generation information access platform.

A set of breakthrough capabilities in the next generation search platform have been validated:

- *Interaction Management*
  - o Ability to drive higher relevance in user experience
  - o Flexible business logic to leverage user interaction within the business context
  - o New paradigm for algorithmically driven user interfaces
- *Contextual Matching*
  - o Significantly enhanced expressiveness for relevance and navigation models
  - o Order of magnitude improvements in corpus size, indexing speed, indexing latency
  - o Expressive data models capturing text, DB schemas and rich media
  - o Expressive query models supporting SQL and XQuery in addition to existing query languages
- *Content Analytics*
  - o Data fusion enablement with a consistent framework across text, data and media
  - o Unified model for analytics at index and query time enabling a shift to more comprehensive on-the-fly computing models
  - o Significantly enhanced developer productivity
  - o Functional richness to drive data quality and data refinement capabilities

The research and development around these 3 pillars inside FAST has been tightly linked to the research tasks and *key achievements* by the academic partners in iAD:

- **iAD sub-project 1 – Schema agnostic indexing service**
  - o Publication of 8 conference papers and 2 journal papers
  - o Won best paper award at WISE conference in Hong Kong
  - o Implementation of 2D indexing and hybrid partitioning prototypes
  - o Development of new Twig Join algorithm that is in average 3 times faster than previous work
  - o Development of prototype for search in social media in collaboration with Cornell University
  - o Started collaboration with Professor Tomáš Skopal at Charles University in Prague in the area of Ptolemaic indexing
  - o Completion of 3 master theses
- **iAD sub-project 2 – Processing high-speed data streams:**
  - o Developed the abstraction of entangled queries for data-driven collaboration.
  - o Developed new similarity search algorithm for low similarity thresholds. Algorithm outperforms previous work by up to two orders of magnitude.
- **iAD sub-project 3:**
  - Reported from UiT:
    - o Published 9 papers.
    - o Graduated PhD Lars Brenna.

- 2 professors in the group spent sabbatical year at Cornell strengthening ties to this world class institution.
- Prototypes developed:
  - vESP – DAVVI enhancement for video and slide presentations – streaming, search, personalization (integrating with FAST FS 14).
  - Hypervisor (Vortex) being put to use for kernlet development and experiments.
  - High-performing MapReduce engine: Cogset

Reported from UiO

- Graduated 2 master students
- Published 14 papers
- vESP prototype – DAVVI enhancement for video and slide presentations – streaming, search, personalization
- Pull-patching – DAVVI enhancement for multicast (patching

**- iAD sub-project 4 - Extreme Precision and Recommendation in Multimedia Access**

- Published 9 iAD related papers in 2010
- Dissemination at UiT in the Information Retrieval Course
- Prototypes developed:
  - TRECVID 2010 iPad & iPhone video search engine
- Sports Summarisation Engine for length-optimised event-based summaries.

**- iAD sub-project 5 – Understanding and managing the disruptive potential of iAD:**

- Newly established cooperation between BI and DNV, combining economic theory and **data mining** of DNV repositories; explore, apply, and validate data mining and analysis technologies including search technologies, machine learning methods, text analytics, and fuzzy logic.
- As part of the research project Knowledge-based Norway, work will be done tracing the development of search technology as well, using the case of MDCN/FAST in comparison to more traditional Norwegian IT firms.

In summary, the academic iAD partners have in 2010:

- Achieved critical mass at the participating universities in terms of personnel.
- The PhD students and researchers have continued to align their activities with long-term opportunities for the commercial partners.
- An impressive set of reference implementations have been completed that provide an excellent foundation for further experiments and publications.
- Established academic results continue to demonstrate clear value in a commercial context.

- The established iAD-Lab located in Oslo continues to be a valuable asset towards recruitment of new partners and investigation of relevant research issues, now also expanded with a location at the University of Tromsø.
- Continued very effective international collaborations.

In summary, the commercial iAD partners have in 2010:

- **Developed new methods/products:** Continued development on Scrum Master for search and digital content on multi-touch devices, and developed additional product prototypes.
- **Contributed to international conferences/workshops:**
  - 8 presentations,
  - 10 virtual classes,
  - several articles in technical magazines,
  - 2 POV on search topics.
- **Performed case studies on:**
  - Nordic innovation
  - FAST as a technology driver
- **Carried out dissemination activities:**
  - Monthly meetings on search with clients in the iAD Lab
  - Lectures for students in the innovation center, at NTNU, UIO and BI.
  - Number of seminars and presentations to Accenture Clients. Large project with large clients on enterprise search strategy work.

## International Collaboration

---

In addition to the per se truly international nature of the iAD Centre due to the formalised participation of Cornell University and DCU, the Norwegian partners bring expertise from their portfolios of existing research projects and external collaborations to the centre. Although brought to completion, the background of the following past projects remain highly relevant to the scope of iAD:

- NFR162349, an RCN-funded collaborative research project with FAST, UiTø, NTNU on next-generation analytical search.
- PHAROS, large scale EC-funded Integrated Project on multimedia search with FAST as Technical Director.
- RUSHES, EC-funded STREP on search of unedited audio-visual footage.
- CHORUS, EC-funded coordination action for audiovisual search.

The iAD Tromsø group has a close collaboration with our two international partners, Dublin City University and Cornell University. Faculty Dag Johansen and Åge Kvalnes spent their sabbatical at Cornell University for a year (July 2009 - June 2010).

The iAD-Tromsø group cooperates with another SFI, the TTL (Telemedicine) project in Tromsø. So far, we have studied whether our iAD technologies and solutions developed for large scale data access can be applied to telemedical problems. We have two joint publications, so it is a realistic problem area worth exploring further.

One of our PhD students, Kjetil Jacobsen, is also co-advised with Professor Keith Marzullo, UC San Diego (currently on leave at NSF Washington as division director).

In the development of Ptolemaic indexing the team at NTNU has collaborated with Prof. Tomáš Skopal at Charles University in Prague. Prof. Skopal has been heading the Similarity Retrieval research group at the Charles University in Prague since 2006 and his research interests is well aligned with the research done on similarity search in iAD.

Dr. Gurrin at DCU maintains a collaborative link with Northeastern University, Shenyang, China and is actively supervising some iAD related projects there. Northeastern University has one of the top ten computer science departments in China.

In collaboration with colleagues in the National Institute of Standards in Technology (NIST) in the US, Colum Foley (postdoctoral fellow, DCU) developed an "Oracle" webservice that was used by all participants of the TRECVID 2010 Interactive search task. This webservice will be used again in the interactive search task of TRECVID 2011. Colum also coordinated an effort involving colleagues at DCU to make a corpus of test data available to participants of the Instance Search task.

Also, Fred Schneider is the Chief Scientist, NSF TRUST Science and Technology Center.

As an example of future collaboration, Yang Zhenyu (Ph.D. candidate) will spend one year at UC Berkeley, starting February 2011.

## Recruitment in 2010

---

Recruitment of outstanding research staff is vital to maintain momentum and ensure sustainability of centre goals. In 2010, the following 4 PhD fellows and more than 19 MSc. students were recruited at the different academic partners in the iAD Centre:

Name	PhD fellows	PostDoc fellows/ Researchers	MSc students	Other
<b>NTNU</b>			3 finished in 2010, 6 new recruited	
<b>UiO</b>		1	Multiple	
<b>UiT</b>	1		4	
<b>BI</b>			8	
<b>Cornell</b>	2	1		
<b>DCU</b>	1	3	1	
<b>Sum</b>	<b>4</b>	<b>5</b>	<b>19</b>	

iAD has reached critical mass in terms of personnel at the academic partners during 2010.

## Communication / Dissemination

---

A number of activities towards communication of iAD research foci and dissemination of iAD results have been performed. These include publications in major international conferences and magazines, as well as through other more informal channels. This is reflected in the list of references.

An example of this from the commercial partners is the Fast Innovation Center/iAD Lab, which acts as a hub for global clients of the commercial partners to connect with the research agenda within iAD. Concrete, active initiatives are in place towards involving large clients in this Center on a regular basis. Several talks, seminars, etc. towards specific groups and the general public are also been carried out by the commercial partners in iAD. This builds on the following initiatives:

- The Fast Innovation center with the iAD-Lab as the research center is recognized as an innovation center in Accenture, Avanade and Microsoft's global network. We are participating in monthly global activities to follow up dissemination of research into the clients' accounts around the world.
- The team in the center has support functions on enterprise search cases with global clients and supporting the local teams with content disseminated as POV, in the design phase and on occasions being pulled in to client work as experts.
- The Centers and lab's prototypes are being installed in other centers in our network.
- We are co-producing new applied research to the global agenda with teams from other centers.

In the academic sphere, dissemination of relevant results is on-going with respect to inclusion into current courses and lectures.

The following activities were performed by UiT towards communication and dissemination during 2010:

- Monthly: co-locating for a day the iAD Tromsø group with the FAST R&D group in town presenting research projects, discussing mutual problems, and the like.
- iAD information retrieval course (Inf-3203) jointly arranged (UiT, Cornell, DCU) in Tromsø 2010-2011. Final exam May 2011.
- Talks at conferences and workshops related to our papers published.
- Jointly arrangement in Tromsø in December with Norwegian Academy of Technological Sciences (NTVA, UiTø and Sparebank1 Nord-Norge where iAD professor Fred B. Schneider was keynote (for industry, government, national politicians, press, academia):
  - Number of interviews at regional and national news broadcast channels.

The following activities were performed by DCU towards communication and dissemination during 2010:

- Weekly internal iAD DCU group meetings for informal discussion of progress and presentation of upcoming conference talks and project ideas.
- iAD information retrieval course (Inf-3203) jointly arranged (UiT, Cornell, DCU) in Tromsø 2010-2011. Final exam May 2011.
- Talks at conferences and workshops related to our papers published.

Four university courses are taught at UiO in which results and experience from iAD are disseminated:

- INF1060 - Introduction to operating systems and data communications
- INF3190 – Data communication
- INF5063 - Programming asymmetric multi-core processors
- INF5071 - Performance in distributed systems

The following activities were performed by Cornell towards communication and dissemination during 2010:

Lectures and invited talks by Johannes Gehrke:

1. Participant on the panel “Future Directions” at the 2010 NSF CISE Information, Integration, and Informatics Workshop. Rosslyn, VA, April 2010.
2. Developing, Optimizing and Hosting Data Driven Web Applications. Invited Talk at the Information Security and Cryptography Group, Saarland University, Saarbruecken, Germany. September 2010.
3. Scalability for MMOs and Transaction Processing. Colloquium at the Ecole Polytechnique Fédérale de Lausanne. Lausanne, Switzerland, November 2010.
4. What Can Database Systems Do For Computer Games and Simulations? Colloquium at the Technical University Munich. Munich, Germany, December 2010.

Lectures and invited talks by Robbert van Renesse:

1. "Building Reliable Reconfigurable Services," The 11th International Conference on Parallel and Distributed Computing, December 2010, Wuhan, China. (keynote)
2. "Efficient Reconciliation and Flow Control for Anti-Entropy Protocols", July 22, 2010, Microsoft Research Asia, Beijing, China.
3. "Efficient Reconciliation and Flow Control for Anti-Entropy Protocols", July 26, 2010, Baidu, Beijing, China.
4. "Efficient Reconciliation and Flow Control for Anti-Entropy Protocols", July 29, 2010, Fudan University, Shanghai, China.

Lectures and invited talks by Fred Schneider:

1. Private Sector Perspectives on Department of Defense Information Technology and Cybersecurity Activities., Hearing at United States House of Representatives, Armed Services Committee, Terrorism and Unconventional Threats and Capabilities Subcommittee, Washington, D.C., February 2010.
2. Principles and Principals for Authorization in Nexus. Triangle Computer Science Distinguished Lecturer Series. North Carolina State University Raleigh, North Carolina, April 2010.
3. Principles and Principals for Authorization in Nexus. Harvard University. Boston, Mass, April 2010.
4. Beyond Hacking: An SOS! Keynote lecture, 32nd International Conference on Software Engineering. Capetown, South Africa, May 2010.

5. Towards a Science of Security. Jasons Meeting, invited talk. San Diego, CA., June 2010.
6. Principles and Principals for Authorization. GTISC Seminar, Georgia Intsitute of Technology, September 2010.
7. Should you be paranoid? A conversation on Cyber Security Cornell Entrepreneur Network of Washington DC November 2010.
8. Towards a Science for Security. Distinguished Lecture Series of the Computer Science Department. University of Massachusetts, Amherst, MA. November 2010.
9. Principles and Principals for Authorization in Nexus. 2010 Samuel D. Conte Distinguished Lecture Series. Purdue University, West Lafayette, IN. December 2010
10. Principles and Principals for Authorization in Nexus. Computer Science Department. University of Tromsø, Tromsø, Norway. December 2010.
11. New Doctrine for Cybersecurity. Inaugural Tromsø NTVA Technical Forum. Tromsø, Norway. December 2010.

Espen Andersen / BI held the following presentations with material and themes relevant to the iAD projects in 2010:

- December 9: Sosiale medier for bedriftsintern samhandling, seminar for RAV Norge, Hotel Continental, Oslo
- November 17: Ut av siloene. Keynote, Dataforeningens seminar Ut av siloene, IBM Headquarters, Mastemyr, Oslo
- November 2: IKT i skolen: Paradokser og poenger, Høstseminar, Hedmark Fylkeskommune, Hamar
- October 14 - 17: Search technology and disruption, in *IT Management and e-Business*, MBA program, Norwegian School of Management/Fudan University, Shanghai, China.
- September 21-25: Technology Strategy, Entrepreneurship and Innovation, (with Hermann Kopp), University of Mannheim EMBA module, BI, Oslo  
Confidential November 2010

- August 24: Å forstå endring: Teknologi, kunder og informasjon, Postbanken lederseminar, Oslo
- August 23: utfordringer og muligheter for norske IKT-selskaper, (telekonferanse) Innovasjon Norges IKT-seksjon.
- May 19: Teknologievolution og markedsutvikling, Tromsø Senter for fjern teknologi, Tromsø
- March 19: Disruptive innovations: When the good technology loses, NetPower seminar, Devoteam DaVinci/SINTEF, Oslo
- March 6-7: Search technology and disruption, in *IT Management and e-Business*, MBA program, Norwegian School of Management/Fudan University, Shanghai, China.
- March 3 - 6: IT Management and e-Business, MBA program, Norwegian School of Management/Fudan University, Shanghai, China. (session on search technology)
- February 17: *Teknologiutvikling og teknologistrategi*, R&D samling, Kongsberg Maritime, Lillehammer
- February 9: Hvor blir alle kundene av? Disruptiv konkurranse i bank og finans, Accenture/Avanade customer seminar, Felix konferansesenter, Oslo
- February 8: *Dataanalyse endrer seg*, keynote BOB/NIB Brukerkonferanse 2011, Holmenkollen Park Hotel, Oslo
- January 31: *Teknologi, organisasjon og kundekommunikasjon*, lederseminar Norsk Tipping AS, Hamar
- January 26: De nye sosiale mediene: Tidstyv eller verdiskaper, Kompetanseforums Årskonferansen 2010, Handelshøyskolen BI, Oslo.
- January 25: Må vi akseptere større forskjeller i norsk skole?, Ringer i vannkonferansen, Hotell 33, Oslo
- January 24: *Hjelper IT oss å oppnå millenium-målene?* Debatt Polyteknisk Forening IT, Håndverkeren, Oslo
- January 20: IKT i skolen: Paradokser og poenger, foredrag ved åpningen av Senter for IKT i utdanningen, Tromsø
- January 15: *Wikipedia - en vakker tanke*, Wikipedias 10-års jubileum, Litteraturhuset, Oslo
- January 8: *Det feilorganiserte sykehuset: Produktivitet i en problemløsende virksomhet*, Årsmøte Norsk forum for helseledelse, Holmenkollen Park Hotel, Oslo

## **Summary of Scientific Activities in 2010**

---

### **Sub-project 1: Schema agnostic indexing services**

#### **Schema agnostic end2end design**

In the area of indexing and retrieval based on similarity the work has mainly been focused around Ptolemaic indexing. This is a novel indexing method for spatial search developed in iAD which is significantly different from the metric indexing methods usually used. Research has also been done on the selection of pivot values for metric indexing. Two papers have been published and several are in the pipeline.

As part of this task we are also looking at semantic search for ubiquitous environments. The goal is to develop and evaluate a system for semantic search in mobile environments using search technology developed by Microsoft as part of iAD. The platform is now up and running and experiments are being conducted to understand the opportunities the new platform offers. One journal paper has been published on semantic search as part of this task.

Traditionally search engine has in some degree used the structure in documents for relevancy, but this has mostly been ad hoc and based on knowledge of a simple schema (what is title, body and metadata). As part of the iAD project we want to extend this and get a deeper understanding of how structure can improve the ranking of documents in results. This work is in its initial stages and existing work is being investigated.

#### **Support extreme scalability for static and streaming data sources**

The main goal of this task is to improve scalability of large scale search systems. The main questions investigated are how to distribute full text indexes over large numbers of computers, how to construct these indexes and how to best evaluate queries over them. Prototypes have been developed demonstrating the 2D index and a hybrid partitioning strategy. The work has resulted in a paper that was given the best paper award at the WISE conference in Hong Kong. During this investigation many new opportunities have been identified and are currently being pursued.

Social media like Facebook and Twitter creates large volumes of data that is difficult to navigate in. In contrast to the World Wide Web where everybody can see everything, data posted on social media is often limited to friends or limited groups of members. It is therefore necessary to create new methods to search in huge volumes of data in the presence of access restrictions. In collaboration with Cornell we have investigated this problem, implemented a prototype and published one paper in this area.

One master thesis has also looked on how to best possible compress XML structure information necessary to execute queries over semi structured data. Compression is important for both reducing the size of the index in memory and on disk in addition to improving overall query performance.

In addition to this work are two master students looking on how search can be implemented in a cloud environment.

### **Consolidation and standardization of the query model**

A core part of evaluating queries over semi-structured data (e.g. XML data) is the ability to match patterns in trees. Twig Joins are the common way of doing this and we have in this activity looked at methods to effectively evaluate these joins for large datasets. The research has tried to unify most work previously done and putting this into a common framework. During this work several opportunities for improving the state of the art were discovered and new algorithms have been developed and implemented. The best new algorithm is worst case optimal and in average 3 times faster on common benchmarks. For some tests it is one order of magnitude faster, but never more than 20% slower than the best previous methods.

This activity has in 2010 resulted in four papers and now reached its end.

## **Sub-project 2: Processing high-speed data streams**

### **Infrastructure for Collaboration**

Many data-driven social and Web applications involve collaboration and coordination. The vision of declarative data-driven coordination (D3C) is to support coordination in the spirit of data management research -- to make coordination data-centric and to specify coordination scenarios through convenient declarative languages. We have proposed the notion of entangled queries, a coordination language that extends SQL by constraints that allow for the coordinated choice of result tuples across queries originating from different users or applications. A substantial challenge arises here: It is nontrivial to define a declarative coordination formalism without immediately arriving at the general (NP-complete) Constraint Satisfaction Problem (CSP) from AI. We proposed an efficiently enforceable syntactic safety condition that we argue is at the sweet spot where interesting declarative power meets applicability in large scale data management systems and applications. The key computational problem of D3C is to match entangled queries to achieve coordination. We developed an efficient matching algorithm which statically analyzes query workloads and merges coordinating entangled queries into compound SQL queries; these SQL queries can be sent to a standard database system and return only coordinated results.

### **Multimedia Search**

Finding similar objects is a key component in many applications, such as document and image clustering, plagiarism detection, near duplicate documents detection, 3D scene reconstruction, similar music and video retrieval, community mining, and personalized recommendations. We need high dimensional descriptors to address the complexity of the objects. A typical article contains hundreds to thousands of words. Many multimedia retrieval algorithms preprocess objects by extracting visual and auditorial descriptors. For images, we may need up to a million descriptors for optimal retrieval performance. Finding all pairs of objects whose similarities exceed a given threshold is a key component in many applications. In particular, there are many important emerging applications that require the use of relatively low similarity thresholds.

We developed ATLAS, a probabilistic similarity search algorithm. ATLAS combines a novel way of filtering candidate pairs of objects using truly random permutations of the vectors with their processing through inverted indexes. At a 97.5% recall rate, our algorithm consistently outperforms all state-of-the-art approaches and achieves a speed-up of up to two orders of magnitude over both exact and probabilistic algorithms.

### **Schema-Agnostic Indexing**

More and more data is accumulated inside social networks. Keyword search provides a simple interface for exploring this content. However, a lot of the content is private, and a search system must enforce the privacy settings of the social network.

We developed a workload-aware keyword search system with access control based on a social network. We made two technical contributions: (1) HeapUnion, a novel union operator that improves processing of search queries with access control by up to a factor of two compared to the best previous solution; and (2) highly accurate cost models that vary in sophistication and accuracy. These cost models provide input to an optimization algorithm that selects the most efficient organization of access control meta-data for a given workload. Our experimental results with real and synthetic data show that our approach outperforms previous work by up to a factor of three.

### **Sub-project 3: Scalable infrastructure for push and pull based computing**

#### **UiTø:**

We are investigating radically different approaches to cloud infrastructures, in particular enterprise clouds for search derived applications. We are interested in the whole software stack, from the operating system and light-weight virtualization technologies, to disruptive applications. In summary, progress we have made the last year includes the following:

#### **Enterprise search run-times**

Hadoop is de facto research and industry standard for the moment for analytic run-times. We have finalized the development of Cogset, an alternative to Hadoop, but with the same interface.

Cogset is an efficient and generic engine for reliable storage and parallel processing of data. It supports a number of high-level programming interfaces, including a MapReduce interface compatible with Hadoop. We have evaluated Cogset's performance as a MapReduce engine, comparing it to Hadoop. Our results show that Cogset generally outperforms Hadoop by a significant margin. We have investigated the underlying causes of this difference in performance and have demonstrated some relatively minor modifications that markedly improve Hadoop's performance, closing some of the gap.

#### **Enterprise cloud run-time**

We have been (and still are) building and evaluating run-time systems for private cloud computing environments. Such enterprise clusters must dynamically support resource elasticity through private cloud sharing, but also in combination with public clouds to timely handle peak resource demands. Performance isolation is fundamental in this context; our conjecture and initial experiences are that existing virtualization platforms have room for improvements in this context. Hence, we are building and evaluating prototypes supporting performance isolation.

We have been developing the *kernlet* idea, a light-weight approach to virtualization aiming for a minimalistic resource footprint. Kernlets challenge the basic approach of current cloud virtualization technologies: instead of hardware interfaces, operating system abstractions should be virtualized. Our conjecture is that by raising the abstraction level, the hypervisor can enact optimizations that are very difficult to achieve with virtual machines running commodity operating systems that assume singular control over the underlying (virtualized) platform. Still, for kernlets to be viable as an alternative to current virtualization technologies, all benefits of said technology must be met or exceeded by kernlets. We have developed our first kernlets and are currently carrying out thorough experiments.

#### **Video search, composition and streaming**

We have also investigated novel application ideas. One such is Davvi, a video search application that composes and streams highly personalized videos in response to traditional queries. Davvi does more than just returning the most relevant results

obtained from an inverted index, it actually extracts video events from across a huge archive of larger videos and composes them into a smooth play-out. The application scenario is soccer, but we have also built this video search capability into the latest enterprise search system from Microsoft, vESP.

Watching video playouts on mobile devices is becoming increasingly popular. Users that access these video archives from mobile devices while moving about, might not be able to efficiently navigate a large search output with a limited screen size and input control. Simply playing back the search output is not a viable option since the user might not have sufficient time or resources available to watch all matching video clips. Consequently, without modifications to the video-search output, the user might miss the video clips that are most important to him. We have also investigated how our internet video archival system, Davvi, best can support access for mobile users. In particular, we have investigated mechanisms for adapting the length of a search generated video playlist to a user specified time constraint.

### UiO:

We are working mainly in TASK 3 of the Annual Working Plan in close collaboration with Tromsø and Dubin building a multimedia infrastructure with run-time support for search and data delivery. Additionally, we have several other ongoing topics:

**P2G:** The computational demands of multimedia data processing are steadily increasing as consumers call for progressively more complex and intelligent multimedia services. New multi-core hardware architectures provide the required resources, but writing parallel, distributed applications capable of benefiting from such resources remains a labored task compared to its sequential counter-part. For this reason frameworks like Google's MapReduce and Microsoft's Dryad were developed, as they allow the developer to think sequentially, yet benefit from parallel and distributed execution. An inherent limitation in the design of these batch-processing frameworks is their inability to express arbitrarily complex workloads. The dependency graphs of the frameworks are often limited to directed acyclic graphs, or even pre-determined stages. This is particularly problematic for video encoding and other algorithms that depend on iterative execution.

Therefore, we *started* working on P2G, a framework designed specifically for processing distributed real-time multimedia data. P2G supports arbitrarily complex dependency graphs with cycles, branches and deadline. P2G is implemented to scale transparently with available resources. Additionally, P2G will support heterogeneous computing resources, such as x86 and GPU processing cores. We have implemented a P2G kernel language to ease with the development effort.

**vESP:** vESP provides a novel and potentially disruptive multimedia service by modifying a widely deployed commercial enterprise search engine. The idea is to transparently integrate rich multimedia data with traditional textual-oriented query results. This includes that the search engine automatically discovers and extracts relevant scenes from a large knowledge repository of existing videos and produces a new, customized video of events matching the user query. To evaluate our prototype,

we have performed experiments using a data set from a knowledge repository in Microsoft consisting of PowerPoint presentations with corresponding videos. Our initial results demonstrate that such integration can be implemented efficiently, and that potential users prefer to have the opportunity to enrich the search results with corresponding video.

#### **Sub-project 4: Extreme precision & recommendation in multimedia access**

#### **Sub-project 4: Extreme Precision and Recommendation in Multimedia Access**

##### **T 4.1 User Profiling, Matching and Recommendation**

###### **T4.1.1 Asynchronous CIR: Content Recommendation**

We have worked on developing mobile sports search and recommendation engine that optimises the result document to match the current user context and needs, as defined by a user profile and contextual query. For more details, please see mDavvi below.

##### **T4.2 Calculating Information Importance**

###### **T4.2.1 Information Importance**

We have worked with UiTø on developing a prototype video search system (Davii) operating over soccer footage. Initially we have found that external manual sources are still more efficient than the automatic approaches, which is inline with other findings in related research. David Scott and David Sadlier have extended this work by analysing the audio track of field sports content to detect the importance of every video shot and to both align the sports content with the external sources and subsequently to accurately segment the video content into a sequence of important events. We are extending this work in 2011 to the generic broadcast video domain.

###### **T4.2.3 MultiModal Interaction**

We have developed a mobile interface to Davii, an early version of which was presented at the demo session of MMM2010 in January 2010. This prototype has since been extended and improved in 2010 so as to improve the search performance and the video playback streaming. Associated with this work we have been refining a new relational model for mobile device access that takes into account recent developments such as capacitive touch screens, larger displays, faster processors and the advent of new devices such as the iPad. The landscape for mobile device access to multimedia content is constantly changing and we are modifying our rationale accordingly. This work is the primary focus of David Scott and Colum Foley.

##### **T4.3 Scalable Semantic Annotation for Multimedia**

Continuing the initial work on sports event segmentation, Jinlin Guo has been working on a tool for identification of on-screen graphics for sports events to solve the problem of missed goals from sports event summaries. Initial results from the screen graphics detection and text location are very promising and now the work turns to the OCR of the text in order to identify the actual scores and messages on screen. This work is to be presented at MMM 2011, in January 2011. In addition, Jinlin has developed an SVM-based content classifier tool to automatically identify visual concepts from generic video content to provide the underlying video search engine for the DCU TRECvid video search prototype, and future video search systems. This tool has been successfully employed to identify over 20 visual

concepts. Attention is now turning to the temporal arrangement of the concepts and how this can aid the detection process for complex real-world events.

#### **T4.3.2 Scalable Semantic Indexing**

Initial exploration of the challenges to be addressed. We are focussing on extending the Multimedia Semantics work (classifier) for TRECVID into a cloud-computing environment. This will produce a scalable and extensible visual feature annotation tool, based on utilising a machine learning technique and operating over the output of a number of image feature primitives. We are working with UiT and UiO on issues of scale for multimedia processing. It is intended to work with two MSc student in early 2011 on actually implementing our visual processing technologies in the cloud.

#### **T4.3.3 Multimedia Archive Demonstrators**

The work on Multimedia archives ties together all other work strands in SP 4. Work has focussed in 2010 on the development of the TRECVID video search system for handheld devices (see 4.2.2); an extensible system that it is planned to reuse in 2011. In addition, we have been involved in the development of Davvi and mDavvi (T4.1.3). Finally we are working closely with Dr Margaret Farren, another DCU academic from the Dept. of Education Studies, in developing a video search framework (Físteoir) that will support novel integration of video into the teaching environment at DCU. We have got access to a large 60TB archive space and have begin to put in place the infrastructure needed to populate this archive. The major benefit of Físteoir is that we can get a large number of real-world users for our recommendation experimentation, as well as supporting pedagogy within both DCU and later our iAD partner universities. This work is ongoing and we have recently begun working with Peter Tiernan, who's PhD is in the area of video in the classroom. Finally we are planning to submit a proposal to the DCU Learning Innovation Fund to additionally support this research.

In 2010, DCU took part in the annual TRECVID video search evaluation. We have developed an effective state-of-the-art WWW video search system which combines both text based and content-based search facilities and operated on iPads and iPhones. The iPad version was evaluated using both novice (BI) and expert (DCU) users and shown to perform above the average when compared to all other participating groups, notwithstanding the obvious challenges of searching through video archives on a handheld device. In addition to simply partaking in TRECVID, DCU iAD have been actively involved in the coordination of the worldwide evaluation effort by; taking part in an initial annotation round to develop a groundtruth; Colum Foley developing the real-time result checker web service that participating groups will use; and Colum Foley helping to crawl the video collection that all participating groups will use.

## **Sub-project 5: Understanding and managing the disruptive potential of iAD**

### **Disruptive technology, search strategy**

**T5.1.1 Information access technology as a disruptive innovation:** Work continues on documenting and understanding the disruptive potential of information access technology.

Milestones:

**M5.1.1.1 Overview paper submitted**

*Work continues*

**M5.1.1.2 Write paper on disruptive technology in the Norwegian music industry**

*Work continues, first draft set up, scheduled presentation for 4/4.*

**M5.1.1.3 M.Sc. thesis on disruption in book publishing produced.**

*Students recruited, project approved, initial interviews done.*

**M5.1.1.4 Recruit a new major partner for the iAD project, based on rich media search and display application/architecture.**

*Work continues. Had meetings with regards to planning for EU application for whole project.*

**T5.2.1 The impact of information access technology on society, culture and behavior:** Behavior data on how users use iAT (in the form of search use records) are available from the business partners of the iAD project. An important initial task for the research group is to evaluate the data and determine how it can be a basis for empirically oriented studies, including issues of quality and privacy.

Milestones:

**M5.2.1.1 Search type paper (user intent) finished**

*In progress*

**M5.2.1.2 Draft paper on search terms as reference submitted**

*Not yet started, will probably start Fall 2011*

Activities:

- Task 5.3 “Economics of iAT” focuses on two main topics, the development of, and the use of, information access technologies, respectively, applying economic theories and empirical research methods to analyse the development and use of information access technologies. Milestones M5.3.1-M5.3.4 have been reached
- Task 5.3.1 Scale economies of the development of information access technologies (to be continued)
- Task 5.3.2 Productivity of the development of information access technologies (to be continued)
- CANCELLED Task 5.3.3 User behaviour using search engines (to be modified); Cancelled because of withdrawal of Schibsted. One finished MSc

thesis by Ole Martin Kjørstad and Nils Marius Sørli. MSc thesis by Sergey Hambardzumyan is still in progress.

- NEW Task 5.3.4 User behaviour and productivity in Multi-tasking information retrieval. Empirical studies of multi-tasking behaviour and how it affects user productivity. Since the 1990s psychologists and neural scientists have studied the limits of human multitasking. Most studies find that the divided and/or switching attention have implications for the way people think, socialize, and perform. We will study the implications for the business community. Ingunn Myrtveit and Ph d student Vedrana Jez had meeting with Opera Software 14.01.2011 – and reached agreement on collaboration for case study.
- NEW Task 5.3.5 “Economic behaviour in the bunker oil market of suppliers and customers”. Newly established cooperation between BI and DNV. We will combine economic theory and **data mining** of DNV repositories; explore, apply, and validate data mining and analysis technologies including search technologies, machine learning methods, text analytics, and fuzzy logic. Data mining and knowledge discovery **across** large, separate, isolated data repositories of vessel, engine damage, and bunker oil data with the aim of identifying nuggets of gold that add business value to DNV’s customers. Ingunn Myrtveit and 6 MSC students, started July 2010: Vegard Lien Kvelstad, Christoffer Lutken, Magnus Nøkleby, Carl Fredrik Schrøder, Andreas Boasson og Gjert Moberg . Workshops arranged with DNV R&I 25.11.10, 15.2.11 and 5.4.11. Students work with DNV R&I every Thursday if possible.

#### **T5.4.1 From search technology cluster to knowledge hub**

A research project called Knowledge-based Norway has been started in the Spring of 2010, to study the knowledge-based cluster development of various industries in Norway. As part of this, work will be done tracing the development of search technology as well, using the case of FAST in comparison to more traditional Norwegian IT firms.

M5.4.1.1: Much time spent on this, but in context of A Knowledge-based Norway project. Report due during March 2011.

## Appendix A: Personnel at the academic partners

<b>Key Personnel (NTNU)</b>					
<b>Key Researchers</b>					
Name	Institution	Main research area			
Svein Erik Bratsberg	NTNU	Similarity search, scalable indexing			
Svein-Olaf Hvasshovd	NTNU	Next generation search engine			
Magnus Lie Hetland	NTNU	Similarity search			
John Krogstie	NTNU	Semantic and location-based search			
Jon Atle Gulla	NTNU	Semantic and location-based search			
Øystein Torbjørnsen	NTNU/ Microsoft	Next generation search engine			
<b>Visiting Researchers</b>					
Name	Affiliation	Nationality	Sex M/F	Duration	Topic
<b>Postdoctoral researchers with financial support from the Centre budget</b>					
Name	Nationality	Period	Sex M/F	Topic	
Stein Løkke Tomassen	Norwegian	12 months	M	Semantic and location-based search	
<b>Postdoctoral researchers working on projects in the centre with financial support from other sources</b>					
Name	Nationality	Period	Sex M/F	Topic	
<b>PhD students with financial support from the Centre budget</b>					
Name	Nationality	Period	Sex M/F	Topic	
Simon Jonassen	Norwegian	2008-2012	M	Scalable distributed indexes	
Mohammad Ali Norozi (50%)	Pakistani	2009-2013	M	Relevancy in structured data	
Truls Amundsen Bjørklund	Norwegian	6 months	M	Indexing/search in column stores	
Nils Grimsmo	Norwegian	6 months	M	XML indexing and evaluation	
<b>PhD students working on projects in the centre with financial support from other sources</b>					
Name	Funding	Nationality	Period	Sex M/F	Topic
Bilegsaikhhan Naidan	NTNU	Mongolian	2009-2012	M	Approx. similarity search
Mohammad Ali Norozi	NTNU 50%	Pakistani	2009-2013	M	Relevancy in semistructured data
Ole Edsberg	NTNU	Norwegian	2009-2013	M	Improving distance-based indexing with machine learning
Hongli Yang	NTNU	Chinese	2009-2013	M	Similarity search
Jon Marius Venstad	NTNU	Norwegian	2009-2013	M	Similarity search
<b>Master degrees</b>					
Name	Sex M/F	Topic			
Knut Moen Sælø	M	Evaluation of database used in distributed systems			
Ola Natvig	M	Compression in XML search engines			

Annual report 2010

iAD – Information Access Disruptions Centre for Research-based Innovation at  
Fast Search & Transfer AS

Eirik Reksten	M	Evaluation of spatial and metric indexing
Kristian Veøy	M	Similarity search (to be finished in 2011)
Håkon Hitland	M	Similarity search (to be finished in 2011)
Vegard Økland	M	Similarity search (to be finished in 2011)
Magnus Haug	M	Similarity search (to be finished in 2011)
Njål Karevoll	M	Search in the cloud (to be finished in 2011)
Stian Erlandsen	M	Similarity search (to be finished in 2011)

## Key Personnel (UiTø)

### Key Researchers

Name	Institution	Main research area
Dag Johansen	UiTø	Distributed systems
Åge Kvalnes	UiTø	Operating systems

### Visiting Researchers

Name	Affiliation	Nationality	Sex M/F	Duration	Topic
Cathal Gurrin (visiting researcher 20%)	Dublin CU	Irish	M	07-11	Video search
Johannes Gehrke, assoc. prof II (20%)	Cornell Univ	German	M	08-12	Stream processing

### Postdoctoral researchers

Name	Nationality	Period	Sex M/F	Topic
Håvard Johansen	Norwegian	08-11	M	Cloud computing

### PhD students with financial support from the Centre budget

Name	Nationality	Period	Sex M/F	Topic
Steffen V. Valvåg	Norwegian	05-10	M	Efficient distributed data processing
Audun Arnesen	Norwegian	10-13	M	Light-weight virtual machines
Joseph Hurley	US	10-14	M	Cloud computing

### PhD students working on projects in the centre with financial support from other sources

Name	Funding	Nationality	Period	Sex M/F	Topic
Kjetil Jacobsen	Private/MSFT	Norwegian	05-11	M	Fault-tolerance

### Master degree students

Name	Sex M/F	Topic
Øyvind Holmstad	M	Real-time video streaming
Erik Bræck Leer	M	Video selection algorithms
Johan Grønvik	M	Cloud Computing
Robert Pettersen	M	Operating systems

<b>Key Personnel (UiO)</b>					
<b>Key Researchers</b>					
Name	Institution	Main research area			
Carsten Griwodz	UiO	Task 3			
Pål Halvorsen	UiO	Task 3			
<b>Visiting Researchers</b>					
Name	Affiliation	Nationality	Sex M/F	Duration	Topic
<b>Postdoctoral researchers</b>					
Name	Nationality	Period	Sex M/F	Topic	
Andreas Petlund	N	2010	M	Thin streams	
<b>PhD students with financial support from the Centre budget</b>					
Name	Nationality	Period	Sex M/F	Topic	
Håvard Espeland	N	2010	M	Distributed processing	
Tomas Kupka	Slovakia	2010	M	Stream scheduling, 3D	
<b>PhD students working on projects in the centre with financial support from other sources</b>					
Name	Funding	Nationality	Period	Sex M/F	Topic
Paul Beskow	IFI	N	2010	M	Scheduling
Pengpeng Ni	IFI	Chinese	2010	F	Video quality assessment
Håkon Kvale Stensland	IFI	N	2010	M	Distributed processing
<b>FINISHED Master degree students in 2010</b>					
Name	Sex M/F	Topic			
Martin Wam	M	Cell programming			
Espen Jacobsen	M	Stream scheuling			

<b>Key Personnel (BI)</b>					
<b>Key Researchers</b>					
Name	Institution	Main research area			
Associate Professor Espen Andersen	BI	iAT as disruption			
Professor Ingunn Myrtveit	BI	Economics of iAT			
Adjunct Professor Erik Stensrud	BI	Economics of iAT			
<b>Visiting Researchers</b>					
Name	Affiliation	Nationality	Sex M/F	Duration	Topic
<b>Postdoctoral researchers with financial support from the Centre budget</b>					
Name	Nationality	Period	Sex M/F	Topic	
<b>Postdoctoral researchers working on projects in the centre with financial support from other sources</b>					
Name	Nationality	Period	Sex M/F	Topic	
<b>PhD students with financial support from the Centre budget</b>					
Name	Nationality	Period	Sex M/F	Topic	
Vedrana Jez		2010-2014	F	Economics of iAT	
Yang Zhenyu		2009-2013	M	iAT as disruption	
<b>PhD students working on projects in the centre with financial support from other sources</b>					
Name	Funding	Nationality	Period	Sex M/F	Topic
<b>Master degrees</b>					
	Sex M/F	Topic			
Ole Martin Kjørstad (completed)	M	Economics of iAT			
Nils Marius Sørli (completed)	M	Economics of iAT			
Sergey Hambardzumyan (started fall 2010)	M	Economics of iAT			
Vegard Lien Kvelstad (started fall 2010)	M	Economics of iAT			
Christoffer Lutken (started fall 2010)	M	Economics of iAT			
Magnus Nøkleby (started fall 2010)	M	Economics of iAT			
Andreas Boasson (started fall 2010)	M	Economics of iAT			
Gjert Moberg (started fall 2010)	M	Economics of iAT			
Carl Fredrik Schrøder (started fall 2010)	M	Economics of iAT			
Anders Gjermshus (started fall 2010)	M	iAT as disruption			
Hilde Marie Wold (started spring 2011)	F	iAT as disruption			
Vibeke Fugle Mailund (started spring 2011)	F	iAT as disruption			
Hani Khoury (started spring 2011)	M	iAT as disruption			
Janneke Støeng (started spring 2011)	F	iAT as disruption			
Andrea Felloni (started spring 2011)	M	iAT as disruption			
Mona Munthe-Kaas (started spring 2011)	F	iAT as disruption			

<b>Key Personnel (Cornell)</b>					
<b>Key Researchers</b>					
Name	Institution	Main research area			
Johannes Gehrke	Cornell University	Database systems, search			
Robbert VanRenesse	Cornell University	Distributed systems			
Fred Schneider	Cornell University	Distributed systems, security			
Christoph Koch	Cornell University	Database systems			
<b>Visiting Researchers</b>					
Name	Affiliation	Nationality	Sex M/F	Duration	Topic
Dag Johansen	Tromso	Norwegian	M	2009-2010	
Age Knaflves	Tromso	Norwegian	M	2009-2010	
<b>Postdoctoral researchers with financial support from the Centre budget</b>					
Name	Nationality	Period	Sex M/F	Topic	
<b>Postdoctoral researchers working on projects in the centre with financial support from other sources</b>					
Name	Nationality	Period	Sex M/F	Topic	
Marcos Vaz Salles	Brasilian	2010	M	Fault tolerance	
Lucja Kot	US	2010	F	Collaboration	
<b>PhD students with financial support from the Centre budget</b>					
Name	Nationality	Period	Sex M/F	Topic	
Tuan Cao	Vietnamese	2010	M	Checkpoint recovery	
Michaela Goetz	German	2010	F	Search	
<b>PhD students working on projects in the centre with financial support from other sources</b>					
Name	Funding	Nationality	Period	Sex M/F	Topic
Ben Sowell	Fellowship	US	2010	M	Indexing
Yin Lou	Fellowship	Chinese	2010	M	Image search
Sudip Roy	Fellowship	Indian	2010	M	Image search
Guozhang Wang	NSF	Chinese	2010	M	Efficient processing
<b>Master degrees</b>					
Name	Sex M/F	Topic			
None					

<b>Key Personnel (DCU)</b>					
<b>Key Researchers</b>					
Name	Institution	Main research area			
Dr Cathal Gurrin	DCU	Information Retrieval			
<b>Visiting Researchers</b>					
Name	Affiliation	Nationality	Sex M/F	Duration	Topic
Havard Johansen	UiT	Norwegian	M	1 M	Search
Joe Hurley	UiT	American	M	2 W	Search
<b>Postdoctoral researchers with financial support from the Centre budget</b>					
Name	Nationality	Period	Sex M/F	Topic	
Peter Wilkins	Australian	Jan-Apr	M	Content Analysis	
Colum Foley	Irish	Apr - Dec	M	Video Search	
Paul Ferguson	Irish	July-Aug	M	Interfaces	
<b>Postdoctoral researchers working on projects in the centre with financial support from other sources</b>					
Name	Nationality	Period	Sex M/F	Topic	
David Sadlier	Irish	July-Sept	M	Sports Summarisation	
Hyowon Lee	Korean	periodic	M	Interfaces	
<b>PhD students with financial support from the Centre budget</b>					
Name	Nationality	Period	Sex M/F	Topic	
David Scott	Irish	Apr 09 – Oct 12	M	User Interaction	
Jinlin Guo	Chinese	Jan 10 – Apr 13	M	Concept Detection	
<b>PhD students working on projects in the centre with financial support from other sources</b>					
Name	Funding	Nationality	Period	Sex M/F	Topic
<b>Master degrees</b>					
Name	Sex M/F	Topic			
Emma Semsero Diaz	F	Interfaces for Video Search on Handheld Devices			

## Appendix B: Statement of Accounts

(All figures in 1000 NOK)

<b>Funding 01.01.2010 - 31.12.2010</b>		
<b>Type of funding</b>	<b>Funding in period (1000 NOK)</b>	<b>Group</b>
The Research Council	11 645,47	
The Host Institution (FAST)	9 078,45	
Public Partners	9 727,72	Norwegian University of Science and Technology, University of Tromsø, University of Oslo
Private (including enterprise) partners	5 845,69	Accenture, Schibsted, BI Norwegian School of Management
International partners	1 522,69	Collaboration partners Cornell University, Dublin City University
<b>Sum of funding</b>	<b>37 820,03</b>	

<b>Costs 01.01.2010 - 31.12.2010</b>		
<b>Type of cost</b>	<b>Cost in period (1000 NOK)</b>	<b>Group</b>
The Host Institution (FAST)	13 152,52	FAST, Collaboration partners Cornell University, Dublin City University
Public Partners	16 425,73	Norwegian University of Science and Technology, University of Tromsø, University of Oslo
Equipment	431,80	
Private (including enterprise) partners	7 809,99	Accenture, Schibsted, BI Norwegian School of Management
<b>Sum of costs</b>	<b>37 820,03</b>	

## Appendix C: Publications

---

### Journal Papers

1. Guozhang Wang, Marcos Antonio Vaz Salles, Benjamin Sowell, Xun Wang, Tuan Cao, Alan J. Demers, Johannes Gehrke, Walker M. White: Behavioral Simulations in MapReduce. *PVLDB* 3(1): 952-963 (2010)
2. Johannes Gehrke: Programming with differential privacy: technical perspective. *Commun. ACM* 53(9): 88 (2010)
3. Robbert van Renesse and Rachid Guerraoui. Replication Techniques for Availability. *Replication: Theory and Practise*. Bernadette Charron-Bost, Fernando Pedone, and Andre Schiper (ed.). Vol. Lecture Notes in Computer Science 5959. Springer-Verlag, Heidelberg. 2010.
4. Michael R. Clarkson, Fred B. Schneider: Quantification of Integrity. *CSF 2010*: 28-43
5. Fred B. Schneider, Lidong Zhou: Implementing Trustworthy Services Using Replicated State Machines. *Replication 2010*: 151-167
6. Michael R. Clarkson, Fred B. Schneider: Hyperproperties. *Journal of Computer Security* 18(6): 1157-1210 (2010)
7. Riccardo Pucella, Fred B. Schneider: Independence from obfuscation: A semantic framework for diversity. *Journal of Computer Security* 18(5): 701-749 (2010)
8. Tom Roeder, Fred B. Schneider: Proactive obfuscation. *ACM Trans. Comput. Syst.* 28(2): (2010)
9. Fred B. Schneider: Fumbling the Future, Again. *IEEE Security & Privacy* 8(4): 3 (2010)
10. João Leitão, Robbert van Renesse, and Luís Rodrigues. Balancing Gossip Exchanges in Networks with Firewalls. *International Workshop on Peer-to-Peer Systems (IPTPS 2010)*. San Jose, CA. April 2010.
11. Michael Chow and Robbert van Renesse. A Middleware for Gossip Protocols. *International Workshop on Peer-to-Peer Systems (IPTPS 2010)*. San Jose, CA. April 2010.
12. "Reducing game latency by migration, core-selection and TCP modifications" Paul B Beskow, Andreas Petlund, Geir A. Erikstad, Carsten Griwodz, Pål Halvorsen *The International Journal of Advanced Media and Communication*, Vol. 4, No. 4, November 2010, p 343 – 363
13. "Cheat Detection Processing: A GPU versus CPU Comparison" Håkon Kvale Stensland, Martin Myrseth, Carsten Griwodz, Pål Halvorsen *TO APPEAR IN Proceedings of the workshop on network and systems support for games (NetGames)*, Taipei, Taiwan, November 2010
14. "Pull-Patching: A Combination of Multicast and Adaptive Segmented HTTP Streaming" (short) Espen Jacobsen, Carsten Griwodz, Pål Halvorsen *Proceedings of the ACM International Multimedia Conference (ACM MM)*, Florence, Italy, October 2010, pp 799 - 802
15. "vESP: Enriching Enterprise Document Search Results with Aligned Video Summarization" (demo) Pål Halvorsen, Dag Johansen, Bjørn Olstad, Tomas Kupka, Sverre Tennøe *Proceedings of the ACM International Multimedia Conference (ACM MM)*, Florence, Italy, October 2010, pp. 1603 - 1604

16. "vESP: A Video-Enabled Enterprise Search Platform" Pål Halvorsen, Dag Johansen, Bjørn Olstad, Tomas Kupka, Sverre Tennøe Proceedings of the International Conference on Data and Knowledge Engineering (ICDKE), Melbourne, Australia, September 2010, pp. 534-541
17. "Frequent Layer Switching for Perceived Quality Improvements of Coarse-Grained Scalable Video" Pengpeng Ni, Alexander Eichhorn, Carsten Griwodz, Pål Halvorsen Springer Multimedia Systems Journal, Vol. 16, No. 3, June 2010, pp. 171 – 182
18. "Quality-Adaptive Scheduling for Live Streaming over Multiple Access Networks" Kristian Evensen, Tomas Kupka, Dominik Kaspar, Pål Halvorsen, Carsten Griwodz Proceedings of the Network and Operating System Support for Digital Audio and Video (NOSSDAV), Amsterdam, The Netherlands, June 2010, pp. 21 – 26
19. "Tips, Tricks and Troubles: Optimizing for Cell and GPU" Håkon Kvale Stensland, Håvard Espeland, Carsten Griwodz, Pål Halvorsen Proceedings of the Network and Operating System Support for Digital Audio and Video (NOSSDAV), Amsterdam, The Netherlands, June 2010, pp. 75 - 80
20. "Low Overhead Container Format for Adaptive Streaming" Haakon Riiser, Pål Halvorsen, Carsten Griwodz, Dag Johansen Proceedings of the ACM Multimedia Systems (MMSys), Scottsdale, AZ USA, February 2010, pp. 193 - 198
21. "Program obfuscation by strong cryptography" Zeljko Vrba, Pål Halvorsen, Carsten Griwodz Proceedings of the International Conference on Availability, Reliability and Security (ARES), Krakow, Poland, February 2010, pp. 242 – 247
22. "A simple improvement of the work-stealing scheduling algorithm" Zeljko Vrba, Pål Halvorsen, Carsten Griwodz Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems (CISIS) - International Workshop on Multi-Core Computing Systems (MuCoCoS), Krakow, Poland, February 2010, 925 - 930
23. "Search in Social Networks with Access Control", Truls A. Bjørklund, Michaela Götz, Johannes Gehrke, Proceedings of KEYS' 2010 The Second International Workshop on Keyword Search on Structured Data, SIGMOD 2010. ISBN: 978-1-4503-0187-9, June 6, 2010
24. "Towards Unifying Advances in Twig Join Algorithms", Nils Grimsmo and Truls Amundsen Bjørklund, Proceedings of the 21st Australasian Database Conference (ADC 2010), January 18-22, 2010
25. "XLeaf: Twig Evaluation with Skipping Loop Joins and Virtual Nodes", Nils Grimsmo, Truls Amundsen Bjørklund and Øystein Torbjørnsen. Proceedings of the Second International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2010), April 11-16, 2010.
26. "Fast Optimal Twig Joins", Nils Grimsmo, Truls Amundsen Bjørklund and Magnus Lie Hetland, Proceedings of 36th International Conference on Very Large Data Bases (VLDB 2010), Singapore, September 13-17, 2010.
27. "Linear Computation of the Maximum Simultaneous Forward and Backward Bisimulation for Node-Labeled Trees", Nils Grimsmo, Truls Amundsen Bjørklund and Magnus Lie Hetland, Proceedings of the 7th International XML

- Database Symposium on Database and XML Technologies (XSym 2010), September 17, 2010.
28. "Dynamic optimization of queries in pivot-based indexing", Svein Erik Bratsberg and Magnus Lie Hetland, Multimedia Tools and Applications (Springer Journal), October 2010.
  29. "A Combined Semi-Pipelined Query Processing Architecture for Distributed Full-Text Retrieval", Simon Jonassen and Svein Erik Bratsberg, The 11th International Conference on Web Information Systems Engineering, WISE 2010, Hong Kong, December 12-14, 2010. Best paper award.
  30. "Extrapolation to Speed-up Query-dependent Link Analysis Ranking Algorithms", Muhammad Ali Norozi, Frontiers of Information Technology 2010, Islamabad, Pakistan, December 21-23, 2010.
  31. "Constructing Feature Vectors for search: investigating intrinsic quality impact on search performance", Stein L. Tomassen and Darijus Strasunskas, International Journal on Web and Grid Services, Vol. 6, No. 3, pp. 289-312.
  32. "Indexing Inexact Proximity Search with Distance Regression in Pivot Space", Ole Edsberg and Magnus Lie Hetland. In Proceedings of the Third International Conference on Similarity Search and Applications, ACM, September 18-19, 2010.
  33. "Searching and Recommending Sports Content on Mobile Devices", David Scott, Cathal Gurrin, Dag Johansen and Håvard Johansen.. In proc. of the 16th International Conference on MultiMedia Modelling, LNCS volume 5916/2010, pages 779-781, Springer, January 8, 2010, Chongqing, China.
  34. "Summarized-Search: The Fusion of Summarization and Search", David Scott, Cathal Gurrin, Peter Wilkins, Colum Foley, iHCI 2010.
  35. "TRECVID 2010 Experiments at Dublin City University". Foley C, Guo J, Scott D, Wilkins P, Gurrin C, Smeaton A.F, Ferguson P, McCusker K, Diaz E, Mc Guinness K, O'Connor N, Giro-i-Nieto X and Margues F. TRECVID 2010 - Text REtrieval Conference TRECVID Workshop, Gaithersburg, MD, 15-17 November 2010
  36. "Proceedings of the second annual SenseCam symposium (SenseCam 2010)", Berry, Emma and Byrne, Daragh and Doherty, Aiden R. and Gurrin, Cathal and Smeaton, Alan F., eds. (2010). 16-17 September 2010. Dublin City University, Dublin, Ireland. ISBN 1872-327-915
  37. "Information access for personal media archives. (2010)", Doherty, Aiden R. and Gurrin, Cathal and Jones, Gareth J.F. and Smeaton, Alan F. SIGIR Forum, Special Interest Group on Information Retrieval.
  38. "Advances in Information Retrieval: 32nd European Conference on IR Research", Cathal Gurrin, Yulan He, Gabriella Kazai, Udo Kruschwitz, Suzanne Little, Thomas Roelleke, Stefan Rüger, Keith van Rijsbergen. ECIR 2010, Milton Keynes, UK, March 28-31, 2010. Proceedings. Berlin: Springer, p. 627-630. (Lecture Notes in Computer Science ; 5993)
  39. "Introduction to Video Search Engines". (Book Review). Journal of Information Retrieval. Vol. 13, Num. 4, (pp398-402) Colum Foley. Information Retrieval.
  40. "Division of Labour and Sharing of Knowledge for Synchronous Collaborative Information Retrieval". Colum Foley and Alan F. Smeaton. Journal of Information Processing & Management. Vol. 46, Num. 6, Pages 762-772.
  41. "Low Overhead Container Format for Adaptive Streaming", Haakon Riiser, Pål Halvorsen, Carsten Griwodz, and Dag Johansen. In proc. of the 2010 ACM

- Conference on Multimedia Systems (MMSys), pp. 193-198, February 2010, Scottsdale, AZ, USA.
42. "Composing Personalized Video Playouts Using Search". Dag Johansen, Håvard Johansen, Pål Halvorsen, Bjørn Olstad, Cathal Gurrin, and Carsten Griwodz. In Proc. of the 2010 IEEE International Conference on Multimedia & Expo, July, 2010, Singapore, pp. 1534 - 1539.
  43. "Simple, yet Paradoxically Complex: Under the Search Surface". Dag Johansen and Bjørn Olstad, PerAda Magazine, August 2010.
  44. "Support for enterprise consolidation of I/O bound services". Åge Kvalnes, Dag Johansen, Pål Halvorsen and Carsten Griwodz, Journal of Software: Practice and Experience, Wiley, vol. 40, November/December 2010, pp. 1035 – 1051.
  45. Cogset vs Hadoop: Measurement and Analysis, Steffen Viken Valvåg, Dag Johansen, and Åge Kvalnes, In IEEE Proceedings of CloudCom 2010, MAPRED'2010 workshop, Indianapolis, USA, December 2010.
  46. "The Nornir run-time system for parallel programs using Kahn process networks on multi-core machines - A flexible alternative to MapReduce". Zeljko Vrba, Pål Halvorsen, Carsten Griwodz, Paul Beskow, Håvard Espeland and Dag Johansen. Journal of Supercomputing, Springer, 2010 (so far online).
  47. Andersen, E: Edging towards the semantic web: Protocols, curation and seeds. *ACM Ubiquity*, November 2010.
  48. Bjerkøe, Richard and Anders Sørbo: The Norwegian Music Industry in the Age of Digitalization, Master Thesis. (this thesis was widely covered on TV, radio and newspapers, and is still generating interest.)

## Reports/Others

### *Published essays, comments, etc.:*

- Andersen, E., Klikkhorer og papirtigre, essay in Kokkvold, P. (ed.) *Antifestskrift til Norsk Presseforbunds 100-årsjubileum*. IJ-forlaget 2010.
- Andersen, E: Det forlagte forleggeriet, *Dagbladet*, 24. september 2010.
- Andersen, E. Innovasjonstull?, *e24*, 18. august
- Andersen, E: Leksikonpolitikk, *Morgenbladet*, 26. mars 2010.
- Andersen, E: NAV: Et IT-problem, *e24*, 1. februar 2010.