Building Intelligent Clinical Decision Support Systems Using a Large Knowledge Base

Mingu Lee, Ph.D.
Vice President, Samsung SDS America

Dongwook Shin, Ph.D.
President, Tranformation
Samsung Group Overview

Samsung is the 19th most valuable brand in the world. Samsung Group’s total revenues of USD 173 billion corresponds to rank 14th on Fortune Global 500.

Employees
- 277,000 around the globe

Sales Revenue
- USD 173 Billion

Brand Value
- USD 17.5 Billion (Global Ranking 19th)

Global Presence
- Over 60 Countries

Global Ranking

<table>
<thead>
<tr>
<th>Rank</th>
<th>Company</th>
<th>Revenues ($ million)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>General Electric</td>
<td>183,207</td>
</tr>
<tr>
<td>13</td>
<td>China National Petroleum</td>
<td>181,123</td>
</tr>
<tr>
<td>14</td>
<td>Volkswagen</td>
<td>166,579</td>
</tr>
<tr>
<td>32</td>
<td>HP</td>
<td>118,364</td>
</tr>
<tr>
<td>81</td>
<td>Sony</td>
<td>76,945</td>
</tr>
</tbody>
</table>

(Best Global Brands 2009, Interbrand)

(Fortune Global 500, 2009)
Samsung Affiliates

Samsung is comprised of companies that are setting new standards in a wide range of businesses. Of its 28 affiliated companies, Samsung Electronics, Samsung Life Insurance and Samsung C&T are listed in the Fortune Global 500, 2009.

Samsung Electronics
- Ranked 40th, Fortune Global 500
- USD 96 billion in revenues
- No.1 market share in the world: DRAM, SRAM, Flash Memory
- 164,600 employees
- 179 offices in 61 countries
- No.2 in Mobile Phone
- Ranked 40th, Fortune Global 500
- No.1 market share in the world: DRAM, SRAM, Flash Memory
- 164,600 employees
- 179 offices in 61 countries
- No.2 in Mobile Phone

Samsung Heavy Industries
- World #1 Shipbuilder (based on order book)
- USD 9.7 billion in revenues
- 12,500 employees
- 105 offices in 44 countries
- Constructed world's tallest structure (Burj Khalifa, Dubai)
- USD 10.7 billion in revenues
- 7,200 employees
- 105 offices in 44 countries
- Constructed world's tallest structure (Burj Khalifa, Dubai)
- USD 10.7 billion in revenues
- 7,200 employees
- 105 offices in 44 countries
- Constructed world's tallest structure (Burj Khalifa, Dubai)
- USD 10.7 billion in revenues
- 7,200 employees
- 105 offices in 44 countries
- Constructed world's tallest structure (Burj Khalifa, Dubai)
- USD 10.7 billion in revenues
- 7,200 employees
- 105 offices in 44 countries

Samsung C&T (Construction & Trading)
- Constructed world's tallest structure (Burj Khalifa, Dubai)
- USD 10.7 billion in revenues
- 7,200 employees
- 105 offices in 44 countries

Samsung Life Insurance
- Largest insurance company in Korea
- USD 22.3 billion sales
- 6,377 employees
- 12 offices in 8 countries
- Largest insurance company in Korea
- USD 22.3 billion sales
- 6,377 employees
- 12 offices in 8 countries
- Largest insurance company in Korea
- USD 22.3 billion sales
- 6,377 employees
- 12 offices in 8 countries
- Largest insurance company in Korea
- USD 22.3 billion sales
- 6,377 employees
- 12 offices in 8 countries
- Largest insurance company in Korea
- USD 22.3 billion sales
- 6,377 employees
- 12 offices in 8 countries

(Samsung Profile, 2009)
Samsung SDS Overview

Samsung SDS is a global ICT service provider with over 11,000 employees and USD 3.2 billion in revenues. With an average annual growth rate of 10% over the last four years, it is ranked as the 3rd largest IT service company in Asia-Pacific region.

<table>
<thead>
<tr>
<th>Employees</th>
<th>11,678 Employees around the globe (1,824 Overseas)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales Revenue</td>
<td>USD 3.2 Billion</td>
</tr>
<tr>
<td>Global Presence</td>
<td>12 Offices in 12 Countries</td>
</tr>
<tr>
<td>Recognition</td>
<td>No. 1 Korean IT services provider with the largest domestic market share (14.7%)</td>
</tr>
<tr>
<td></td>
<td>3rd largest IT services provider in Asia-Pacific region</td>
</tr>
</tbody>
</table>

(Market Share: IT Services, Asia/Pacific, 2008, Gartner)

* Numbers are as of May, 2010

Revenue Growth

<table>
<thead>
<tr>
<th>Year (Projection)</th>
<th>Revenue (USD billion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>2.23</td>
</tr>
<tr>
<td>2006</td>
<td>2.52</td>
</tr>
<tr>
<td>2007</td>
<td>2.69</td>
</tr>
<tr>
<td>2008</td>
<td>3.17</td>
</tr>
<tr>
<td>2009</td>
<td>3.25</td>
</tr>
<tr>
<td>2010 (Projection)</td>
<td>3.82</td>
</tr>
</tbody>
</table>

CAGR = 11.4%*

* Includes 2010 projection
Service Offering Overview

**Smart Infrastructure Engineering (SIE)**
- Transportation
- Construction

**Enterprise Application Services**
- Business and ICT Consulting
- ERP/SCM/PLM/MES/CRM

**Mobile Communication Services**
- Mobile Service
- Unified Communication

**e-Government**
- Citizen Services
- National Services
- Government Services

**ICT Infrastructure**
- Data Center
- Network/Data Communication
- Cloud computing

**Business Process Outsourcing (BPO)**
- Testing Services
# Samsung SDS America Public Sector Capabilities

## Mobile
- Mobile Groupware
- SAP Mobile BI Dashboard
- Oracle/Siebel Mobile CRM for Pharmaceutical Sales
- Mobile Device Management
- Mobile Applications (Android OS)

## ERP
- DoD BTA ERP
- Samsung Electronics Global ERP (Largest ERP implementation in the world)
- Mobilizing ERP applications

## Health IT
- HHS NIH Clinical Decision Support Systems (CDSS) – Natural Language Processing

## Portfolio Management
- Oracle Primavera
- EPA Enterprise Architecture - Best Practices
- EPA Toxic Release Inventory - White Paper

## System Integration
- e-Procurement Systems (Vietnam and Costa Rica)
- Customs Services (Kazakhstan, Dominican Republic, & Mongolia)
- National Tax Service Networks (Sri Lanka)

## Security
- Kuwait Oil Facilities Integrated Security System (US $300 M)
- EPA Portal and Web Access Management
Mobile EMR by Samsung SDS

Mobile Electronic Medical Record (EMR) provides healthcare professionals instant access to patient information at the point of care.

<table>
<thead>
<tr>
<th>Features</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Search patient records</td>
<td>• Compact 7” tablet form factor significantly enhances mobility</td>
</tr>
<tr>
<td>• View clinical charts/ analytics</td>
<td>• Improve quality of service thru access to more accurate, timely information</td>
</tr>
<tr>
<td>• Make annotations</td>
<td>• Process automation thru mobility significantly improves data quality and enhances efficiency of hospital staff</td>
</tr>
<tr>
<td>• View lab results and findings</td>
<td></td>
</tr>
<tr>
<td>• View and compare images (e.g., X-ray film, CT Scan, Ultrasound, etc.)</td>
<td></td>
</tr>
<tr>
<td>• Always connected – data maintained securely on server</td>
<td></td>
</tr>
<tr>
<td>• Remotely manage devices (i.e., remote lock/wipe of data, black-list/white-list apps)</td>
<td></td>
</tr>
</tbody>
</table>
Project Overview

- U.S. Department of Health and Human Services (HHS) National Institutes of Health (NIH) National Library of Medicine (NLM)
- Clinical Decision Support Systems (CDSS) with Artificial intelligence (AI) approach to discover Adverse Drug Events accurately
- Performed by Samsung SDS America & Tranformation
While his doctor is out-of-town, an elderly asthma patient who has developed severe knee pain sees another physician in his doctor’s office. An EMR provided documentation of the last visit, including recent laboratory results and a list of the patient’s medications. This information easily brought the doctor up to date on the patient’s condition. The doctor entered an order for medicine for the knee pain into the system, printed out a (legible) prescription for the patient, and sent him on his way. Unfortunately, within 2 months, the patient wound up in the emergency room with a bleeding ulcer caused by interaction of the pain medicine with the patient’s asthma medicine.

- Referred From Agency for Healthcare Research and Quality (AHRQ)
Technical Components

- I. DailyMed: Drug information (package insert):
  - A website Operated by NLM to publish drug labels
  - Contents provided by the U.S. Food and Drug Administration (FDA)
- II. Natural Language Programming (NLP): Convert drug information in DailyMed to the corresponding logical forms
- III. Unified Medical Language System (UMLS): A medical ontology developed by NLM
  - MetaMap: A tool to find UMLS concepts from sentences
- IV. OpenCyc: A large knowledge base system with logical inference
- V. Google Health: Personal Health Record (PHR) system
- VI. Adverse Drug Event Detection Services
- VII. Android Mobile App & UI
Conceptual Architecture

External Services

- **Drug Information Services**
  - Drug Insert - DailyMed

- **UMLS Ontology Services**
  - MetaMap

- **Knowledge Management Services**
  - Knowledge base & reasoning engine - OpenCyc

- **Personal Health Record Services**
  - Google Health

Business Process Management

- NLP Services
  - NLP Services
  - Human Review

- **Adverse Drug Event Detection Services**
  - ADE Services

- **User Interface**
  - Web Application
  - Mobile – Android App
Conceptual Architecture

Conversion using NLP

Assertion from patient info

Assertion in logic form

OpenCyc

Inference

Alerting to patient/physician

Conversion

Review the accuracy

DailyMed

UMLS

Google Health based PHR
**WARNINGS**

**Stomach bleeding warning:**

*This product contains a nonsteroidal anti-inflammatory drug (NSAID), which may cause stomach bleeding. The chance is higher if you:*

- are age 60 or older
- have had stomach ulcers or bleeding problems
- take a blood thinning (anticoagulant) or steroid drug
- take other drugs containing an NSAID [aspirin, ibuprofen, naproxen, or others]
- have 3 or more alcoholic drinks every day while using this product
- take more or for a longer time than directed

---

**I. Drug Information Services**
Pre-Processing of DailyMed

- Using XML format of DailyMed SPL
  - Process “Warning”, “Boxed warning”, “Indication”, “Contraindication”, “Precaution” and “Adverse reaction” part
  - These parts can be identified by the corresponding LOINC code

```xml
<section>
  <code code="34071-1" codeSystem="2.16.840.1.113883.6.1" displayName="WARNINGS SECTION"/>
  <text>
    <paragraph> Stomach bleeding warning: This product contains an NSAID, which may cause severe stomach bleeding. The chance is higher if you
    <list>
      <item>are age 60 or older</item>
      <item>have had stomach ulcers or bleeding problems</item>
      <item>take a blood thinning (anticoagulant) or steroid drug</item>
      <item>take other drugs containing prescription or nonprescription NSAIDs [aspirin, ibuprofen, naproxen, or others]</item>
    </list>
  </text>
</section>
```
Stomach bleeding warning: This product contains an NSAID, which may cause severe stomach bleeding. The chance is higher if you are age 60 or older; have had stomach ulcers or bleeding problems; take a blood thinning (anticoagulant) or steroid drug; take other drugs containing prescription or nonprescription NSAIDs [aspirin, ibuprofen, naproxen, or others].

PMID: 1f01c10a-9434-91a4-2ee4-352315a6b610|34071-1|advil|ibuprofen
AB - Stomach bleeding warning: This product contains an NSAID, which may cause severe stomach bleeding and the chance is higher if you are age 60 or older;
Unified Medical Language System

• UMLS is a medical ontology developed in National Library of Medicine
  – Having more than 1 million concepts
  – Concepts are connected with several relations: ISA, ASSOCIATED_WITH, AFFECTS...

• Relations defined in UMLS can connect concepts with specific relations, otherwise unrelated.
Using MetaMap

II. UMLS Ontology Services
DESCRIPTION

ADVAIR DISKUS 100/50, ADVAIR DISKUS 250/50, and ADVAIR DISKUS 500/50 are combinations of fluticasone propionate and salmeterol xinafoate. One active component of ADVAIR DISKUS is fluticasone propionate, a corticosteroid having the chemical name S-(fluoromethyl) 6α,9-difluoro-11β,17-dihydroxy-16α-methyl-3-oxoandrosta-1,4-diene-17β-carbothioate, 17-propionate and the following chemical structure:


III. NLP Services
Certainty Factor

• Expert knowledge is often expressed with certainty factor.
  – Using words, “most probable”, “probably”, “probably not”, “improbable”.
  – In DailyMed, certainty factors are expressed as: “chance is higher”, “more likely”, “low chance”, “less likely”.
    • “chance is higher”, “more likely” -> CAN_CAUSE(A, B, “high”).
    • “low chance”, “less likely” -> CAN_CAUSE(A,B, “low”).
    • No certainty factor -> CAN_CAUSE(A,B, “middle”).

III. NLP Services
Using NLP with certainty factor

SYNONYM(“Advil”, “Ibuprofen”).
CONTAINS(“Advil”, “NSAID”).

\[\text{CAN\_CAUSE}(\text{Patient “Gastric bleeding”, “middle”}) \text{ IF}
\quad \text{TAKES}(\text{Patient, “NSAID”}).\]

\[\text{CAN\_CAUSE}(\text{Patient “Gastric bleeding”, “high”}) \text{ IF}
\quad \text{TAKES}(\text{Patient, “NSAID”}) \text{ and IS\_OLDER}(\text{Patient, 60}).\]

\[\text{CAN\_CAUSE}(\text{Patient “Gastric bleeding”, “high”}) \text{ IF}
\quad \text{TAKES}(\text{Patient, “NSAID”}) \text{ and TAKES}(\text{Patient, “Steroid drug”}).\]
PUTTING THINGS TOGETHER FOR INFERENCE

SYNONYM(“Advil”, “Ibuprofen”).
CONTAINS(“Advil”, “NSAID”).

CAN_CAUSE($Patient “Gastric bleeding”, “middle”) IF TAKES($Patient, “NSAID”).
CAN_CAUSE($Patient “Gastric bleeding”, “high”) IF TAKES($Patient, “NSAID”) and age($Patient) > 60.
CAN_CAUSE($Patient “Gastric bleeding”, “high”) IF TAKES($Patient, “NSAID”) and TAKES($Patient, “Steroid drug”).
CONTAINS (“Advair”, “Fluticasone propionate”).
ISA(“Fluticasone propionate”, “Corticosteroid”).
ISA(“Corticosteroid”, “Steroid drugs”).

Inference reasoning triggered by prescription of Advil

Advil can cause gastric bleeding for the patient who takes Advair and the risk is high.
• Rule for contraindication
  – Sample sentence: Aldactone is contraindicated for patients with anuria, acute renal insufficiency, significant impairment of renal excretory function, or hyperkalemia.

• Pattern:

• Generate logical statements from this pattern
  – CONTRAINDICATE($X, phsu) IF HAS_DISORDER($X, dysn or patf).
  – CONTRAINDICATE($X, "bronchospasm", "middle") IF HAS_DISORDER ($X, "anuria") or ... or HAS_DISORDER ($X, "hyperkalemia")
  – phsu: Pharmacological Substance
  – dsyn: Disease and Syndrome
  – patf: Pathologic Function

III. NLP Services
• Rule for allergy
  – Sample sentence: Ibuprofen may cause a severe allergic reaction, especially in people allergic to aspirin.

• Pattern:
  – [phsu or orch]–[cause]–[dsyn]–[in]–[people]–[allergic]–[to]–[phsu or orch]

• Generate logical statements from this pattern
  – CAN_CAUSE($X, "allergic reaction“, “high”) IF TAKES($X, "ibuprofen") and ALLERGIC_TO($X, “aspirin”)

III. NLP Services
Physician involvement in NLP

• NLP conversion is a complicated process and needs expert’s intervention and feedback.
  – All the rules are reviewed by physicians who have expertise in medical informatics.

• The result of NLP conversion is also reviewed by experts and thoroughly tested with comprehensive test set.
Maintaining Microtheories

- Microtheory is a set of rules and knowledge treated independently from one another in OpenCyc.
  - Each microtheory can have knowledge conflicting with one another.
  - DailyMed information is converted into rules and maintained as a big microtheory.
  - Each patient information is translated in a separate microtheory and interact with DailyMed microtheory.
Google Health will retire on January 1st of 2012.
Detailed Medical Information in Google PHR

- **Profile ID**: PJXQp1n6bl
- **Demographics**: Age: 55 years old, Sex: Male
- **Problem List**:
  - Familial Combined Hyperlipidemia
  - Glaucoma
  - Hypertension
  - Hypothyroidism
  - Problems:
  - Take two tablets of curcumin every day
  - Apply tear drop
  - Check the blood pressure every day
  - Take synthroid every morning
- **Medication List**:
  - Metoprolol Succinate
  - Synthroid
  - Timolol Maleate
  - Medications:
  - 50 mg, By mouth - 1 tablet sustained release 24 hr, 1 time per day in the morning
  - 137 mcg, By mouth - 1 tablet, 1 time per day
  - 0.25 %, To eyes - 1 drops, 1 time per day in the morning
- **Allergy List**: Peanut, Penicillin G
XML output from Google PHR

- <Profile ID="90x5Fk_dg"/>
- <Profile ID="7lyoOYp_T4"/>
- <Profile ID="P7XJQp1n6F1">
  - <DateOfBirth>1955-10-18</DateOfBirth>
  - <Gender>Male</Gender>
- <Problems Count="4">
  - <Problem:Familial Combined Hyperlipidemia|277.4:ICD9_Broader|9.137221:Google</Problem>
  - <Problem:Hypothyroidism|244.9:ICD9|584:FDB|259.9:ICD9_Broader|9.7212:Google|244.8:ICD9_Narrower</Problem>
</Problems>
- <Medications Count="9">
</Medications>
- <Alerts Count="2">
  - <Alert:Peanut|568-6:FDB|8.10547:Google</Alert>
  - <Alert:Penicillin_G|4977-6:FDB|8.10627:Google</Alert>
</Alerts>
- <Results Count="2">
  - <Result:Exercise Treadmill Test|36.5390:Google|93015:CPT</Result>
  - <Result:Thyroid Stimulating Hormone (TSH)|7.14129:Google|84443:CPT|3016-3LOINC</Result>
</Results>
- <Procedures Count="2">
  - <Procedure:Colonoscopy|37.3454:Google|45379:CPT</Procedure>
  - <Procedure:Heart Bypass Surgery (CABG)|5.3678:Google|13510:CPT</Procedure>
</Procedures>

V. PHR Services
CAN_CAUSE($Patient, “Gastric bleeding”, “middle”) IF TAKES($Patient, “NSAID”).
CAN_CAUSE($Patient, “Gastric bleeding”, “high”) IF TAKES($Patient, “NSAID”) and IS_OLDER($Patient, 60).
CAN_CAUSE($Patient, “Gastric bleeding”, “high”) IF TAKES($Patient, “NSAID”) and TAKES($Patient, “Steroid drugs”).
COMPONENT(“Advair”, “Fluticasone propionate”).
ISA(“Fluticasone propionate”, “Corticosteroid”).
ISA(“Corticosteroid”, “Steroid drugs”).

TAKES(“John”, “Advair”).
TAKES(“John”, “Advil”).
:- CAN_CAUSE(“John”, $X, $Y).

Doctor prescribes “Advil”

CAN_CAUSE(“John”, “Gastric bleeding”, “high”).

John’s Medical Record
Age: 62
Medication: Advair

VI. ADE Services
Current Workflow

Google Health

Download John’s data

Process

John calls our application

John can have knee pain if he takes Advil and the risk is high.

Upload result

John gets the result

Java Interface

Assert to OpenCyc:
(#$take #$John #$Advair)
(#take #$John #$Advil)

Input query to OpenCyc:
"(#$canCause #$John ?X ?Y)"

Java code getting output:
while (iterator.hasNext())
{
    item = (CycConstant)iterator.next();
    results.add(item.getName());
}

OpenCyc

Cyc Microtheory 1

Assert to OpenCyc:
(#$take #$John #$Advair)
(#take #$John #$Advil)

Cyc Microtheory 2

(#$implies
((#$take ?Patient #$Steroid)
 #$and
 (#$take ?Patient #$Advair))
(#$canCause ?Patient #$kneePain #$high))

Translation done in advance using NLP

DailyMed

* All the data transfer format is ASCII and synchronous

VI. ADE Services
VII. User Interface
Thank you