

EVALUASI

TINJAUAN

- Mengevaluasi Pengujian Kegunaan dan Fungsi dari Sistem Interaktif
- Evaluasi dapat dilakukan pada :
 - Laboratorium
 - Lapangan Pekerjaan
 - Kerja Sama dengan User
- Beberapa Pendekatan Perancangan Evaluasi :
 - Metode Analitik
 - Metode Review
 - Metode Model Dasar
- Beberapa Pendekatan Evaluasi yang Telah diterapkan :
 - Metode Experimental
 - Metode Observasi
 - Metode Query
- Metode Evaluasi Yang Dipilih Secara Cermat Dan Harus Cocok Dengan Pekerjaan

EVALUASI

- Evaluasi digunakan untuk melihat apakah hasil rancangan dengan proses uji coba sistem yang telah dibuat sesuai dengan permintaan user
- Proses ini tidak dikerjakan dalam satu fase proses perancangan tetapi melalui perancangan dengan prinsip life cycle.

Tujuan Evaluasi

Ada 3 tujuan utama evaluasi :

- ***Menilai seberapa jauh sistem berfungsi.***
Desain sistem memungkinkan user melakukan tugas yang dibutuhkan dengan lebih mudah. Ini tidak hanya membuat fungsionalitas yang

sesuai ada di sistem, tetapi membuat mudah dicapai oleh user, user dapat melakukan aksi untuk melaksanakan tugas. Juga mencakup kesesuaian penggunaan sistem terhadap harapan user pada tugas tertentu. Evaluasi pada tahap ini meliputi pengukuran unjuk kerja dari user pada sistem, untuk melihat keefektifan sistem dalam mendukung tugas.

- ***Menilai efek (pengaruh) interface bagi user***

Hal ini mencakup pertimbangan aspek dari kemudahan sistem dipelajari, usability, dan perilaku user. Penting juga untuk mengidentifikasi area desain yang berlebih dari user dengan menggunakan sejumlah informasi yang berlebih.

- ***Mengidentifikasi masalah khusus yang terjadi pada sistem***

Ketika menggunakan konteks yang diinginkan menyebabkan hasil yang tidak diinginkan, atau terjadi kekacauan diantara user. Ini tentunya berhubungan dengan usability dan fungsionalitas dari desain (bergantung pada sebab masalah). Tujuan ini merupakan aspek negatif dari desain.

Jenis- Jenis Evaluasi

Ada 2 jenis utama dari evaluasi, yaitu yang dilaksanakan pada kondisi laboratorium dan yang dilaksanakan pada lingkungan kerja atau 'lapangan'

- Dalam kondisi percobaan (Laboratory)
- Dalam kondisi lokasi kerja sebenarnya
- Participatory Design

Pada Kondisi Percobaan (Laboratory)

Penggunaan pengujian sistem ini pada ruang percobaan mempunyai beberapa kondisi diantaranya :

- Laboratorium yang bagus biasanya memiliki fasilitas perekaman audio/visual yang baik, komputer beserta perlengkapannya yang mungkin tidak ada pada lokasi kerja sebenarnya.
- Operator bebas dari gangguan yang menghambat pekerjaan.
- Beberapa situasi hanya dapat dilakukan di lab, seperti sistem yang akan digunakan ditempatkan pada lokasi yang berbahaya atau lokasi yang terpencil, contoh stasiun ruang angkasa.

- Dapat memanipulasi situasi untuk memecahkan masalah dan melihat sedikit penggunaan prosedur atau membandingkan beberapa alternatif perancangan dengan situasi yang sebenarnya.
- Situasi pada laboratorium tidak dapat menggambarkan situasi ruang kerja sebenarnya dan terdapat beberapa orang yang tidak bisa bekerja pada kondisi di laboratorium

Pada Kondisi Lokasi Kerja Sebenarnya

Penggunaan pengujian sistem ini pada lokasi kerja sebenarnya mempunyai beberapa kondisi diantaranya :

- Tingkat gangguan yang melebihi ambang batas, tingkat-tingkat pergerakan yang besar dan interupsi yang tetap, seperti panggilan telepon menyebabkan observasi ini sulit dilakukan.
- Situasi yang lebih “terbuka” antara sistem dan pengguna, dimana kondisi ini tidak ditemukan pada kondisi di laboratorium
- Observasi pada lokasi kerja sebenarnya lebih baik dilakukan daripada di laboratorium dan gangguan-gangguan yang terjadi pada lokasi ini digunakan sebagai situasi yang mewakili situasi sebenarnya dan digunakan untuk proses penyimpanan dan pengambilan selama tugas.

Participatory Design

- Adalah suatu pemikiran yang melibatkan keseluruhan alur perancangan dan tidak hanya proses evaluasi saja.
- Perancangan ini dilakukan pada ruang kerja yang melibatkan pengguna yang tidak hanya digunakan sebagai subyek percobaan tetapi juga sebagai anggota yang aktif dalam team perancangan.

Mempunyai tiga karakteristik :

- Meningkatkan lingkungan kerja dan tugas
- Mempunyai sifat kerja sama, yakni pengguna dilibatkan dalam anggota team dan mempunyai kontribusi pada setiap tingkat perancangan.
- Mempunyai pendekatan iterative, perancangan adalah suatu subyek untuk evaluasi dan revisi pada setiap tingkatan.

Beberapa metode untuk membantu mengkomunikasikan informasi antara pengguna dan perancang :

- Brainstorming
- Storyboarding
- Workshops
- Pensil dan kertas percobaan

Metode-metode di atas tidak semuanya digunakan secara eksekutif dalam participatory design. Metode ini digunakan untuk memberikan pemahaman antara perancang dan user.

Mengevaluasi Perancangan

Evaluasi terjadi setelah proses perancangan. Evaluasi pertama sistem idealnya dilakukan sebelum implementasi dimulai. Jika perancangan dievaluasi, kesalahan dapat dihindari karena perancangan diubah/diperbaiki sebelumnya. Sejumlah metode dibuat untuk mengevaluasi perancangan sebelum implementasinya, yaitu :

- Cognitive Walkthrough
- Heuristic Evaluation

Cognitive Walkthrough

Suatu usaha yang dilakukan untuk mengenalkan teori psikologi ke dalam bentuk informal dan subyektif atau dengan kata lain mempunyai tujuan untuk mengevaluasi perancangan dengan melihat seberapa besar dukungan yang diberikan ke pengguna untuk mempelajari beberapa tugas yang diberikan. Pendekatan ini dikemukakan oleh Polson.

Dalam pendekatan ini terdapat beberapa issue yang timbul seperti :

- Pengaruh apa yang timbul setelah tugas ini diberikan ke pengguna?
- Proses cognitive apa yang tersedia?
- Masalah pembelajaran apa yang seharusnya timbul?

Untuk melakukan Cognitive Walkthrough harus mempunyai informasi yang dibutuhkan :

- Deskripsi dari suatu interface yang dibutuhkan itu sendiri
- Deskripsi dari tugas termasuk usaha yang benar untuk melakukannya dan struktur tujuan untuk mendukungnya

Dengan Informasi ini maka Evaluator dapat melakukan langkah dari walkthrough :

- Memilih Tugas
- Mendiskripsikan tujuan dari user
- Melakukan kegiatan / aksi yang tepat
- Menganalisa proses keputusan untuk setiap kegiatan

Contoh Cognitive Walkthrough :

Memprogram video dengan Remote Control. Misalkan akan memprogram video ke waktu dimulai dari jam 18.00 dan berakhir jam 19.15 pada channel 4 pada tanggal 4. Maka tugas tersebut akan meliputi :

- Set waktu awal
- Set waktu akhir
- Set channel
- Set tanggal

Ini adalah tujuan dari user, memasukkan pengalaman dan pengetahuan sebelumnya. Ekspresi di atas adalah pada tingkat tinggi, tidak diberikan detail pada level interface. Dengan melihat tujuan user yang diidentifikasi, setiap aksi digambarkan dengan tugas yang lengkap dan menentukan apakah interface mendukung pembangkitan struktur tujuan yang dibutuhkan untuk aksi yang lengkap. Setiap aksi dianalisa pada formulir walkthrough. Formulir menampilkan sejumlah pertanyaan bagi evaluator sebagai pertimbangan .

Dari contoh di atas hal pertama yang dilakukan adalah melakukan penekan tombol "*Timed Recording*".

Aksi #1

Deskripsi : tekan tombol "*Timed Recording*"

.....
.....

Langkah selanjutnya adalah membandingkan struktur tujuan yang tersedia dengan struktur tujuan pada langkah awal. Hal ini dilakukan dengan empat pertanyaan:

- Akankah user gagal untuk menambah tujuan yang diperlukan ?
- Akankah user gagal untuk memindahkan tujuan yang tidak diinginkan ?
- Akankah user mengadopsi beberapa dasar tujuan yang palsu dari interface ?
- Akankah user membatalkan tujuan yang masih dibutuhkan ?

Heuristic Evaluation

Hampir sama dengan Cognitive Walkthrough tetapi sedikit terstruktur dan sedikit terarah. Dalam sistem ini terdapat beberapa kriteria :

- Perilaku Sistem dapat dipastikan.
- Perilaku Sistem konsisten
- Kemampuan memori user tidak melebihi batas
- Dialog merupakan orientasi tugas

Tujuan dari Heuristic Evaluation adalah untuk memperbaiki perancangan secara efektif. Orang yang melakukan evaluasi melalui kinerja dari serangkaian tugas dengan perancangan dan penilaiannya sesuai dengan kriteria setiap tingkatan. Jika ada kesalahan terdeteksi maka perancangan dapat ditinjau ulang untuk memperbaiki masalah ini sebelum pada tingkatan implementasi.

Mengevaluasi Implementasi (model yang telah diterapkan)

Perbedaan yang besar dengan evaluasi perancangan adalah keberadaan implementasi sistem yang ada dalam berbagai bentuk. Hal ini dapat dimulai dari simulasi kemampuan interaktif dari suatu sistem, sebagai contoh Wizard of Oz, melalui fungsi prototipe dasar sampai dengan sistem yang telah diimplementasi secara keseluruhan.

Metode Empirik : Evaluasi Eksperimen

Suatu metode yang sangat kuat dalam mengevaluasi perancangan atau aspek dari perancangan adalah menggunakan eksperimen yang terkontrol. Ini menyediakan kenyataan empiris untuk mendukung klaim umum atau hipotesis.

Setiap eksperimen memiliki dasar yang sama. Evaluator memiliki hipotesa untuk diuji yang dapat ditentukan dengan mengukur beberapa atribut dari perilaku subyek. Sejumlah kondisi eksperimen ditentukan yang berbeda dalam hal nilai dari variabel tetap yang terkontrol. Setiap perubahan dalam pengukuran perilaku diatributkan ke kondisi yang berbeda. Dalam bentuk dasar ada sejumlah faktor penting terhadap keseluruhan kehandalan eksperimen yang harus dipertimbangkan dalam perancangan eksperimen.

Faktor-faktornya adalah :

- Subyek
 - Pemilihan subyek sangat penting dalam beberapa eksperimen.
 - Harus setepat mungkin dengan keinginan user.
 - Jika subyek bukan user sebenarnya maka subyek yang dipilih harus sama usia dan tingkatan pendidikan serta pengalaman menggunakan komputer secara umum dan keterhubungan sistem yang sedang dalam pengujian
- Variabel
 - Independent Variabel
Karakteristik suatu eksperimen yang memanipulasi untuk menghasilkan kondisi yang berbeda untuk perbandingan.
 - Dependent Variabel
Variabel yang dapat diukur dalam eksperimen.
- Hypothesis
 - Merupakan prediksi yang dihasilkan dalam eksperimen.
 - Masih menggunakan variabel independent dan dependent, dimana variasi di dalam independent variable akan menyebabkan perbedaan pada dependent variable.
 - Tujuan dari eksperimen adalah untuk menunjukkan bahwa perkiraan benar dengan menyangkal adanya null hypothesis, yang menyatakan bahwa tidak ada perbedaan dalam dependent variable diantara tingkatan dari suatu independent variable.

Perancangan Eksperimen

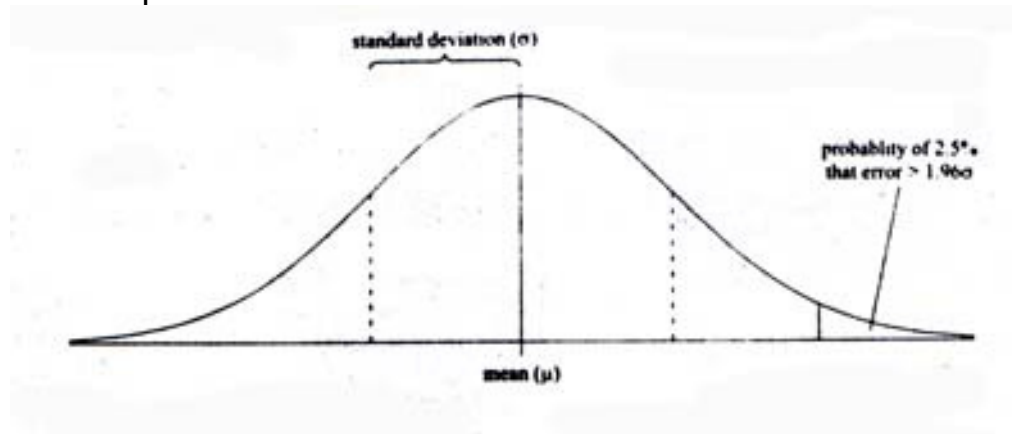
- Between-Groups (Randomized)
 - Masing-masing subyek diberikan kondisi yang berbeda yakni kondisi eksperimen dan control.
 - Keuntungan perancangan ini adalah setiap user menghasilkan satu kondisi
 - Kerugiannya adalah dengan semakin banyak jumlah subyek yang tersedia akan menyebabkan hasilnya akan berkurang dan perbedaan antar setiap individu akan membuat bias hasil. Hal ini dapat diatasi dengan memilih dengan hati-hati subyek yang dipilih dan menjamin setiap kelompok di masyarakat terwakili.
- Within-Groups
 - Setiap user akan menampilkan kondisi yang berbeda
 - Jumlah user yang tersedia lebih sedikit
 - Pengaruh dari subyek lebih sedikit

Pengukuran Statistik

- Dua aturan dalam analisa dengan statistik yakni : melihat dan menyimpan data
- Variabel yang digunakan :
 - Discrete Variables
 - Menggunakan jumlah yang terbatas dari suatu nilai atau tingkatan
Contoh: Warna layar monitor yakni : red, green atau blue.
 - Merupakan suatu independent variable
 - Continous Variable
 - Menggunakan jumlah berapapun
Contoh : Ketinggian seseorang atau waktu yang diperlukan untuk menyelesaikan pekerjaan.
 - Merupakan suatu dependent variable

	Independent Variable	Dependent Variable	
Parametric			
	2 valued	Normal	Student's test on difference of means
	Discrete	Normal	ANOVA (Analysis Of Variance)
	Continous	Normal	Linear (or non-linear) regression factor analys
Non-parametric			
	2 valued	Continous	Wilcoxon (or Mann-Whitney) rank-sum test
	Discrete	Continous	Rank-sum version of ANOVA
	Continous	Continous	Spearman's rank correlation
Contingency test			
	2 valued	Discrete	No special test, see next entry
	Discrete	Discrete	Contingency table and X squared test
	Continous	Discrete	(rare) group independent variable and then as above

Contoh parametric test



Contoh non-parametric statistik

Kondisi A : 33,42,25,79,52

Kondisi B : 87,65,92,93,91,55

Setelah diurut mulai 25,33,42,...,92,93 kemudian setelah di ranking maka 25 jadi 1, 33 jadi 2 dan seterusnya.

Kondisi A : 2,3,1,7,4

Kondisi B : 8,6,10,11,9,5

Dengan Wilcoxon test, setiap kondisi dijumlah rankingnya kemudian dikurangkan dengan nilai terkecilnya sehingga didapatkan nilai statistic U nya.

Kondisi A : $1+2+3+4+5 = 15$

Kondisi B : $1+2+3+4+5+6 = 21$

	Rank sum	smallest	U
Kondisi A	: (2+3+1+7+4)	- 15	= 2
Kondisi B	: (8+6+10+11+9+5)	- 21	= 28

Ternyata jumlah dari kedua statistik U tersebut $2 + 28 = 30$ merupakan perkalian dari jumlah nilai data dari masing-masing kondisi 5×6 .

Contoh mengevaluasi perancangan icon



Dimisalkan kita akan membuat paket pengolahan kata baru yang menggunakan icon untuk presentasinya. Ada 2 jenis perancangan icon akan digunakan yang paling mudah bagi user. Jenis pertama menggunakan gambar natural (berdasarkan metapora dari dokumen kertas), jenis yang lain menggunakan gambar abstrak seperti terlihat pada gambar di atas.

Yang pertama harus dilakukan adalah membentuk hipotesis : tampilan yang seperti apa yang diinginkan? Pada kasus ini kebanyakan user akan mengambil gambar natural sebagai icon yang lebih mudah dalam pemanggilan, karena lebih familiar bagi user. Hipotesis ini mengidentifikasi variabel independent untuk eksperimen, natural dan abstrak. Pada pertimbangan variabel dependent, pengukuran "lebih mudah" dilihat dari jumlah kesalahan dalam memilih dan waktu yang digunakan untuk memilih sebuah icon. Kontrol eksperimen dilakukan dengan membuat interface yang identik dengan perancangan icon dan tugas pemilihan dibuat yang dapat diulang untuk setiap kondisi. Kemudian metode eksperimen dilakukan dengan menggunakan kelompok user yang dimaksud. Mengingat pembelajaran terlibat dalam eksperimen ini, maka perancangan within-groups digunakan dimana setiap user menggunakan kedua set icon tersebut. Kemudian hasil eksperimen secara detail dianalisis.

Teknik Observasi

- **Think aloud**

Cara yang populer dalam mengumpulkan informasi dengan melihat interaksi user.

Evaluators look at and record user actions (with various techniques). Consequently, users are asked to elaborate on their actions with *"thinking aloud"*, that is, explain: what they believe is happening, why they are taking actions, what they are trying to do.

Think aloud has advantages, namely simple, requires little knowledge to use, and provides useful information about the interface. It can also be used to observe how the system is used.

Another variation is cooperative evaluation, where:

- The process has fewer barriers, so it is easier to learn for those who evaluate.
- Users have motivation to test the system.
- Those who evaluate can describe the confusion that occurs and maximize the approach that is effective to identify the problem.

- ***Analisa Protokol***

There are several methods to record user activities:

- Paper and pencil
- Audio recording
- Video recording
- Computer logging
- User notebook

- ***Automatic protocol analysis tools***

- EVA (Experimental Video Annotator)
System prototype that runs on multi-media workstation and is connected directly to video recorder
- Example: Workplace project at Xerox PARC

- ***Post-task walkthroughs***

There are several conditions that cause subjects to be unable to talk during observation, as if the subject is busy with a very important and many tasks. In this condition, post-task walkthroughs can observe user behavior more objectively.

Teknik Query

- ***Interview***

Menginterview user tentang pengalaman mereka dengan sistem interaktif yang menyediakan informasi secara langsung dan terstruktur.

- ***Questionnaire***

Metode alternatif yang agak kurang fleksibel dibandingkan dengan interview tetapi dapat meraih subyek yang banyak dan membutuhkan waktu yang tidak lama.

Terdapat beberapa jenis questionnaire :

- General
- Open-ended
- Scalar
- Multi-choice
- Ranked

Memilih Metode Evaluasi

Faktor yang membedakan teknik evaluasi :

Ada 8 faktor yang membedakan teknik evaluasi yang berbeda yang membantu kita dalam memilih teknik yang sesuai, yaitu :

- Tingkat siklus
- Jenis evaluasi
- Tingkat obyektivitas dan subyektivitas
- Jenis ukuran yang tersedia
- Informasi yang tersedia
- Kesiapan dari suatu respon
- Tingkat gangguan yang tidak secara langsung
- Sumber yang tersedia

Design VS. Implementation

Evaluasi pada tingkatan perancangan hanya membutuhkan ahlinya dan akan dianalisa sedangkan evaluasi pada tingkat implementasi membawa user sebagai subyek dari eksperimen.

Laboratory VS. Field studies

Secara ideal proses perancangan mencakup dua jenis evaluasi ini dimana pada laboratorium merupakan awal tingkatan sedangkan field studies ditempatkan pada tingkatan implementasi.

Subyektif VS. Obyektif

Pada teknik yang subyektif seperti pada cognitive walkthrough maupun think aloud mempercayakan pada evaluator atau orang yang melakukan evaluasi. Hal ini dapat menyebabkan penilaian evaluator menjadi bias sehingga diperlukan lebih dari satu orang evaluator. Sedangkan teknik obyektif seharusnya menghasilkan hasil yang berulang yang tidak tergantung pada evaluator. Eksperimen yang terkontrol merupakan contoh pengukuran obyektif. Secara ideal pengukuran secara subyektif dan obyektif seharusnya digunakan keduanya.

Qualitative VS Quantitative measures

Pengukuran menggunakan numerik lebih mudah digunakan teknik statistik sedangkan non-numerik lebih sulit dilakukan akan tetapi menghasilkan sesuatu yang penting secara lengkap yang tidak dapat dilakukan secara numerik.

Informasi yang tersedia

Tingkatan informasi yang tersedia untuk evaluator pada setiap tingkat proses perancangan dimulai pada informasi yang rendah sampai dengan tingkatan yang tinggi.

Kesiapan suatu tindakan

Faktor lain yang membedakan teknik evaluasi adalah kesiapan mengambil tindakan yang tersedia. Seperti metode think aloud yang merekam perilaku user pada waktu interaksi. Sedangkan post task walkthrough mempercayakan pada pengumpulan kembali kegiatan user.

Intrusiveness

Merupakan teknik yang menghasilkan pengukuran segera yang dihasilkan user selama interaksi dan mempunyai resiko mempengaruhi perilaku user. Evaluator harus sensitif untuk mengurangi masalah ini tetapi tidak dapat mengubah semuanya.

Sumber-sumber

Hal terakhir yang perlu diperhatikan dalam memilih teknik evaluasi yakni : peralatan, waktu, biaya, subyek serta evaluator yang ahli.